

Robust Methods in Univariate Time Series Models

Emily J. Whitehouse



The University of
Nottingham

*Thesis submitted to the University of Nottingham for the degree of Doctor of
Philosophy*

School of Economics
April 2017

Abstract

The size and power properties of a hypothesis test typically depend on a series of factors which are unobservable in practice. A branch of the econometric literature therefore considers robust testing methodologies that achieve good size-control and competitive power across a range of differing circumstances. In this thesis we discuss robust tests in three areas of time series econometrics: detection of explosive processes, unit root testing against nonlinear alternatives, and forecast evaluation in small samples.

Recent research has proposed a method of detecting explosive processes that is based on forward recursions of OLS, right-tailed, Dickey-Fuller [DF] unit root tests. In Chapter 2 an alternative approach using GLS DF tests is considered. We derive limiting distributions for both mean-invariant and trend-invariant versions of OLS and GLS variants of the Phillips, Wu and Yu (2011) [PWY] test statistic under a temporary, locally explosive alternative. These limits are dependent on both the value of the initial condition and the start and end points of the temporary explosive regime. Local asymptotic power simulations show that a GLS version of the PWY statistic offers superior power when a large proportion of the data is explosive, but that the OLS approach is preferred for explosive periods of short duration. These power differences are magnified by the presence of an asymptotically non-negligible initial condition. We propose a union of rejections procedure that capitalises on the respective power advantages of both OLS and GLS-based approaches. This procedure achieves power close to the effective envelope provided by the two individual PWY tests across all settings of the initial condition and length of the explosive period considered in this chapter. We show that these results are also robust to the point in the sample at which the temporary explosive regime occurs. An application of the union procedure to NASDAQ daily prices confirms the empirical value of this testing strategy.

Chapter 3 examines the local power of unit root tests against globally stationary exponential smooth transition autoregressive [ESTAR] alternatives under two sources of uncertainty: the degree of nonlinearity in the ESTAR model, and the presence of a linear deterministic trend. First, we show that the Kapetanios, Shin and Snell (2003, *Journal of Econometrics* 112, 359–379) [KSS] test for nonlinear stationarity has local asymptotic power gains over standard Dickey-Fuller [DF] tests for certain degrees of nonlinearity in the ESTAR model, but that for other degrees of nonlinearity, the linear DF test has superior power. Second, we derive limiting distributions of demeaned, and demeaned and detrended KSS and DF tests under a local ESTAR alternative when a local trend is present in the DGP. We show that the power of the demeaned tests outperforms that of the detrended tests when no trend is present in the DGP, but deteriorates as the magnitude of the trend increases. We propose a union of rejections testing procedure that combines all four individual tests and show that this captures most of the power available from the individual tests across different degrees of nonlinearity and trend magnitudes. We also show that incorporating a trend detection procedure into this union testing strategy can

result in higher power when a large trend is present in the DGP. An empirical application of our proposed union of rejections procedures to energy consumption data in 180 countries shows the value of these procedures in practical cases.

In Chapter 4 we show that when computing standard Diebold-Mariano-type tests for equal forecast accuracy and forecast encompassing, the long-run variance can frequently be negative when dealing with multi-step-ahead predictions in small, but empirically relevant, sample sizes. We subsequently consider a number of alternative approaches to dealing with this problem, including direct inference in the problem cases and use of long-run variance estimators that guarantee positivity. The finite sample size and power of the different approaches are evaluated using an extensive Monte Carlo simulation exercise. Overall, for multi-step-ahead forecasts, we find that the recently proposed Coroneo and Iacone (2015) test, which is based on a weighted periodogram long-run variance estimator, offers the best finite sample size and power performance.

Acknowledgments

First and foremost, I would like to thank my supervisors, Professor Steve Leybourne and Professor David Harvey, for the invaluable guidance and mentorship they have provided. Their good humour, unpretentious advice and constant encouragement have made my PhD studies a rewarding and enjoyable experience. I couldn't have asked for better supervisors.

I am also grateful to Dr Patrick Marsh for his sound advice and encouragement throughout my time at Nottingham. I was fortunate enough to spend three months of my PhD as a visiting graduate student at the University of California, San Diego and I am grateful to Professor Graham Elliott for hosting my visit.

Financial support in the form of an ESRC Studentship and Overseas Visit Award are gratefully acknowledged.

Finally, an acknowledgement must go to my mother and father. Whilst their interest in time series econometrics is limited, their support has been endless.

Contents

1	Introduction	1
2	Explosive asset price bubble detection with uncertainty over the bubble length and initial condition	6
2.1	Introduction	7
2.2	Testing for asset price bubbles	9
2.3	The presence of an initial condition	11
2.4	Right-tailed unit root tests	12
2.5	Asymptotic behaviour of tests	13
2.6	A union of rejections strategy	19
2.7	Finite sample power comparison	21
2.8	Empirical application	22
2.9	Conclusion	23
	Appendix: Proof of Lemma 2.1	25
3	Testing for a unit root against ESTAR stationarity	52
3.1	Introduction	53
3.2	Nonlinear stationarity	54
3.3	Trend uncertainty	57
3.4	Unit root tests and the ESTAR model	58
3.5	Asymptotic properties of tests	60
	3.5.1 Asymptotic behaviour under a local trend	62
	3.5.2 Asymptotic behaviour under a fixed trend	64
3.6	A union of rejections strategy	65
	3.6.1 Union of DF and KSS	65
	3.6.2 Union of demeaned, and demeaned and detrended tests . .	68
	3.6.3 Asymptotic power under fixed c	70
3.7	Trend detection	71
3.8	Finite sample simulations	73
3.9	Empirical example	74
3.10	Conclusion	78
	Appendix: Proof of Lemma 3.1	100
	Appendix: Proof of Lemma 3.2	108

4	Testing equal forecast accuracy and forecast encompassing in small samples	112
4.1	Introduction	113
4.2	Forecast evaluation	115
4.3	Standard tests for equal accuracy and encompassing	118
4.4	Frequency of negative long-run variance estimates	121
4.5	Adjusted Diebold-Mariano type tests	124
4.6	Finite sample size and power	129
	4.6.1 Forecast accuracy	130
	4.6.2 Forecast encompassing	135
	4.6.3 Simulations using calibrations from empirical data	136
4.7	Conclusion	138
5	Conclusion	151

List of Tables

2.1	Asymptotic critical values for PWY_{OLS}^μ , PWY_{GLS}^μ , PWY_{OLS}^β , PWY_{GLS}^β , and λ_ζ^μ and λ_ζ^β values	43
2.2	Finite sample critical values, $T = 150$	43
2.3	Application of PWY_{OLS}^μ , PWY_{GLS}^μ , PWY_{OLS}^β , PWY_{GLS}^β , U^μ , U^β to NASDAQ prices	43
3.1	Asymptotic critical values of KSS_μ , DF_μ , KSS_τ , DF_τ and union of rejection scaling values λ_μ^ζ , λ_τ^ζ and λ_4^ζ for significance level ζ . .	80
3.2	Finite sample sizes of KSS_μ , DF_μ , KSS_τ , DF_τ and union of rejection procedures U_μ , U_τ , U_4 and U^* for significance level ζ and $T = 100$	80
3.3	Finite sample sizes of U_4 with GS and $MAIC$ lag selection methods for $\zeta = 0.05$	80
3.4	Total primary energy consumption per capita: DF_μ , DF_τ , KSS_μ , KSS_τ , U_μ , U_τ , U_4 and U^* unit root test results	81
3.5	Total primary energy consumption per capita: countries for which no unit root test rejections found	83
4.1	Frequency of negative long-run variance estimates in tests for equal forecast accuracy	140
4.2	Frequency of negative long-run variance estimates in tests for forecast encompassing	140
4.3	Empirical size of nominal 0.10-level tests for equal forecast accuracy, $\rho = 0$	141
4.4	Empirical size of nominal 0.10-level tests for equal forecast accuracy, $\rho = 0.5$	142
4.5	Empirical size of nominal 0.10-level tests for equal forecast accuracy, $\rho = 0.9$	143
4.6	Empirical size of nominal 0.10-level tests for equal forecast accuracy with heavy-tailed errors, $\rho = 0$	144
4.7	Empirical size of nominal 0.10-level tests for equal forecast accuracy with negative MA coefficients, $\rho = 0$	145
4.8	Size-adjusted power of nominal 0.10-level tests for equal forecast accuracy, $\rho = 0$	146
4.9	Empirical size of nominal 0.10-level tests for forecast encompassing	147
4.10	Size-adjusted power of nominal 0.10-level tests for forecast encompassing	148
4.11	Fitted MA coefficients of fe_t^4 , fe_t^8 and fe_t^{12}	149

4.12	Frequency of negative long-run variance estimates in tests for equal forecast accuracy for $h = 4, 8, 12$	149
4.13	Empirical size of 0.10 level tests under calibrated MA errors . . .	149
4.14	Size-adjusted power of 0.10 level tests under calibrated MA errors	149

List of Figures

2.1	Asymptotic power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.45, 0.55]$, where $i = \{\mu, \beta\}$	44
2.2	Asymptotic power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.2, 0.8]$, where $i = \{\mu, \beta\}$	45
2.3	Asymptotic power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.45, 0.55]$, where $i = \{\mu, \beta\}$	46
2.4	Asymptotic power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.2, 0.8]$, where $i = \{\mu, \beta\}$	46
2.5	Finite sample power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.45, 0.55]$, where $i = \{\mu, \beta\}$	47
2.6	Finite sample power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.2, 0.8]$, where $i = \{\mu, \beta\}$	48
2.7	Finite sample power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.15, 0.25]$, where $i = \{\mu, \beta\}$	49
2.8	Finite sample power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.75, 0.85]$, where $i = \{\mu, \beta\}$	50
2.9	Logarithms of NASDAQ closing prices	51
3.1	Asymptotic size and local power of KSS_μ , DF_μ and U_μ for fixed g^2 ($\kappa = 0$)	84
3.2	Asymptotic size and local power of KSS_τ , DF_τ and U_τ for fixed g^2	85
3.3	Asymptotic size and local power of U_μ , U_τ and U_4 for fixed $g^2 = 2$	86
3.4	Asymptotic size and local power of U_μ , U_τ and U_4 for fixed $g^2 = 10$	87
3.5	Asymptotic size and local power of U_μ , U_τ and U_4 for fixed $g^2 = 50$	88
3.6	Asymptotic size and local power of KSS_μ , DF_μ and U_μ for fixed c ($\kappa = 0$)	89
3.7	Asymptotic size and local power of KSS_τ , DF_τ and U_τ for fixed c	90
3.8	Asymptotic size and local power of U_μ , U_τ and U_4 for fixed $c = 20$	91
3.9	Asymptotic size and local power of U_μ , U_τ and U_4 for fixed $c = 60$	92
3.10	Asymptotic size and local power of U_μ , U_τ and U_4 for fixed $c = 100$	93
3.11	Finite sample size and local power of KSS_μ , DF_μ and U_μ for fixed g^2 ($\beta = 0$)	94
3.12	Finite sample size and local power of KSS_τ , DF_τ and U_τ for fixed g^2	95
3.13	Finite sample size and local power of U_μ , U_τ , U_4 and U^* for fixed $g^2 = 2$	96
3.14	Finite sample size and local power of U_μ , U_τ , U_4 and U^* for fixed $g^2 = 10$	97

3.15	Finite sample size and local power of U_μ , U_τ , U_4 and U^* for fixed $g^2 = 50$	98
3.16	Total energy consumption per capita (logs of million BTU per person)	99
4.1	$\widehat{C}(k)$ and γ_k values for $T = 8, h = 6$	150
4.2	Bias Correction Factor (CF)	150

Chapter 1

Introduction

Hypothesis testing relies on assumptions which often fail to hold in practice. These assumptions may be about the distribution of the data, the presence of nuisance parameters in an econometric model, or the underlying dynamics of a time series. Whilst a practitioner may be aware of the set of conditions under which an econometric test performs adequately, it is not always possible to confirm that the data series of interest satisfies those assumptions.

When evaluating the performance of a test, we are typically concerned with two properties: size and power. Size is defined as the probability that the test incorrectly rejects the null hypothesis; whilst power is the probability that the null hypothesis is correctly rejected when the alternative is true. Comparing two or more competing test procedures, a practitioner would favour a test with superior size-control and higher power.

Deriving the asymptotic distribution of a test statistic allows us to identify the nuisance parameters which affect the size and power properties of that test. When these nuisance parameters are unobservable, it will not be clear which of a group of competing test procedures achieves the best overall performance, and therefore which test a practitioner should choose to employ. For example, as demonstrated by Harvey et al. (2009), the relative performance of mean-invariant and trend-invariant unit root tests depends on the presence or otherwise of a linear trend in the time series. A branch of the econometric literature has therefore focused on designing statistical test procedures which are robust to uncertainty about whether underlying assumptions such as these hold in practice. In Chapters 2 and 3 of this thesis, we consider the limiting distributions of different test procedures for explosive processes and unit root tests respectively, and identify a series of factors which can have an adverse effect on the performance of these tests. In both cases, we propose union of rejections testing procedures, in line with Harvey et al. (2009, 2012), which are robust to uncertainty about these factors.

The size and power properties of a consistent test procedure will be dependent

upon the sample size of observations used to compute the test statistic, with increases in the sample size yielding tests with better size-control and higher power. Tests which rely on asymptotic critical values can often suffer from size-distortion when only a small number of observations are available. The issue of small sample sizes is of particular relevance to forecast evaluation exercises when a practitioner is typically required to split their data series into an in sample period for model estimation, and an out of sample period for forecast evaluation. We explore this problem in Chapter 4, in the context of tests for equal forecast accuracy and forecast encompassing. A modified version of the Diebold and Mariano (1995) procedure, proposed by Coroneo and Iacone (2015), is shown to be robust to small, but empirically relevant, sample sizes.

This thesis therefore contains three substantive and self-contained chapters which each focus on an area of time series econometrics in which we identify a failing of current statistical tests. Each chapter suggests a new test procedure that is robust to the problems identified. Chapter 2 considers the detection of explosive processes; Chapter 3 considers unit root testing against an alternative hypothesis of nonlinear stationarity; and Chapter 4 examines forecast evaluation techniques in small samples.

In Chapter 2 we propose a test procedure for detecting temporary explosive processes which is robust to the length of the explosive period as a proportion of the total sample, and the value of the initial condition. Recent events such as the Dot-Com bubble of the late 1990s and the global financial crisis of 2007-2009 have highlighted the need for improved theoretical modelling of asset bubbles and econometric techniques which can detect those bubbles. A key finding of the financial asset bubble literature is that time series in which an asset price bubble is present follows an explosive process and can therefore be modelled by a simple first order autoregression [AR]. In recent work, Phillips et al. (2011) [PWY] propose a method of detecting explosive processes that is based on forward recursions of OLS, right-tailed, Dickey-Fuller [DF] unit root tests. These right-tailed unit root tests consider the null hypothesis of a unit root against the alternative of an explosive process.

In this chapter an alternative approach using GLS DF tests is considered. The asymptotic distributions of both mean-invariant and trend-invariant versions of OLS and GLS variants of the PWY test statistic are derived under a temporary, locally explosive alternative. We show that these limits are dependent on both the value of the initial condition and the start and end points of the temporary explosive regime. This implies that the value of the initial condition and the length of the explosive period will affect test power. In fact, local asymptotic

power simulations show that a GLS version of the PWY statistic offers superior power when a large proportion of the data is explosive, but that the OLS approach is preferred for explosive periods of short duration. An increase in the value of the initial condition increases the power differences between the two tests. We therefore propose a union of rejections procedure that captures the relative power advantages of both OLS and GLS based approaches, achieving power close to the effective envelope provided by the two individual PWY tests across all settings of the initial condition and length of the explosive period considered in this chapter. These results are also shown to be robust to the point in the sample at which the temporary explosive regime occurs. An application of the union procedure to NASDAQ daily prices confirms the empirical value of this testing strategy.

Chapter 3 proposes a test for a unit root against an alternative hypothesis of a nonlinear, but globally stationary, exponential smooth transition autoregressive (ESTAR) process. The application of standard Dickey-Fuller [DF] unit root tests to certain economic time series, such as real exchange rates or real interest rates, has sometimes led to surprising results. Economic theory suggests that these series should be stationary processes, but many studies have been unable to reject the unit root null leading to concern that the DF unit root test has low power against some classes of stationary alternatives. To address this problem, Kapetanios et al. (2003) [KSS] propose a test of the null hypothesis of a unit root against an ESTAR alternative. This procedure is shown to have power gains over standard DF tests under some ESTAR alternatives, and the test has subsequently been frequently employed in the applied econometric literature.

In this chapter we examine the local power of the KSS and DF unit root tests against nonstationary ESTAR alternatives under two sources of uncertainty: the degree of mean reversion in the ESTAR model; and the presence of a linear deterministic trend. First, we demonstrate that the KSS test for nonlinear stationarity has local asymptotic power gains over standard DF tests for some degrees of nonlinearity in the ESTAR model, but that for other degrees of nonlinearity, the linear DF test has superior power. Second, we derive limiting distributions of mean-invariant and trend-invariant KSS and DF tests under a local ESTAR alternative when a local trend is present in the DGP. The power of the mean-invariant tests outperforms that of the trend-invariant tests when no trend is present in the DGP, but deteriorates as the magnitude of the trend increases. We propose a union of rejections testing procedure which combines inference from all four individual tests. This procedure captures most of the power available from the individual tests across different degrees of nonlinearity and trend magnitudes. Incorporating a trend detection procedure into this union testing strategy is shown

to result in higher power when a large trend is present in the DGP. An empirical application of the proposed union of rejections procedures to energy consumption data in 180 countries demonstrates the practical use of these tests.

In Chapter 4 we examine the Diebold and Mariano (1995) [DM] and Harvey et al. (1997) [HLNa] tests for equal forecast accuracy and the Harvey et al. (1998) [HLNb] test for forecast encompassing when only a small number of out of sample forecast observations are available. Forecasting plays a critical role in economic and financial research and policy-making. The evaluation of two or more competing forecasts of the same economic time series has therefore become an extensive field in the econometric literature. One of the most common evaluation exercises undertaken is to compare the accuracy of competing forecasts on the basis of some measure of loss associated with the forecast errors, such as mean squared forecast error. DM formalise this approach by proposing a statistical test which considers a null hypothesis of equal accuracy against the alternative that one forecast is more accurate than another. Harvey et al. (1997) [HLNa] demonstrate that the DM test can be oversized when only a small number of out of sample forecasts are available, and propose a finite sample bias correction and use of different critical values to better control size in these small samples. Another forecast evaluation exercise frequently undertaken is to test for whether one set of forecasts encompasses another. It may be the case that regardless of one forecast's superior predictive accuracy, a second, less favoured, forecast still contains information which is relevant and valuable. In these circumstances a combination of two or more forecasts may be optimal, rather than choosing to use a single forecast in isolation. Harvey et al. (1998) propose a DM-type test where the loss differential is redefined to test a forecast encompassing null hypothesis. Application of DM/HLNa and HLNb tests have become prevalent in empirical forecasting research, such that it is now routine for the results of these tests to be reported alongside any forecast comparisons.

These tests are fundamentally comprised of a mean of a loss differential series divided by an estimate of the long-run variance. The first contribution of this chapter is to show that this long-run variance estimate can frequently be negative when dealing with multi-step-ahead forecast horizons and a small number of out of sample observations. In practical applications, it is not uncommon for tests of equal accuracy and encompassing to be conducted using sample size and forecast horizon settings that lie in the region where negative variance estimates occur frequently. We therefore require a test procedure which delivers reliable small sample size and power properties. The second contribution of this chapter is to assess a number of different strategies for dealing with the problem of negative

long-run variance estimation in tests of equal accuracy and encompassing. An extensive Monte Carlo simulation exercise is conducted to establish the small sample size and power properties of these different approaches. In general, we find that the recently proposed Coroneo and Iacone (2015) approach outperforms the other tests considered, with the attractive finite sample properties reported in their paper for moderate sample sizes and forecast horizons extending to the small sample and long horizon region under focus in this chapter. The empirical relevance of our results is confirmed through an application of the tests in this chapter to errors calibrated from forecast models of Euro-area inflation.

Chapter 5 summarises the main results of the thesis and concludes.

Chapter 2

Explosive asset price bubble detection with uncertainty over the bubble length and initial condition

Recent research has proposed a method of detecting explosive processes that is based on forward recursions of OLS, right-tailed, Dickey-Fuller [DF] unit root tests. In this chapter an alternative approach using GLS DF tests is considered. We derive limiting distributions for both mean-invariant and trend-invariant versions of OLS and GLS variants of the Phillips, Wu and Yu (2011) [PWY] test statistic under a temporary, locally explosive alternative. These limits are dependent on both the value of the initial condition and the start and end points of the temporary explosive regime. Local asymptotic power simulations show that a GLS version of the PWY statistic offers superior power when a large proportion of the data is explosive, but that the OLS approach is preferred for explosive periods of short duration. These power differences are magnified by the presence of an asymptotically non-negligible initial condition. We propose a union of rejections procedure that capitalises on the respective power advantages of both OLS and GLS-based approaches. This procedure achieves power close to the effective envelope provided by the two individual PWY tests across all settings of the initial condition and length of the explosive period considered in this chapter. We show that these results are also robust to the point in the sample at which the temporary explosive regime occurs. An application of the union procedure to NASDAQ daily prices confirms the empirical value of this testing strategy.

2.1 Introduction

Detection of asset price bubbles in financial and macroeconomic time series data is an increasingly prominent issue in economics. Recent events such as the Dot-Com bubble of the late 1990s and the global financial crisis of 2007-2009 have highlighted the detrimental and wide-ranging effect that a bubble in a single asset market can have on the whole economy. A substantial body of literature has therefore emerged which attempts both to improve our theoretical modelling of asset price bubbles and to design statistical tests which can detect these bubbles. A key finding of this literature is that time series in which an asset price bubble is present follows an explosive process and can therefore be modelled by a simple first order autoregression [AR].

In recent work, Phillips et al. (2011) [PWY] propose a test for the presence of explosive behaviour in time series that is based on the supremum of recursive, right-tailed unit root tests. Their test allows for temporary explosiveness, such that a bubble emerges and collapses within the sample period. The PWY test employs OLS Dickey-Fuller [DF] unit root tests. When considering full sample explosiveness, Harvey and Leybourne (2014) show that a DF unit root test based on Generalised Least Squares [GLS] demeaning or detrending has a power advantage over OLS DF tests. This result holds when the initial condition of the series, i.e. the deviation of the first observation of the sample away from the deterministic of the process, is both asymptotically negligible and non-negligible.

In this chapter we evaluate the properties of both OLS and GLS versions of the PWY statistic in detecting temporary explosiveness. Whilst PWY focus on mean-invariant tests only, we also consider a trend-invariant procedure. Many time series that are potentially subject to asset price bubbles, such as stock market prices or house prices, may follow a linear trend. It is well known that the power of a demeaned unit root test is significantly affected by the presence of a trend. Therefore trend-invariant tests for explosive processes are likely to be required in addition to mean-invariant tests. We derive the limiting distributions of the two mean-invariant and the two trend-invariant PWY statistics under a temporary, locally explosive alternative, in the presence of an asymptotically non-negligible initial condition. These limits are shown to be dependent on both the value of the initial condition and the start and end points of the temporary explosive regime. We simulate asymptotic and finite sample power for OLS and GLS PWY statistics under varying initial conditions, lengths of the explosive period, and points in the sample at which the explosive period occurs. Our results show that the GLS version of the PWY statistic outperforms its OLS

counterpart when a long explosive period is considered, in line with the findings of Harvey and Leybourne (2014). However, when considering an explosive period of shorter length, we show that the OLS version of the PWY statistic has superior power. The presence of an asymptotically non-negligible initial condition does not affect the ranking of the two procedures, but increases the power difference between tests.

As the relative power performance of OLS and GLS based PWY tests depends on two factors which are likely to be unobserved in practice: the length of the explosive regime and the value of the initial condition, we propose a union of rejections procedure to capitalise on the respective power advantages of the two tests. This union procedure displays power close to the effective envelope provided by the two individual tests for all settings of the initial condition and bubble length considered in this chapter. An application of the union procedure to NASDAQ prices confirms the empirical usefulness of this testing strategy.

Throughout this chapter, we refer to both the *length* of the explosive period and the *magnitude* of the explosive process. The length of the explosive period describes the proportion of the sample for which the data follows an explosive process, as opposed to a unit root. The magnitude of the explosive process describes the growth rate of the explosive series. The magnitude is affected by both the parameter in the AR model, as well as the value of the initial condition.

Here and throughout the thesis we use the following notation: $\lfloor \cdot \rfloor$ to denote the integer part of the argument; $\xrightarrow{p}$ and $\xrightarrow{d}$ to denote convergence in probability and weak convergence, respectively as the sample size diverges.

In the following section we outline some of the key contributions of the previous literature on testing for asset price bubbles. Section 2.3 describes the role the initial condition plays in explosive processes. Section 2.4 outlines the PWY test procedure and discusses a GLS variant. In Section 2.5 local asymptotic distributions of mean-invariant and trend-invariant OLS and GLS tests are derived and results from asymptotic simulations are reported for differing values of the initial condition and length of the explosive period. Section 2.6 proposes a union of rejections procedure and reports the local asymptotic power of this procedure. Section 2.7 evaluates the finite sample performance of all test statistics considered in this chapter. An empirical application of the test procedures in this chapter to NASDAQ prices is considered in Section 2.8. Section 2.9 concludes.

2.2 Testing for asset price bubbles

The efficient market hypothesis states that the price of an asset perfectly reflects the intrinsic value of that asset. Holding an asset represents a claim on all future cash-flows from that asset. In the context of stock markets, present discounted values of expected future dividends are therefore considered to represent the intrinsic or fundamental value of the stock. Any movement in a stock's price are said to be a result of receiving new information which adjusts the investors' expectations of future dividends.

A number of studies find that the behaviour of stock prices and dividends are inconsistent with this hypothesis. Volatility tests of Shiller (1981), Leroy and Porter (1981) and West (1988) show that the volatility of stock prices is too high to be explained by uncertainty about future dividends. While deviations of prices away from their market fundamentals has often been construed as evidence of irrationality, Blanchard and Watson (1982) suggest that it is perfectly possible to have rational deviations from fundamentals in the form of rational bubbles. These rational bubbles represent a belief that there are factors other than market fundamentals which influence prices. However, *inter alios*, Flood and Garber (1980); Hamilton and Whiteman (1985); Hamilton (1986) note that these results do not provide sufficient evidence that rational bubbles exist. Instead, it may be that the true market fundamentals are unobservable. Alternative methods for detecting rational bubbles that examine the properties of both prices and fundamentals have therefore been developed.

The present value theory of finance can be used to demonstrate the influence of a rational bubble on prices. Starting with the standard no arbitrage condition, the real stock price, P_t , can be written as

$$P_t = \frac{1}{1+r} E_t (P_{t+1} + D_{t+1})$$

where r is the risk-free discount rate ($r > 0$), D_t is the real dividend and E_t is the expectation at time t . Through recursive substitution of this condition it is shown that

$$P_t = P_t^f + B_t$$

where P_t^f denotes the market fundamentals component of prices, and B_t a bubble component. The fundamentals component is given by the discounted sum of future dividends.

$$P_t^f = \sum_{i=1}^{\infty} (1+r)^{-i} E_t (D_{t+i})$$

If the stochastic difference equation

$$B_{t+1} = (1 + r)B_t + u_t$$

holds, where $E_{t-i}(u_t) = 0$ for all $i \geq 0$ then a rational bubble is said to exist. In the absence of bubbles, where $B_t = 0$, prices will be determined by expected future dividends. However, if a bubble is present, such that $B_t \neq 0$, the explosive behaviour of B_t dominates and prices will be explosive regardless of whether dividends follow a stationary or unit root process. Rational bubbles therefore appear as explosive behaviour in a price series.

In seminal work, Diba and Grossman (1988) propose a method of detecting rational bubbles that examines the first difference of a price series. As shown above, if a bubble exists, a price series will be explosive. The first difference of an explosive series cannot be stationary. They therefore apply unit root tests to the first differenced price series where rejection of the null hypothesis of a unit root implies that a rational bubble is not present in prices. Although this method of bubble detection does not allow us to conclude that an asset bubble exists (nonstationarity in first differenced prices does not imply the presence of explosive behaviour), the opposite conclusion is possible, such that we can rule out the existence of a rational bubble. Applying unit root tests to annual data on the S&P 500 composite stock price index from 1871 to 1986, they find that stock prices are nonstationary in levels but stationary in first differences, and thus conclude that a rational bubble is not present.

Evans (1991) shows that methods which rely on full sample statistical tests are likely to have low power, due to the temporary nature of explosiveness. Typically, asset price bubbles are temporary events, collapsing after a certain period of time. Price series which contain these periodically collapsing bubbles may appear much like unit root or even stationary processes across the full sample. For such processes, methods that examine the stochastic properties of a first differenced price series will have low power. PWY suggest an alternative procedure for detecting explosive behaviour. They examine the price series directly for explosiveness by using the supremum of recursive, right-tailed unit root tests, which consider the null of a unit root against the alternative of explosiveness. Applying this test to monthly observations of the NASDAQ composite stock index with a sample period of February 1973 - June 2005, they find evidence of explosiveness in prices but not in the real dividend series, and thus conclude that a rational bubble occurred in the NASDAQ stock market in this sample period.

2.3 The presence of an initial condition

Consider the following DGP

$$\begin{aligned}y_t &= \mu + u_t \\u_t &= (1 + \delta)u_{t-1} + v_t \quad t = 2, \dots, T\end{aligned}$$

where v_t is assumed to follow a martingale difference sequence with conditional variance σ^2 and $\sup_t E(v_t^4) < \infty$. Suppose that y_t is a price index, which has been normalized at the first value, such that $y_1 = 100$. We can model this initial value in two ways. First, we could set $\mu = 100$ and $u_1 = 0$, such that y_t has a non-zero mean. Alternatively, we could set $\mu = 0$ and $u_1 = 100$, such that y_t has zero mean, but the process u_t is initialised at a non-zero value. Of course, some combination of non-zero μ and u_1 is also possible, such that $\mu + u_1 = 100$. If y_t is a stationary series ($\delta < 0$), the process is mean-reverting and there is a clear sense of the value of the underlying mean of the process. However, in an explosive context ($\delta > 0$), the distinction between the underlying mean and the value of u_1 is less clear cut. In practice, it may be difficult to determine which of these models is appropriate for an explosive process with non-zero initial value.

The initial condition of a process is defined as the deviation of the first observation of the series from the deterministic of that series. Setting $u_1 \neq 0$ implies that the process has a non-zero initial condition. For an explosive process with a non-zero initial value, it is therefore unclear whether the process has an initial condition, or indeed what size such an initial condition would be.

The effect of the initial condition on the performance of unit root tests is well-documented in both the stationary and explosive contexts. When examining the null hypothesis of a unit root against an alternative of stationarity, Elliott et al. (1996) show that a GLS version of the DF test has superior power to its OLS counterpart when the initial condition is asymptotically negligible. However, as demonstrated by Muller and Elliott (2003), this power advantage does not hold when the initial condition is non-negligible, such that, for large values of the initial condition, the OLS DF test has superior power to the GLS variant. Harvey and Leybourne (2014) examine whether the results found in the stationary case extend to the explosive, right-tailed context. Focusing on full-sample explosiveness, they find that GLS DF unit root tests retain their superior power over OLS DF tests in the case of negligible initial conditions, although this power advantage is smaller than that exhibited in the left-tailed context. DF GLS tests are shown to also display superior power to OLS DF tests for non-negligible initial conditions, in contrast to the results seen for left-tailed unit root tests.

The difficulty in distinguishing between non-zero means and non-zero initial conditions in an explosive context motivates us to examine the effect that the presence of an asymptotically non-negligible initial condition has on the PWY test for detecting temporary explosiveness. Given that the PWY test is comprised of the supremum of a series of right-tailed, sub-sample OLS DF tests, we consider whether the full-sample results of Harvey and Leybourne (2014) extend to the case of temporary explosiveness. Specifically, we examine the performance of OLS and GLS versions of the sub-sampling PWY test under a temporary explosive alternative.

2.4 Right-tailed unit root tests

Consider a time series y_t where $t = 1, \dots, T$. We are interested in testing the null that y_t follows a unit root AR(1) process for the full sample, against the alternative that y_t is temporarily explosive, i.e. y_t follows an explosive AR(1) process for some sub-period of the full sample. In order to detect this temporary explosiveness, PWY propose a supremum test based on forward recursions of right-tailed, OLS DF unit root tests. Choosing some initial value τ_0 where $\tau_0 \in [0, 1]$, a right-tailed, OLS DF test is calculated over the first $\lfloor \tau_0 T \rfloor$ observations. The sample size is then increased by an additional observation, the DF test re-estimated, and so on for $T - \lfloor \tau_0 T \rfloor$ recursions. Whilst PWY consider the mean case only, we examine both demeaned, and demeaned and detrended DF tests.

The PWY-type statistics are therefore given by

$$\begin{aligned} PWY_{OLS}^{\mu} &= \sup_{\tau \in [\tau_0, 1]} DF_{OLS}^{\tau, \mu} \\ PWY_{OLS}^{\beta} &= \sup_{\tau \in [\tau_0, 1]} DF_{OLS}^{\tau, \beta} \end{aligned}$$

where $DF_{OLS}^{\tau, \mu}$ and $DF_{OLS}^{\tau, \beta}$ denote the right-tailed demeaned, and demeaned and detrended OLS DF tests respectively. The DF test is the standard t-test

$$DF_{OLS}^{\tau, i} = \frac{\hat{\delta}_{\tau}}{s.e.(\hat{\delta}_{\tau})}$$

where $i = \{\mu, \beta\}$ and $\hat{\delta}_{\tau}$ is the OLS estimate from the auxiliary regression

$$\Delta \hat{u}_{\tau, t} = \delta_{\tau} \hat{u}_{\tau, t-1} + v_{\tau, t} \quad (2.1)$$

where $\hat{u}_{\tau, t} = y_t - z_t' \hat{\theta}$ are the residuals from the OLS regression of y_t on $z_t = 1$,

$\theta = \mu$ in the case of $DF_{OLS}^{\tau,\mu}$ and $z_t = (1, t)'$, $\theta = (\mu, \beta)'$ in the case of $DF_{OLS}^{\tau,\beta}$ over the subsample period $t = 1, \dots, \lfloor \tau T \rfloor$, with the standard error of $\hat{\delta}_\tau$ given by $s.e.(\hat{\delta}_\tau)$.

To explore whether a GLS variant of the PWY test can provide power gains in testing for temporary explosiveness, we modify the PWY test statistic such that it is based on forward recursions of right-tailed, GLS demeaned, or demeaned and detrended DF tests. These tests are given by

$$\begin{aligned} PWY_{GLS}^\mu &= \sup_{\tau \in [\tau_0, 1]} DF_{GLS}^{\tau,\mu} \\ PWY_{GLS}^\beta &= \sup_{\tau \in [\tau_0, 1]} DF_{GLS}^{\tau,\beta} \end{aligned}$$

where $DF_{GLS}^{\tau,\mu}$ and $DF_{GLS}^{\tau,\beta}$ denote the right-tailed, GLS demeaned, and demeaned and detrended DF tests respectively. The DF test is the standard t-test

$$DF_{GLS}^{\tau,i} = \frac{\hat{\delta}_\tau}{s.e.(\hat{\delta}_\tau)}$$

where $i = \{\mu, \beta\}$ and $\hat{\delta}_\tau$ is the GLS estimate from the auxiliary regression

$$\Delta \tilde{u}_{\tau,t} = \tilde{\delta}_\tau \tilde{u}_{\tau,t-1} + \tilde{v}_{\tau,t}, \quad (2.2)$$

where on setting $\bar{\rho} = 1 + \bar{c}/T$ for some chosen constant $\bar{c}$, $\tilde{u}_{\tau,t} = y_t - z_t' \tilde{\theta}$ where $\tilde{\theta}$ is obtained from the GLS regression of $y_c = (y_1, y_2 - \bar{\rho}y_1, \dots, y_{\tau T} - \bar{\rho}y_{\tau T-1})'$ on $z_c = (z_1, z_2 - \bar{\rho}z_1, \dots, z_{\tau T} - \bar{\rho}z_{\tau T-1})'$ where $z_t = 1$ in the case of $DF_{GLS}^{\tau,\mu}$ and $z_t = (1, t)'$ in the case of $DF_{GLS}^{\tau,\beta}$, with the standard error of $\hat{\delta}_\tau$ given by $s.e.(\hat{\delta}_\tau)$.

2.5 Asymptotic behaviour of tests

To examine the behaviour of the four tests discussed in 2.4 we consider the following DGP

$$y_t = \mu + \beta t + u_t \quad (2.3)$$

$$u_t = \begin{cases} T^{1/2}\sigma\alpha & t = 1 \\ u_{t-1} + v_t & t = 2, \dots, \lfloor \tau_1 T \rfloor \\ (1 + \delta)u_{t-1} + v_t & t = \lfloor \tau_1 T \rfloor + 1, \dots, \lfloor \tau_2 T \rfloor \\ u_{t-1} + v_t & t = \lfloor \tau_2 T \rfloor + 1, \dots, T \end{cases} \quad (2.4)$$

where v_t is assumed to follow a martingale difference sequence with conditional variance σ^2 and $\sup_t E(v_t^4) < \infty$. A unit root is imposed on y_t up to time $\lfloor \tau_1 T \rfloor$. We set $\delta \geq 0$ such that y_t follows an explosive process when $\delta > 0$ between time $\lfloor \tau_1 T \rfloor + 1$ and $\lfloor \tau_2 T \rfloor$. In the third regime, the series reverts back to a unit root process. We assume that on reversion back to a unit root there is no crash, such that the observation $u_{\lfloor \tau_2 T \rfloor + 1} = u_{\lfloor \tau_2 T \rfloor} + v_{\lfloor \tau_2 T \rfloor + 1}$. As Evans (1991) argues, this assumption may be unrealistic in that an empirically plausible rational bubble must have a significant chance of collapsing. Periodically collapsing bubbles of this type therefore follow more complex AR(1) processes with additional regimes. One simple solution is to model an instantaneous crash by re-initialising u_t when it reverts back to a unit root. PWY discuss doing this by setting the first observation of the new unit root regime to be equal to the last observation before the explosive regime began plus some stochastic element, such that $u_{\lfloor \tau_2 T \rfloor + 1} = u_{\lfloor \tau_1 T \rfloor} + v_{\lfloor \tau_2 T \rfloor + 1}$. However, as Harvey et al. (2015) note, the supremum of forward recursions of DF unit root tests will tend to occur when the sub-sample of observations used contains only the pre-collapse period of the data. As a result, PWY-type tests are unlikely to be affected by the inclusion of an instantaneous crash in the DGP.

The null hypothesis that y_t follows a unit root throughout is given by $H_0 : \delta = 0$, and the alternative that y_t exhibits temporary explosiveness is given by $H_1 : \delta > 0$. We focus on local alternative hypotheses of the form $\delta = c/T$ where $c \geq 0$. Following Harvey and Leybourne (2014), we consider two possibilities for the initial condition, u_1 : it is either asymptotically negligible, $u_1 = o_p(T^{1/2})$ (for simplicity we set $\alpha = 0$ in this case), or asymptotically non-negligible, where $\alpha \neq 0$ is a finite constant. Under the null hypothesis, α acts as a mean shift, such that increasing α or increasing μ both increase the mean of y_t . Under the alternative, when $\delta > 0$, the initial condition adds an explosive series to the data. In this sense, α and the explosive parameter c both contribute to the magnitude of the explosive process. Note that to illustrate the results, our analysis assumes that v_t is not serially correlated. Following PWY, serial correlation is permitted to enter the model provided that the usual ADF lag augmentation is applied to the DF tests by including lags of $\Delta \hat{u}_{\tau,t}$ and $\Delta \tilde{u}_{\tau,t}$ in the auxiliary regressions 2.1 and 2.2 respectively. The asymptotic distributions of PWY_{OLS}^i and PWY_{GLS}^i where $i = \{\mu, \beta\}$ are given in the following lemma.

Lemma 2.1 *Let y_t be generated according to (2.3)-(2.4) where we assume $\beta = 0$ in the case of the two demeaned tests. For $c \geq 0$,*

$$\begin{aligned}
PWY_{OLS}^\mu &\xrightarrow{d} \sup_{\tau \in [\tau_0, 1]} \frac{K_{c,\alpha}^\mu(\tau, \tau, \tau_1, \tau_2)^2 - K_{c,\alpha}^\mu(0, \tau, \tau_1, \tau_2)^2 - \tau}{2\sqrt{\int_0^\tau K_{c,\alpha}^\mu(r, \tau, \tau_1, \tau_2)^2 dr}} \equiv L_{c,\tau_1,\tau_2}^{OLS^\mu} \\
PWY_{GLS}^\mu &\xrightarrow{d} \sup_{\tau \in [\tau_0, 1]} \frac{K_{c,\alpha}^{\mu,G}(\tau, \tau_1, \tau_2)^2 - \tau}{2\sqrt{\int_0^\tau K_{c,\alpha}^{\mu,G}(r, \tau_1, \tau_2)^2 dr}} \equiv L_{c,\tau_1,\tau_2}^{GLS^\mu} \\
PWY_{OLS}^\beta &\xrightarrow{d} \sup_{\tau \in [\tau_0, 1]} \frac{K_{c,\alpha}^\beta(\tau, \tau, \tau_1, \tau_2)^2 - K_{c,\alpha}^\beta(0, \tau, \tau_1, \tau_2)^2 - \tau}{2\sqrt{\int_0^\tau K_{c,\alpha}^\beta(r, \tau, \tau_1, \tau_2)^2 dr}} \equiv L_{c,\tau_1,\tau_2}^{OLS^\beta} \\
PWY_{GLS}^\beta &\xrightarrow{d} \sup_{\tau \in [\tau_0, 1]} \frac{K_{c,\tilde{c},\alpha}^{\beta,G}(\tau, \tau, \tau_1, \tau_2)^2 - \tau}{2\sqrt{\int_0^\tau K_{c,\tilde{c},\alpha}^{\beta,G}(r, \tau, \tau_1, \tau_2)^2 dr}} \equiv L_{c,\tau_1,\tau_2}^{GLS^\beta}
\end{aligned}$$

where

$$\begin{aligned}
K_{c,\alpha}^\mu(r, \tau, \tau_1, \tau_2) &= K_{c,\alpha}(r, \tau_1, \tau_2) - \tau^{-1} \int_0^\tau K_{c,\alpha}(s, \tau_1, \tau_2) ds \\
K_{c,\alpha}^{\mu,G}(r, \tau_1, \tau_2) &= K_{c,\alpha}(r, \tau_1, \tau_2) - \alpha \\
K_{c,\alpha}^\beta(r, \tau, \tau_1, \tau_2) &= K_{c,\alpha}(r, \tau_1, \tau_2) - \frac{2}{\tau} \left(2 - \frac{3}{\tau}r\right) \int_0^\tau K_{c,\alpha}(s, \tau_1, \tau_2) ds \\
&\quad + \frac{6}{\tau^2} \left(1 - \frac{2}{\tau}r\right) \int_0^\tau s K_{c,\alpha}(s, \tau_1, \tau_2) ds \\
K_{c,\tilde{c},\alpha}^{\beta,G}(r, \tau, \tau_1, \tau_2) &= K_{c,\alpha}(r, \tau_1, \tau_2) - \alpha - \\
&\quad \left[(\tau - \bar{c}\tau^2 + \bar{c}^2\tau^3/3)^{-1} \{ (1 - \bar{c}\tau) K_{c,\alpha}(\tau, \tau_1, \tau_2) \right. \\
&\quad \left. + \bar{c}^2 \int_0^\tau s K_{c,\alpha}(s, \tau_1, \tau_2) ds \} - \frac{1 - \bar{c}\tau + \bar{c}^2\tau^2/2}{\tau - \bar{c}\tau^2 + \bar{c}^2\tau^3/3} \alpha \right] r
\end{aligned}$$

with

$$K_{c,\alpha}(r, \tau_1, \tau_2) = \begin{cases} \alpha & r = 0 \\ \alpha + W(r) & r < \tau_1 \\ e^{c(r-\tau_1)}\alpha + e^{c(r-\tau_1)}W(\tau_1) + \int_{\tau_1}^r e^{c(r-s)}dW(s) & \tau_1 < r < \tau_2 \\ W(r) - W(\tau_2) + e^{c(\tau_2-\tau_1)}\alpha \\ + e^{c(\tau_2-\tau_1)}W(\tau_1) + \int_{\tau_1}^{\tau_2} e^{c(\tau_2-s)}dW(s) & r > \tau_2 \end{cases}$$

and $W(r)$ a standard Brownian motion process.

Proof: See Appendix.

The limit distributions of PWY_{OLS}^μ , PWY_{OLS}^β , PWY_{GLS}^μ and PWY_{GLS}^β under the null hypothesis are given by $L_{0,\tau_1,\tau_2}^{OLS^\mu}$, $L_{0,\tau_1,\tau_2}^{OLS^\beta}$, $L_{0,\tau_1,\tau_2}^{GLS^\mu}$ and $L_{0,\tau_1,\tau_2}^{GLS^\beta}$ respectively. These are the limit distributions obtained from Lemma 2.1, with $c = 0$.

Elliott et al. (1996) choose $\bar{c}$ such that when testing the null of a unit root against the alternative of stationarity, the Gaussian point optimal invariant test of $c = 0$ against $c = \bar{c}$, which forms the asymptotic Gaussian local power envelope, has a power of 0.50. Using a nominal 0.05 level test, Harvey and Leybourne (2014) repeat the exercise in the context of testing the unit root null against an explosive alternative, yielding approximate values of $\bar{c}_\mu = 1.6$ and $\bar{c}_\tau = 2.4$ for the demeaned and detrended DF tests respectively. In the context of temporary explosiveness, the optimal choice of $\bar{c}_\mu$ and $\bar{c}_\tau$ will change depending on the bubble start and end points, τ_1 and τ_2 , as well as the length of the bubble, $\tau_2 - \tau_1$. Of course, these bubbles start and end points are likely to be unknown in practice, making it impossible to know the ideal setting of $\bar{c}_\mu$ and $\bar{c}_\tau$. For simplicity, we therefore choose to adopt the full sample values employed by Harvey and Leybourne (2014) in what follows.

The value of τ_0 , the length of the first sub-sample, should be chosen such that it gives a sufficiently large number of observations to ensure that the initial estimation of the sub-sample DF statistic is adequate, but not too large that a temporary explosive period early in the sample is missed. In practice, the optimal choice of τ_0 is likely to be driven by the sample size T , with smaller values of τ_0 being possible for larger sample sizes. Throughout this paper, we set $\tau_0 = 0.1$ so that our work is in line with that of PWY. Asymptotic null critical values, given in Table 2.1, are generated by direct simulation of the limit distributions using *IID* $N(0, 1)$ random variates, with the integrals approximated by normalised sums of 1,000 steps. Simulations are conducted using 10,000 Monte Carlo replications throughout the chapter.

To evaluate the performance of the tests for different bubble lengths, we consider two $[\tau_1, \tau_2]$ pairs: $[0.45, 0.55]$ and $[0.2, 0.8]$. We therefore have a short bubble where 10% of the data is explosive, and a longer bubble where the explosive interval covers 60% of the data. Both of these non-collapsing bubbles are centred within the sample. To consider the effect of the initial condition on the performance of the tests, we set $\alpha = \{0, 2, 10\}$. Figures 2.1 and 2.2 plot local asymptotic power curves of nominal 0.05-level PWY_{OLS}^μ , PWY_{OLS}^β , PWY_{GLS}^μ and PWY_{GLS}^β tests for different values of c , obtained via direct simulation of the limiting distributions above.

Figure 2.1 reports asymptotic power for the short explosive regime, $[\tau_1, \tau_2] = [0.45, 0.55]$. Figures 2.1(a) and 2.1(b) consider power where $\alpha = 0$, with 2.1(a)

considering the two mean-invariant tests and 2.1(b) considering the two trend-invariant tests. Figures 2.1(c) and 2.1(d) report powers in the mean and trend case respectively, where $\alpha = 2$, and Figures 2.1(e) and 2.1(f) examine $\alpha = 10$. Consider first Figure 2.1(a) where the initial condition is asymptotically negligible. Asymptotic power of both PWY_{OLS}^μ and PWY_{GLS}^μ increases as the magnitude of the explosive bubble, c , increases. For small values of c , PWY_{GLS}^μ has slightly higher power than PWY_{OLS}^μ , but this power ranking is reversed at approximately $c = 14$, beyond which PWY_{OLS}^μ retains a small power advantage. The difference in power between the two tests is always very small, with a maximum disparity of 0.033 observed at $c = 16.8$. In Figure 2.1(b), PWY_{OLS}^β has a power advantage over PWY_{GLS}^β across all values of c . These power differences are larger than those observed in the mean case. For $c = 11.2$, PWY_{OLS}^β has power of 0.423 and PWY_{GLS}^β of 0.292 yielding a power advantage of 0.131.

Consider now Figures 2.1(c) and 2.1(d), where $\alpha = 2$ such that the initial condition is asymptotically non-negligible. In both the mean and trend cases, with $i = \{\mu, \beta\}$, PWY_{OLS}^i has superior power to PWY_{GLS}^i across almost all values of c . The power difference between the two tests is greater than that exhibited when $\alpha = 0$, with PWY_{OLS}^μ and PWY_{OLS}^β having a maximum power advantage of 0.105 and 0.231 respectively. Figures 2.1(e) and 2.1(f) examine power for $\alpha = 10$. The relative power advantage of PWY_{OLS}^i over PWY_{GLS}^i has increased from the $\alpha = 2$ case, with observed maximum power differences of 0.167 and 0.318 respectively. These results demonstrate that when a short explosive period is considered, PWY_{OLS}^i generally has superior power performance to PWY_{GLS}^i for both asymptotically negligible and non-negligible initial conditions. This is contrary to the full sample results of Harvey and Leybourne (2014), where PWY_{GLS}^i was shown to outperform PWY_{OLS}^i at all values of the initial condition. The relative power advantage of PWY_{OLS}^i increases as the value of the initial condition increases. This difference in powers is greater when comparing the two trend-invariant tests than in the mean-invariant case.

Figure 2.2 reports power results where the longer explosive regime, $[\tau_1, \tau_2] = [0.2, 0.8]$, is considered. Figures 2.2(a) and 2.2(b) report powers for the mean-invariant and trend-invariant tests respectively, where $\alpha = 0$. Figures 2.2(c) and 2.2(d) consider $\alpha = 2$ and 2.2(e) and 2.2(f) consider $\alpha = 10$. Consider first Figures 2.2(a) and 2.2(b). As before, the asymptotic local power of all tests is increasing in c . However, for this longer explosive regime, PWY_{GLS}^μ has a small power advantage compared to PWY_{OLS}^μ for c values up to approximately 4.5, beyond which the powers of the two tests are near identical. The greatest power difference observed between the two tests is 0.115 at $c = 2$. In the trend case,

PWY_{GLS}^β outperforms PWY_{OLS}^β at almost all settings of c , with near identical power observed for the few remaining c settings. As with the short explosive period considered in Figure 2.1, the power difference between the two tests is greater when considering the trend-invariant case, with PWY_{GLS}^β having a maximum power advantage of 0.182 at $c = 3.5$.

Consider now Figures 2.2(c) and 2.2(d), where $\alpha = 2$. The relative power performance of the mean-invariant tests closely corresponds to that exhibited in Figure 2.2(a). In the trend case, PWY_{GLS}^β retains a power advantage over PWY_{OLS}^β across all values of c beyond 0.6, with a maximum power difference of 0.528 observed at $c = 2$, much larger than that obtained for $\alpha = 0$. In Figure 2.2(e) and 2.2(f), we consider $\alpha = 10$. PWY_{GLS}^μ has a power advantage over PWY_{OLS}^μ across all c values. The power difference between tests is similar to that observed for $\alpha = 2$, with powers of 0.739 and 0.630 respectively for PWY_{GLS}^μ and PWY_{OLS}^μ at $c = 0.4$, yielding a maximum power advantage of 0.109. In the trend case PWY_{GLS}^β has a power advantage over PWY_{OLS}^β across the vast majority of c values. The maximum power advantage observed under these settings is 0.489 at $c = 0.9$, slightly lower than that exhibited at $\alpha = 2$. We therefore find that, for a longer explosive period, PWY_{GLS}^i generally outperforms PWY_{OLS}^i irrespective of the value of the initial condition. These results coincide with those found in the full sample case by Harvey and Leybourne (2014). We note that the initial condition plays a role in the relative superiority of PWY_{GLS}^β , with a much larger power advantage observed in the case of $\alpha = 2$ and $\alpha = 10$, than for an asymptotically negligible initial condition. However, unlike in the case of a short explosive period, the power differences of the two mean-invariant tests appear to be somewhat less affected by the size of the initial condition.

To further investigate the impact that the initial condition has on asymptotic power, we examine the performance of PWY_{OLS}^i and PWY_{GLS}^i for a given c value across different values of α . Figures 2.3 and 2.4 display power for the two $[\tau_1, \tau_2]$ pairs: $[0.45, 0.55]$ and $[0.2, 0.8]$ respectively. We set $\alpha = \{1, \dots, 10\}$ and we select a value of c that yields power approaching one for the largest value of α . Consider first Figure 2.3, where $c = 2$. In both the mean and trend case, the powers of PWY_{OLS}^i and PWY_{GLS}^i are increasing as α increases. However, beyond approximately $\alpha = 2.5$ in the mean case and $\alpha = 1$ in the trend case, the power of the two tests begin to grow at different rates, such that the relative power advantage of PWY_{OLS}^i to PWY_{GLS}^i is increasing with α . Now considering Figure 2.4, we set $c = 0.8$ in Figure 2.4(a) and $c = 1$ in Figure 2.4(b). Again, in both the mean and trend case, the power of both tests is increasing as α increases. The relative power ranking of the mean-invariant tests is generally unaffected by

the size of the initial condition. However, in the trend case, we observe that the power advantage of PWY_{GLS}^β over PWY_{OLS}^β increases substantially with α . This is in contrast to the full sample, left-tailed results of Muller and Elliott (2003) where a GLS-based unit root test is shown to have decreasing power as the size of the initial condition increases.

Overall, the results in this section highlight the important role that the length of the explosive period plays in the performance of recursive, right-tailed unit root tests. Whilst for longer explosive periods PWY_{GLS}^i has superior power performance to PWY_{OLS}^i , this power ranking is reversed for short explosive periods. With the exception of mean-invariant tests under the longer explosive period, the power difference between tests is shown to be small when the initial condition is asymptotically negligible, but much larger under the presence of an asymptotically non-negligible initial condition. As both the length of the explosive period and the value of the initial condition are unlikely to be known in practice, a researcher will be unsure which test procedure to employ. This suggests that a composite procedure which capitalises on the respective power advantages of both tests is required.

2.6 A union of rejections strategy

We consider the union of rejections procedure of Harvey et al. (2009, 2012) to combine inference from the two mean-invariant tests, or the two trend-invariant tests. The union of rejections approach is a simple decision rule where the null hypothesis of a unit root is rejected if either of the individual tests reject. We can write our proposed union of rejections strategies as

$$\begin{aligned} U^\mu &: \text{Reject } H_0 \text{ if } PWY_{OLS}^\mu > \lambda_\zeta cv_{OLS}^{\zeta,\mu} \text{ or } PWY_{GLS}^\mu > \lambda_\zeta cv_{GLS}^{\zeta,\mu} \\ U^\beta &: \text{Reject } H_0 \text{ if } PWY_{OLS}^\beta > \lambda_\zeta cv_{OLS}^{\zeta,\beta} \text{ or } PWY_{GLS}^\beta > \lambda_\zeta cv_{GLS}^{\zeta,\beta} \end{aligned}$$

where $cv_{OLS}^{\zeta,\mu}$, $cv_{GLS}^{\zeta,\mu}$, $cv_{OLS}^{\zeta,\beta}$ and $cv_{GLS}^{\zeta,\beta}$ denote the asymptotic null critical values of PWY_{OLS}^μ , PWY_{GLS}^μ , PWY_{OLS}^β and PWY_{GLS}^β respectively for a significance level ζ . If we were to use the decision rule

$$\text{Reject } H_0 \text{ if } PWY_{OLS}^i > cv_{OLS}^{\zeta,i} \text{ or } PWY_{GLS}^i > cv_{GLS}^{\zeta,i}$$

such that both test statistics are compared to the critical values given in Table 2.1, U^i would be oversized. We therefore incorporate a scaling constant, λ_ζ , which is calculated such that the asymptotic size of U^μ and U^β is equivalent to

the nominal size ζ . The decision rules can also be written as:

$$U^\mu : \text{Reject } H_0 \text{ if } \max \left(PWY_{OLS}^\mu, \frac{cv_{OLS}^{\zeta,\mu}}{cv_{GLS}^{\zeta,\mu}} PWY_{GLS}^\mu \right) > \lambda_\zeta cv_{OLS}^{\zeta,\mu}$$

$$U^\beta : \text{Reject } H_0 \text{ if } \max \left(PWY_{OLS}^\beta, \frac{cv_{OLS}^{\zeta,\beta}}{cv_{GLS}^{\zeta,\beta}} PWY_{GLS}^\beta \right) > \lambda_\zeta cv_{OLS}^{\zeta,\beta}$$

Under the null hypothesis, where $\delta = 0$,

$$\max \left(PWY_{OLS}^\mu, \frac{cv_{OLS}^{\zeta,\mu}}{cv_{GLS}^{\zeta,\mu}} PWY_{GLS}^\mu \right) \xrightarrow{d} \max \left(L_{c,\tau_1,\tau_2}^{OLS^\mu}, \frac{cv_{OLS}^{\zeta,\mu}}{cv_{GLS}^{\zeta,\mu}} L_{c,\tau_1,\tau_2}^{GLS^\mu} \right) \quad (2.5)$$

$$\max \left(PWY_{OLS}^\beta, \frac{cv_{OLS}^{\zeta,\beta}}{cv_{GLS}^{\zeta,\beta}} PWY_{GLS}^\beta \right) \xrightarrow{d} \max \left(L_{c,\tau_1,\tau_2}^{OLS^\beta}, \frac{cv_{OLS}^{\zeta,\beta}}{cv_{GLS}^{\zeta,\beta}} L_{c,\tau_1,\tau_2}^{GLS^\beta} \right) \quad (2.6)$$

To obtain the appropriate value for the scaling constant λ_ζ , we can simulate the limit distribution of U^i (i.e. the RHS of equations 2.5 and 2.6) and calculate the ζ level critical value $cv_U^{\zeta,\mu}$ or $cv_U^{\zeta,\beta}$. Computing $\lambda_\zeta = cv_U^{\zeta,i} / cv_{OLS}^{\zeta,i}$ will then give the value for the scaling constant that ensures U^i is asymptotically correctly sized. Asymptotic scaling constants calculated in this way are given in Table 2.1.

Figures 2.1 and 2.2 display the powers of U^μ and U^β at the nominal 0.05 level for the explosive intervals $[\tau_1, \tau_2] = [0.45, 0.55]$ and $[0.2, 0.8]$ respectively for initial conditions $\alpha = \{0, 2, 10\}$. Consider first Figures 2.1(a) and 2.1(b). In both the mean and trend case, the power of U^i is near identical to that exhibited by PWY_{OLS}^i , the better performing of the two individual test procedures. This pattern extends to Figures 2.1(c)-2.1(f), where asymptotically non-negligible initial conditions are considered. Therefore, when a short explosive period is present in the data, applying the union procedure is costless in the sense that there is no power loss associated with using U^i over PWY_{OLS}^i , but a power gain is made by using U^i instead of PWY_{GLS}^i . Figure 2.2 considers the longer explosive period where we previously showed that PWY_{GLS}^i generally outperforms PWY_{OLS}^i . Figures 2.2(a) and 2.2(b) display the power of U^i when the initial condition is asymptotically negligible. In both the mean and trend cases, the power of U^i is either very similar, or tracking slightly below that of PWY_{GLS}^i . U^i outperforms PWY_{OLS}^i (or has near identical power) at all values of c . When considering a non-negligible initial condition, as in Figures 2.2(c)-2.2(f), U^i retains its power advantage over PWY_{OLS}^i , exhibiting power either equivalent to or slightly lower than that obtained by PWY_{GLS}^i .

Figures 2.3 and 2.4 display the power of U^i for explosive intervals of $[0.45, 0.55]$ and $[0.2, 0.8]$ respectively across different values of α . In all cases, we see that the

power of U^i is either equivalent to, or slightly below that of the best-performing individual test, whilst always greater than that of the worst-performing test.

The results from this section demonstrate the obvious advantage of employing a union of rejection strategy when testing for temporary explosive behaviour. When the powers of PWY_{OLS}^i and PWY_{GLS}^i are similar, U^i will have near identical power to the better-performing of the two tests, regardless of the length of the explosive period. As the size of the initial condition increases, such that the power differences between the two individual tests also increases, the power of U^i always closely tracks that of the better-performing test, providing a substantial power advantage over either PWY_{GLS}^i when a short explosive interval is considered, or PWY_{OLS}^i when a longer explosive interval is considered. A practitioner employing a union of rejections procedure would overcome any potential power loss arising from uncertainty over both the length of the explosive period and the size of the initial condition.

Throughout this chapter we abstract from the issue of whether or not a trend is present in the time series and assume that the practitioner is able to make an informed judgement about this. Of course, all four individual tests could be combined in a union procedure in line with Harvey et al. (2012), who demonstrate the need for a higher scale value when combining an increasing number of tests, and the subsequent impact on power that this has.

2.7 Finite sample power comparison

To assess the extent to which local asymptotic power is an accurate representation of finite sample behaviour, we consider a number of finite sample simulations. Finite sample critical values for conventional levels of significance and a sample size of $T = 150$ are given in Table 2.2. As in the asymptotic simulations, we set $\alpha = \{0, 2, 10\}$. Finite sample results for $T = 150$ are given in Figures 2.5 and 2.6 for the $[\tau_1, \tau_2]$ pairs: $[0.45, 0.55]$ and $[0.2, 0.8]$ respectively. For both the short and long explosive period, the finite sample powers of PWY_{OLS}^i and PWY_{GLS}^i are closely aligned with their local asymptotic counterparts. As before, the power of U^i is always greater than that of the worst-performing individual test (PWY_{GLS}^i in the $[0.45, 0.55]$ case and PWY_{OLS}^i in the $[0.2, 0.8]$ case), whilst exhibiting power that closely tracks that obtained by the better-performing test. In Figure 2.6(f), the power of U^β tracks further underneath that of PWY_{GLS}^β than was observed for equivalent settings in the asymptotic simulations. Note that whilst we use finite sample critical values, in line with PWY, we continue to use asymptotic union scale values as is standard in a union of rejections context. A practitioner

could choose to simulate appropriate finite sample scale values if the minor loss of power compared to the asymptotic case is of concern.

To assess whether the location of the explosive period affects the power performance of the test statistics, we also consider two additional $[\tau_1, \tau_2]$ pairs: $[0.15, 0.25]$ and $[0.75, 0.85]$. We therefore have two bubbles of short length, located near the beginning and end of the sample respectively. Figures 2.7 and 2.8 display the power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for both asymptotically negligible and non-negligible initial conditions, setting $\alpha = \{0, 2, 10\}$ as before.

For these short explosive intervals we again find that PWY_{OLS}^i is generally superior to PWY_{GLS}^i . We note an exception to this in Figures 2.7(d) and 2.7(f) where the two trend-invariant tests are considered for the interval $[0.15, 0.25]$ with $\alpha = 2$ and $\alpha = 10$ respectively. Here, despite the small interval length, there is a power ranking reversal as c increases such that PWY_{GLS}^i has a small power advantage over PWY_{OLS}^i across most values of c . This result does not occur when the same explosive interval length is placed later in the sample at $[0.75, 0.85]$ in Figures 2.8(d) and 2.8(f), suggesting that the relative power of PWY_{OLS}^i to PWY_{GLS}^i is adversely affected if the explosive period is early in the sample. The power performance of U^i is unaffected by the location of the explosive interval within the sample. As such, U^i offers a method for detecting explosive processes that is robust to the length of the explosive period, the point in the sample at which the explosive period occurs, and the value of the initial condition.

2.8 Empirical application

To illustrate the differing performances of the test statistics considered in this chapter, we apply the four individual tests PWY_{OLS}^μ , PWY_{GLS}^μ , PWY_{OLS}^β and PWY_{GLS}^β , as well as the two union procedures U^μ and U^β to logarithms of daily NASDAQ closing prices. The sample period used is 2 January 1996 - 31 December 2015, yielding 5036 observations (note that there are approximately 252 trading days in each year). We also consider a sub-sample of 2519 observations that ends ten years earlier on 30 December 2005. These two samples are plotted in Figure 2.9. Both samples cover the period of the Dot-Com bubble of the late 1990s. PWY employ PWY_{OLS}^μ on the NASDAQ stock index using monthly data from February 1973 - June 2005 and conclude that the series exhibits explosive behaviour.

PWY_{OLS}^i and PWY_{GLS}^i tests, where $i = \{\mu, \beta\}$, are computed using augmented DF tests where the optimal number of lags is selected using the Bayesian Information Criterion up to a maximum lag of 14. As before, we set $\tau_0 = 0.1$. Ta-

ble 2.3 reports the four individual test statistics and the rejections obtained from the two union procedures. For the full sample, ending in 2015, PWY_{OLS}^μ fails to reject the unit root null at conventional levels of significance, whilst PWY_{GLS}^μ rejects the null at a 0.01 level. The mean-invariant union, U^μ , is able to pick up this rejection at a 0.05 level. Considering the trend-invariant tests, it is the GLS-based procedure, PWY_{GLS}^β , that fails to reject the null hypothesis of a unit root. In contrast, PWY_{OLS}^β rejects the null at a 0.01 level of significance. The trend-invariant union procedure, U^β , picks up this rejection, also at a 0.01 level. This demonstrates how, in practice, inference from PWY tests can depend on whether the procedure employs OLS or GLS DF tests. In both the mean and trend case (where a rejection comes respectively from the GLS test only and the OLS test only), the union of rejections procedure is able to reject the null hypothesis.

When considering the smaller sub-sample of data that ends in 2005, the PWY_{OLS}^μ , PWY_{GLS}^μ and PWY_{OLS}^β test statistics remain the same (at 3 decimal places). The supremum of forward recursions of DF tests will occur when the sub-sample of observations used contains the explosive period of the data. The Dot-Com bubble is generally agreed to have collapsed in the early 2000s. Therefore, by ending the sample in 2005, we have excluded only the post-collapse period of the data.

In this smaller sample, PWY_{GLS}^β now rejects the unit root null at a 0.01 level of significance; whilst the results for the other three individual tests are unchanged from the full sample. This supports the conclusion made in Sections 2.5 and 2.7: the relative power performance of PWY_{GLS}^β to PWY_{OLS}^β improves as the bubble length increases as a proportion of the sample. Again, for this shorter sub-sample, rejections are obtained by the two union of rejections procedures.

2.9 Conclusion

This chapter examines the power performance of the recently proposed PWY test for detecting explosive behaviour in comparison to a GLS-based variant. Limit distributions for both mean-invariant and trend-invariant versions of the two tests are derived; and it is shown that these limits are dependent on the value of the initial condition and the start and end date of the explosive period. Asymptotic and finite sample simulations show that the GLS-based test offers superior power when an explosive period of long length is considered, whilst the original OLS PWY test has a power advantage for explosive periods of short length. The power rankings of the two tests are unaffected by the presence of an

asymptotically non-negligible initial condition, but, generally, power differences between the tests increase as the value of the initial condition increases.

In practice, the exact length of an explosive period and the value of the initial condition are unlikely to be known with certainty, thus it will not be obvious which of the two individual tests should be relied upon. To capitalise on the relative power advantages of the two tests, we propose a union of rejections procedure which is shown to provide power close to the effective power envelope provided by the two individual tests, regardless of the length of the explosive period and the value of the initial condition. Further simulations show that the union procedure is robust to the point in the sample at which the explosive period occurs.

Appendix: Proof of Lemma 2.1

By backward substitution in 2.4 we obtain

$$u_t = \begin{cases} T^{1/2}\sigma\alpha & t = 1 \\ T^{1/2}\sigma\alpha + \sum_{i=2}^t v_i & t = 2, \dots, \lfloor \tau_1 T \rfloor \\ (1 + \delta)^{t - \lfloor \tau_1 T \rfloor} \left(T^{1/2}\sigma\alpha + \sum_{i=2}^{\lfloor \tau_1 T \rfloor} v_i \right) & t = \lfloor \tau_1 T \rfloor + 1, \dots, \lfloor \tau_2 T \rfloor \\ + \sum_{i=\lfloor \tau_1 T \rfloor + 1}^t (1 + \delta)^{i - \lfloor \tau_1 T \rfloor - 1} v_{t - (i - \lfloor \tau_1 T \rfloor - 1)} & \\ u_{\lfloor \tau_2 T \rfloor + 1} + \sum_{i=\lfloor \tau_2 T \rfloor + 2}^t v_i & t = \lfloor \tau_2 T \rfloor + 2, \dots, T \end{cases}$$

The third of these parts comes from the backward recursion:

$$\begin{aligned} u_{\lfloor \tau_1 T + 1 \rfloor} &= (1 + \delta)u_{\lfloor \tau_1 T \rfloor} + v_{\lfloor \tau_1 T \rfloor + 1} \\ u_{\lfloor \tau_1 T + 2 \rfloor} &= (1 + \delta)u_{\lfloor \tau_1 T \rfloor + 1} + v_{\lfloor \tau_1 T \rfloor + 2} \\ &= (1 + \delta)^2 u_{\lfloor \tau_1 T \rfloor} + (1 + \delta)v_{\lfloor \tau_1 T \rfloor + 1} + v_{\lfloor \tau_1 T \rfloor + 2} \\ u_{\lfloor \tau_1 T + 3 \rfloor} &= (1 + \delta)u_{\lfloor \tau_1 T \rfloor + 2} + v_{\lfloor \tau_1 T \rfloor + 3} \\ &= (1 + \delta)^3 u_{\lfloor \tau_1 T \rfloor} + (1 + \delta)^2 v_{\lfloor \tau_1 T \rfloor + 1} + (1 + \delta)v_{\lfloor \tau_1 T \rfloor + 2} + v_{\lfloor \tau_1 T \rfloor + 3} \\ &\vdots \\ u_t &= (1 + \delta)^{t - \lfloor \tau_1 T \rfloor} u_{\lfloor \tau_1 T \rfloor} + \sum_{i=\lfloor \tau_1 T \rfloor + 1}^t (1 + \delta)^{t - i} v_i \quad t = \lfloor \tau_1 T \rfloor + 1, \dots, \lfloor \tau_2 T \rfloor \\ &= (1 + \delta)^{t - \lfloor \tau_1 T \rfloor} \left(T^{1/2}\sigma\alpha + \sum_{i=2}^{\lfloor \tau_1 T \rfloor} v_i \right) \\ &\quad + \sum_{i=\lfloor \tau_1 T \rfloor + 1}^t (1 + \delta)^{t - i} v_i \quad t = \lfloor \tau_1 T \rfloor + 1, \dots, \lfloor \tau_2 T \rfloor \end{aligned}$$

and subsequently, using $\delta = c/T$

$$T^{-1/2}u_{\lfloor rT \rfloor} = \begin{cases} \sigma\alpha & \lfloor rT \rfloor = 1 \\ \sigma\alpha + T^{-1/2} \sum_{i=2}^{\lfloor rT \rfloor} v_i & \lfloor rT \rfloor = 2, \dots, \lfloor \tau_1 T \rfloor \\ (1 + c/T)^{\lfloor rT \rfloor - \lfloor \tau_1 T \rfloor} \left(\sigma\alpha + T^{-1/2} \sum_{i=2}^{\lfloor \tau_1 T \rfloor} v_i \right) & \lfloor rT \rfloor = \lfloor \tau_1 T \rfloor + 1, \dots, \lfloor \tau_2 T \rfloor \\ + T^{-1/2} \sum_{i=\lfloor \tau_1 T \rfloor + 1}^{\lfloor rT \rfloor} (1 + c/T)^{\lfloor rT \rfloor - i} v_i & \\ T^{-1/2}u_{\lfloor \tau_2 T \rfloor + 1} + T^{-1/2} \sum_{i=\lfloor \tau_2 T \rfloor + 2}^{\lfloor rT \rfloor} v_i & \lfloor rT \rfloor = \lfloor \tau_2 T \rfloor + 1, \dots, T \end{cases}$$

Weak convergence of standardised partial sums of v_t to the standard Brownian motion process $W(r)$ is a standard result, so we can write

$$\begin{aligned} T^{-1/2} \sum_{i=2}^{\lfloor rT \rfloor} v_i &\xrightarrow{d} \sigma W(r) \\ T^{-1/2} \sum_{i=2}^{\lfloor \tau_1 T \rfloor} v_i &\xrightarrow{d} \sigma W(\tau_1) \end{aligned}$$

$$T^{-1/2} \sum_{i=\lfloor \tau_2 T \rfloor + 2}^{\lfloor rT \rfloor} v_i \xrightarrow{d} \sigma W(r) - \sigma W(\tau_2)$$

Following Phillips (1987), from the third part of the above

$$T^{-1/2} \sum_{i=\lfloor \tau_1 T \rfloor + 1}^{\lfloor rT \rfloor} (1 + c/T)^{\lfloor rT \rfloor - i} v_i$$

converges in distribution to the Ornstein-Uhlenbeck process

$$W_c(r) = \int_{\tau_1}^r e^{c(r-s)} dW(s)$$

We can therefore show that

$$\begin{aligned} T^{-1/2} u_{\lfloor rT \rfloor} &\xrightarrow{d} \sigma \begin{cases} \alpha & r = 0 \\ \alpha + W(r) & r < \tau_1 \\ e^{c(r-\tau_1)}\alpha + e^{c(r-\tau_1)}W(\tau_1) + \int_{\tau_1}^r e^{c(r-s)}dW(s) & \tau_1 < r < \tau_2 \\ W(r) - W(\tau_2) + e^{c(\tau_2-\tau_1)}\alpha + e^{c(\tau_2-\tau_1)}W(\tau_1) & r > \tau_2 \\ + \int_{\tau_1}^{\tau_2} e^{c(\tau_2-s)}dW(s) \end{cases} \\ &\equiv \sigma K_{c,\alpha}(r, \tau_1, \tau_2) \end{aligned}$$

Asymptotic distribution of PWY_{OLS}^μ

Here we have $\beta = 0$ and the recursive test statistic calculated for the sub-sample $t = 1, \dots, \lfloor \tau T \rfloor$ is based on the t -ratio for $\hat{\delta}_\tau$ from the estimated regression

$$\Delta \hat{u}_{\tau,t} = \hat{\delta}_\tau \hat{u}_{\tau,t-1} + \hat{v}_{\tau,t}, \quad t = 1, \dots, \lfloor \tau T \rfloor$$

where

$$\hat{u}_{\tau,t} = y_t - \lfloor \tau T \rfloor^{-1} \sum_{s=1}^{\lfloor \tau T \rfloor} y_s$$

i.e.

$$t_\tau = \frac{\hat{\delta}_\tau}{s.e.(\hat{\delta}_\tau)}$$

where

$$\begin{aligned} \hat{\delta}_\tau &= \frac{\sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \hat{u}_{\tau,t} \hat{u}_{\tau,t-1}}{\sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2} \\ s.e.(\hat{\delta}_\tau)^2 &= \frac{\hat{\sigma}_\tau^2}{\sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2} \\ \hat{\sigma}_\tau^2 &= \lfloor \tau T \rfloor^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{v}_{\tau,t}^2 \end{aligned}$$

Note that in what follows we can set $\mu = 0$ without loss of generality, so that $y_t = u_t$. Consider the recursively demeaned y_t , i.e. $\hat{u}_{\tau,t}$. Since $y_t = u_t$ we have

$$\begin{aligned} \hat{u}_{\tau,t} &= u_t - \lfloor \tau T \rfloor^{-1} \sum_{s=1}^{\lfloor \tau T \rfloor} u_s \\ T^{-1/2} \hat{u}_{\tau, \lfloor \tau T \rfloor} &= T^{-1/2} u_{\lfloor \tau T \rfloor} - \lfloor \tau T \rfloor^{-1} \sum_{\lfloor sT \rfloor=1}^{\lfloor \tau T \rfloor} T^{-1/2} u_{\lfloor sT \rfloor} \\ &\xrightarrow{d} \sigma K_{c,\alpha}(r, \tau_1, \tau_2) - \sigma \tau^{-1} \int_0^\tau K_{c,\alpha}(s, \tau_1, \tau_2) ds \\ &\equiv \sigma K_{c,\alpha}^\mu(r, \tau, \tau_1, \tau_2) \end{aligned}$$

where

$$T^{-3/2} \sum_{\lfloor sT \rfloor=1}^{\lfloor \tau T \rfloor} u_{\lfloor sT \rfloor} \xrightarrow{d} \sigma \int_0^\tau K_{c,\alpha}(s, \tau_1, \tau_2) ds$$

by the Functional Central Limit Theorem.

The recursive parameter estimate $\hat{\delta}_\tau$:

$$\begin{aligned}\hat{\delta}_\tau &= \frac{\sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \hat{u}_{\tau,t} \hat{u}_{\tau,t-1}}{\sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2} \\ T \hat{\delta}_\tau &= \frac{T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \hat{u}_{\tau,t} \hat{u}_{\tau,t-1}}{T^{-2} \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2}\end{aligned}$$

For the denominator we have

$$\begin{aligned}T^{-2} \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2 &= T^{-1} \sum_{\lfloor \tau T \rfloor=2}^{\lfloor \tau T \rfloor} (T^{-1/2} \hat{u}_{\tau, \lfloor \tau T \rfloor - 1})^2 \\ &\xrightarrow{d} \sigma^2 \int_0^\tau K_{c,\alpha}^\mu(r, \tau, \tau_1, \tau_2)^2 dr\end{aligned}$$

Now consider the numerator

$$T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \hat{u}_{\tau,t} \hat{u}_{\tau,t-1} = \frac{1}{2} \left\{ (T^{-1/2} \hat{u}_{\lfloor \tau T \rfloor})^2 - (T^{-1/2} \hat{u}_1)^2 - T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta \hat{u}_{\tau,t})^2 \right\}$$

using

$$\begin{aligned}\hat{u}_{\tau,t} &= \Delta \hat{u}_{\tau,t} + \hat{u}_{\tau,t-1} \\ \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t}^2 &= \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta \hat{u}_{\tau,t})^2 + \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2 + 2 \sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \hat{u}_{\tau,t} \hat{u}_{\tau,t-1} \\ \sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \hat{u}_{\tau,t} \hat{u}_{\tau,t-1} &= \frac{1}{2} \left\{ \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t}^2 - \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2 - \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta \hat{u}_{\tau,t})^2 \right\} \\ &= \frac{1}{2} \left\{ \hat{u}_{\lfloor \tau T \rfloor}^2 - \hat{u}_1^2 - \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta \hat{u}_{\tau,t})^2 \right\}\end{aligned}$$

Then

$$\begin{aligned}T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \hat{u}_{\tau,t} \hat{u}_{\tau,t-1} &= \frac{1}{2} \left\{ (T^{-1/2} \hat{u}_{\lfloor \tau T \rfloor})^2 - (T^{-1/2} \hat{u}_1)^2 - T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta \hat{u}_{\tau,t})^2 \right\} \\ &\xrightarrow{d} \frac{1}{2} \left\{ \sigma^2 K_{c,\alpha}^\mu(\tau, \tau, \tau_1, \tau_2)^2 - \sigma^2 K_{c,\alpha}^\mu(0, \tau, \tau_1, \tau_2)^2 - \tau \sigma^2 \right\}\end{aligned}$$

The limit of the last term comes from

$$\begin{aligned}
T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta \hat{u}_{\tau,t})^2 &= T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\hat{u}_{\tau,t} - \hat{u}_{\tau,t-1})^2 \\
&= T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \left\{ \left(u_t - \lfloor \tau T \rfloor^{-1} \sum_{s=1}^{\lfloor \tau T \rfloor} u_s \right) - \left(u_{t-1} - \lfloor \tau T \rfloor^{-1} \sum_{s=1}^{\lfloor \tau T \rfloor} u_s \right) \right\}^2 \\
&= T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta u_t)^2 \\
&= \tau (\tau T)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta u_t)^2 \\
&= \tau \begin{cases} (\tau T)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} v_t^2 & \lfloor \tau T \rfloor \leq \lfloor \tau_1 T \rfloor \\ (\tau T)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} v_t^2 + o_p(1) & \lfloor \tau_1 T \rfloor < \lfloor \tau T \rfloor \leq \lfloor \tau_2 T \rfloor \\ (\tau T)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} v_t^2 + o_p(1) & \lfloor \tau T \rfloor > \lfloor \tau_2 T \rfloor \end{cases} \\
&\xrightarrow{p} \tau \sigma^2
\end{aligned}$$

since

$$\begin{aligned}
\Delta u_t &= \begin{cases} v_t & t = 2, \dots, \lfloor \tau_1 T \rfloor \\ cT^{-1} u_{t-1} + v_t & t = \lfloor \tau_1 T \rfloor + 1, \dots, \lfloor \tau_2 T \rfloor \\ v_t & t = \lfloor \tau_2 T \rfloor + 1, \dots, T \end{cases} \\
&= \begin{cases} v_t & t = 2, \dots, \lfloor \tau_1 T \rfloor \\ v_t + o_p(1) & t = \lfloor \tau_1 T \rfloor + 1, \dots, \lfloor \tau_2 T \rfloor \\ v_t & t = \lfloor \tau_2 T \rfloor + 1, \dots, T \end{cases}
\end{aligned}$$

So

$$\begin{aligned}
T \hat{\delta}_\tau &\xrightarrow{d} \frac{\frac{1}{2} \{ \sigma^2 K_{c,\alpha}^\mu(\tau, \tau, \tau_1, \tau_2)^2 - \sigma^2 K_{c,\alpha}^\mu(0, \tau, \tau_1, \tau_2)^2 - \tau \sigma^2 \}}{\sigma^2 \int_0^\tau K_{c,\alpha}^\mu(r, \tau, \tau_1, \tau_2)^2 dr} \\
&= \frac{K_{c,\alpha}^\mu(\tau, \tau, \tau_1, \tau_2)^2 - K_{c,\alpha}^\mu(0, \tau, \tau_1, \tau_2)^2 - \tau}{2 \int_0^\tau K_{c,\alpha}^\mu(r, \tau, \tau_1, \tau_2)^2 dr}
\end{aligned}$$

Next consider $\hat{\sigma}_\tau^2$ and then $s.e.(\hat{\delta}_\tau)$. We obtain

$$\begin{aligned}
\hat{\sigma}_\tau^2 &= [\tau T]^{-1} \sum_{t=2}^{[\tau T]} \hat{v}_{\tau,t}^2 \\
&= [\tau T]^{-1} \sum_{t=2}^{[\tau T]} (\Delta \hat{u}_{\tau,t} - \hat{\delta}_\tau \hat{u}_{\tau,t-1})^2 \\
&= [\tau T]^{-1} \sum_{t=2}^{[\tau T]} (\Delta \hat{u}_{\tau,t})^2 + \hat{\delta}_\tau^2 [\tau T]^{-1} \sum_{t=2}^{[\tau T]} \hat{u}_{\tau,t-1}^2 - 2\hat{\delta}_\tau [\tau T]^{-1} \sum_{t=2}^{[\tau T]} \Delta \hat{u}_{\tau,t} \hat{u}_{\tau,t-1} \\
&= [\tau T]^{-1} \sum_{t=2}^{[\tau T]} (\Delta u_t)^2 + O_p(T^{-2})O_p(T) - 2O_p(T^{-1})O_p(1) \\
&= [\tau T]^{-1} \sum_{t=2}^{[\tau T]} (\Delta u_t)^2 + o_p(1) \\
&\xrightarrow{p} \sigma^2
\end{aligned}$$

and

$$\begin{aligned}
\{Ts.e.(\hat{\delta}_\tau)\}^2 &= \frac{\hat{\sigma}_\tau^2}{T^{-2} \sum_{t=2}^{[\tau T]} \hat{u}_{\tau,t-1}^2} \\
&\xrightarrow{d} \frac{\sigma^2}{\sigma^2 \int_0^\tau K_{c,\alpha}^\mu(r, \tau, \tau_1, \tau_2)^2 dr} \\
&= \frac{1}{\int_0^\tau K_{c,\alpha}^\mu(r, \tau, \tau_1, \tau_2)^2 dr}
\end{aligned}$$

Putting it all together we find

$$\begin{aligned}
t_\tau &= \frac{T\hat{\delta}_\tau}{Ts.e.(\hat{\delta}_\tau)} \\
&\xrightarrow{d} \frac{\frac{K_{c,\alpha}^\mu(\tau, \tau, \tau_1, \tau_2)^2 - K_{c,\alpha}^\mu(0, \tau, \tau_1, \tau_2)^2 - \tau}{2 \int_0^\tau K_{c,\alpha}^\mu(r, \tau, \tau_1, \tau_2)^2 dr}}{\sqrt{\frac{1}{\int_0^\tau K_{c,\alpha}^\mu(r, \tau, \tau_1, \tau_2)^2 dr}}} \\
&= \frac{K_{c,\alpha}^\mu(\tau, \tau, \tau_1, \tau_2)^2 - K_{c,\alpha}^\mu(0, \tau, \tau_1, \tau_2)^2 - \tau}{2\sqrt{\int_0^\tau K_{c,\alpha}^\mu(r, \tau, \tau_1, \tau_2)^2 dr}}
\end{aligned}$$

Asymptotic distribution of PWY_{GLS}^μ

Here we have $\beta = 0$ and the recursive test statistic calculated for the sub-sample $t = 1, \dots, \lfloor \tau T \rfloor$ is based on the t -ratio for $\tilde{\delta}_\tau$ from the estimated regression

$$\Delta \tilde{u}_{\tau,t} = \tilde{\delta}_\tau \tilde{u}_{\tau,t-1} + \tilde{v}_{\tau,t}, \quad t = 1, \dots, \lfloor \tau T \rfloor$$

where

$$\tilde{u}_{\tau,t} = y_t - \tilde{\mu}_\tau$$

with $\tilde{\mu}_\tau$ the recursive GLS estimate of the mean with quasi-differencing parameter $\bar{\rho} = 1 + \bar{c}T^{-1}$, i.e.

$$\begin{aligned} \tilde{\mu}_\tau &= \frac{y_1 + (1 - \bar{\rho}) \sum_{t=2}^{\lfloor \tau T \rfloor} (y_t - \bar{\rho} y_{t-1})}{1 + (\lfloor \tau T \rfloor - 1)(1 - \bar{\rho})^2} \\ &= \frac{y_1 - \bar{c}T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta y_t - \bar{c}T^{-1} y_{t-1})}{1 + (\lfloor \tau T \rfloor - 1)\bar{c}^2 T^{-2}} \\ &= y_1 + o_p(1) \end{aligned}$$

so the recursive t -ratio here is

$$t_\tau = \frac{\tilde{\delta}_\tau}{s.e.(\tilde{\delta}_\tau)}$$

where

$$\begin{aligned} \tilde{\delta}_\tau &= \frac{\sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \tilde{u}_{\tau,t} \tilde{u}_{\tau,t-1}}{\sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{u}_{\tau,t-1}^2} \\ s.e.(\tilde{\delta}_\tau)^2 &= \frac{\tilde{\sigma}_\tau^2}{\sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{u}_{\tau,t-1}^2} \\ \tilde{\sigma}_\tau^2 &= \lfloor \tau T \rfloor^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{v}_{\tau,t}^2 \end{aligned}$$

Note that in what follows we can set $\mu = 0$ without loss of generality, so that $y_t = u_t$. First consider the recursively demeaned y_t , i.e. $\tilde{u}_{\tau,t}$. Since $y_t = u_t$ we have

$$\begin{aligned} \tilde{u}_{\tau,t} &= u_t - u_1 + o_p(1) \\ T^{-1/2} \tilde{u}_{\tau, \lfloor rT \rfloor} &= T^{-1/2} (u_{\lfloor rT \rfloor} - u_1) + o_p(1) \\ &\xrightarrow{d} \sigma K_{c,\alpha}(r, \tau_1, \tau_2) - \sigma \alpha \\ &= \sigma K_{c,\alpha}^{\mu,G}(r, \tau_1, \tau_2) \end{aligned}$$

Now consider the recursive parameter estimate $\tilde{\delta}_\tau$:

$$\begin{aligned}\tilde{\delta}_\tau &= \frac{\sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \tilde{u}_{\tau,t} \tilde{u}_{\tau,t-1}}{\sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{u}_{\tau,t-1}^2} \\ T\tilde{\delta}_\tau &= \frac{T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \tilde{u}_{\tau,t} \tilde{u}_{\tau,t-1}}{T^{-2} \sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{u}_{\tau,t-1}^2}\end{aligned}$$

For the denominator we have

$$\begin{aligned}T^{-2} \sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{u}_{\tau,t-1}^2 &= T^{-1} \sum_{\lfloor rT \rfloor=2}^{\lfloor \tau T \rfloor} (T^{-1/2} \tilde{u}_{\tau,t-1})^2 \\ &\xrightarrow{d} \sigma^2 \int_0^\tau K_{c,\alpha}^{\mu,G}(r, \tau_1, \tau_2)^2 dr\end{aligned}$$

by the Continuous Mapping Theorem. Now consider the numerator

$$\begin{aligned}T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \tilde{u}_{\tau,t} \tilde{u}_{\tau,t-1} &= \frac{1}{2} \left\{ (T^{-1/2} \tilde{u}_{\lfloor \tau T \rfloor})^2 - (T^{-1/2} \tilde{u}_1)^2 - T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta \tilde{u}_{\tau,t})^2 \right\} \\ &\xrightarrow{d} \frac{1}{2} \{ \sigma^2 K_{c,\alpha}^{\mu,G}(\tau, \tau_1, \tau_2)^2 - \sigma^2 K_{c,\alpha}^{\mu,G}(0, \tau_1, \tau_2)^2 - \tau \sigma^2 \}\end{aligned}$$

with the limit of the last term coming from

$$\begin{aligned}T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta \tilde{u}_{\tau,t})^2 &= T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\tilde{u}_{\tau,t} - \tilde{u}_{\tau,t-1})^2 \\ &= T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \{ (u_t - u_1) - (u_{t-1} - u_1) \}^2 + o_p(1) \\ &= T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta u_t)^2 + o_p(1) \\ &= \tau(\tau T)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta u_t)^2 + o_p(1) \\ &\xrightarrow{p} \tau \sigma^2\end{aligned}$$

Note that $K_{c,\alpha}^{\mu,G}(0, \tau_1, \tau_2) = W(0) = 0$ so we can simplify to

$$T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \tilde{u}_{\tau,t} \tilde{u}_{\tau,t-1} \xrightarrow{d} \frac{1}{2} \{ \sigma^2 K_{c,\alpha}^{\mu,G}(\tau, \tau_1, \tau_2)^2 - \tau \sigma^2 \}$$

So

$$\begin{aligned}
T\tilde{\delta}_\tau &\xrightarrow{d} \frac{\frac{1}{2}\{\sigma^2 K_{c,\alpha}^{\mu,G}(\tau, \tau_1, \tau_2)^2 - \tau\sigma^2\}}{\sigma^2 \int_0^\tau K_{c,\alpha}^{\mu,G}(r, \tau_1, \tau_2)^2 dr} \\
&= \frac{K_{c,\alpha}^{\mu,G}(\tau, \tau_1, \tau_2)^2 - \tau}{2 \int_0^\tau K_{c,\alpha}^{\mu,G}(r, \tau_1, \tau_2)^2 dr}
\end{aligned}$$

Next consider $\tilde{\sigma}_\tau^2$ and then $s.e.(\tilde{\delta}_\tau)$. We obtain

$$\begin{aligned}
\tilde{\sigma}_\tau^2 &= [\tau T]^{-1} \sum_{t=2}^{[\tau T]} \tilde{v}_{\tau,t}^2 \\
&= [\tau T]^{-1} \sum_{t=2}^{[\tau T]} (\Delta \tilde{u}_{\tau,t} - \tilde{\delta}_\tau \tilde{u}_{\tau,t-1})^2 \\
&= [\tau T]^{-1} \sum_{t=2}^{[\tau T]} (\Delta \tilde{u}_{\tau,t})^2 + \tilde{\delta}_\tau^2 [\tau T]^{-1} \sum_{t=2}^{[\tau T]} \tilde{u}_{\tau,t-1}^2 - 2\tilde{\delta}_\tau [\tau T]^{-1} \sum_{t=2}^{[\tau T]} \Delta \tilde{u}_{\tau,t} \tilde{u}_{\tau,t-1} \\
&= [\tau T]^{-1} \sum_{t=2}^{[\tau T]} (\Delta u_t)^2 + O_p(T^{-2})O_p(T) - 2O_p(T^{-1})O_p(1) \\
&= [\tau T]^{-1} \sum_{t=2}^{[\tau T]} (\Delta u_t)^2 + o_p(1) \\
&\xrightarrow{p} \sigma^2
\end{aligned}$$

and

$$\begin{aligned}
\{T s.e.(\tilde{\delta}_\tau)\}^2 &= \frac{\tilde{\sigma}_\tau^2}{T^{-2} \sum_{t=2}^{[\tau T]} \tilde{u}_{\tau,t-1}^2} \\
&\xrightarrow{d} \frac{\sigma^2}{\sigma^2 \int_0^\tau K_{c,\alpha}^{\mu,G}(r, \tau_1, \tau_2)^2 dr} \\
&= \frac{1}{\int_0^\tau K_{c,\alpha}^{\mu,G}(r, \tau_1, \tau_2)^2 dr}
\end{aligned}$$

Putting it all together we find

$$\begin{aligned}
t_\tau &= \frac{T\tilde{\delta}_\tau}{T s.e.(\tilde{\delta}_\tau)} \\
&\xrightarrow{d} \frac{K_{c,\alpha}^{\mu,G}(\tau, \tau_1, \tau_2)^2 - \tau}{2\sqrt{\int_0^\tau K_{c,\alpha}^{\mu,G}(r, \tau_1, \tau_2)^2 dr}}
\end{aligned}$$

Asymptotic distribution of PWY_{OLS}^β

In this case the recursive test statistic calculated for the sub-sample $t = 1, \dots, \lfloor \tau T \rfloor$ is based on the t -ratio for $\hat{\delta}_\tau$ from the estimated regression

$$\Delta \hat{u}_{\tau,t} = \hat{\delta}_\tau \hat{u}_{\tau,t-1} + \hat{v}_{\tau,t}, \quad t = 1, \dots, \lfloor \tau T \rfloor$$

where

$$\hat{u}_{\tau,t} = y_t - \hat{\mu}_\tau - \hat{\beta}_\tau t$$

with $\hat{\mu}_\tau$ and $\hat{\beta}_\tau$ the recursively detrending estimates. The t -ratio is then

$$t_\tau = \frac{\hat{\delta}_\tau}{s.e.(\hat{\delta}_\tau)}$$

with

$$\begin{aligned} \hat{\delta}_\tau &= \frac{\sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \hat{u}_{\tau,t} \hat{u}_{\tau,t-1}}{\sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2} \\ s.e.(\hat{\delta}_\tau)^2 &= \frac{\hat{\sigma}_\tau^2}{\sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2} \\ \hat{\sigma}_\tau^2 &= \lfloor \tau T \rfloor^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{v}_{\tau,t}^2 \end{aligned}$$

In what follows we can set $\mu = \beta = 0$, so that $y_t = u_t$ throughout. First consider the properties of the recursively detrending estimates:

$$\begin{aligned} \begin{bmatrix} \hat{\mu}_\tau \\ \hat{\beta}_\tau \end{bmatrix} &= \begin{bmatrix} \lfloor \tau T \rfloor & \sum_{t=1}^{\lfloor \tau T \rfloor} t \\ \sum_{t=1}^{\lfloor \tau T \rfloor} t & \sum_{t=1}^{\lfloor \tau T \rfloor} t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{\lfloor \tau T \rfloor} u_t \\ \sum_{t=1}^{\lfloor \tau T \rfloor} t u_t \end{bmatrix} \\ \begin{bmatrix} T^{-1/2} \hat{\mu}_\tau \\ T^{1/2} \hat{\beta}_\tau \end{bmatrix} &= \begin{bmatrix} T^{-1} \lfloor \tau T \rfloor & T^{-2} \sum_{t=1}^{\lfloor \tau T \rfloor} t \\ T^{-2} \sum_{t=1}^{\lfloor \tau T \rfloor} t & T^{-3} \sum_{t=1}^{\lfloor \tau T \rfloor} t^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-3/2} \sum_{t=1}^{\lfloor \tau T \rfloor} u_t \\ T^{-5/2} \sum_{t=1}^{\lfloor \tau T \rfloor} t u_t \end{bmatrix} \\ &\xrightarrow{d} \begin{bmatrix} \tau & \tau^2/2 \\ \tau^2/2 & \tau^3/3 \end{bmatrix}^{-1} \begin{bmatrix} \sigma \int_0^\tau K_{c,\alpha}(r, \tau_1, \tau_2) dr \\ \sigma \int_0^\tau r K_{c,\alpha}(r, \tau_1, \tau_2) dr \end{bmatrix} \\ &= \begin{bmatrix} \frac{4}{\tau} & -\frac{6}{\tau^2} \\ -\frac{6}{\tau^2} & \frac{12}{\tau^3} \end{bmatrix} \begin{bmatrix} \sigma \int_0^\tau K_{c,\alpha}(r, \tau_1, \tau_2) dr \\ \sigma \int_0^\tau r K_{c,\alpha}(r, \tau_1, \tau_2) dr \end{bmatrix} \\ &= \begin{bmatrix} \frac{4\sigma}{\tau} \int_0^\tau K_{c,\alpha}(r, \tau_1, \tau_2) dr - \frac{6\sigma}{\tau^2} \int_0^\tau r K_{c,\alpha}(r, \tau_1, \tau_2) dr \\ \frac{12\sigma}{\tau^3} \int_0^\tau r K_{c,\alpha}(r, \tau_1, \tau_2) dr - \frac{6\sigma}{\tau^2} \int_0^\tau K_{c,\alpha}(r, \tau_1, \tau_2) dr \end{bmatrix} \end{aligned}$$

where

$$T^{-5/2} \sum_{t=1}^{rT} tu_t \xrightarrow{d} \int_0^\tau r K_{c,\alpha}(r, \tau_1, \tau_2) dr$$

is a standard result given in Hamilton (1994).

Next consider the recursively detrended y_t , i.e. $\hat{u}_{\tau,t}$. Since $y_t = u_t$ we have

$$\begin{aligned} \hat{u}_{\tau,t} &= u_t - \hat{\mu}_\tau - \hat{\beta}_\tau t \\ T^{-1/2} \hat{u}_{\tau, \lfloor rT \rfloor} &= T^{-1/2} u_{\lfloor rT \rfloor} - T^{-1/2} \hat{\mu}_\tau - T^{-1/2} \hat{\beta}_\tau t \\ &= T^{-1/2} u_{\lfloor rT \rfloor} - T^{-1/2} \hat{\mu}_\tau - T^{1/2} \hat{\beta}_\tau r \\ &\xrightarrow{d} \sigma K_{c,\alpha}(r, \tau_1, \tau_2) - \frac{4\sigma}{\tau} \int_0^\tau K_{c,\alpha}(s, \tau_1, \tau_2) ds + \frac{6\sigma}{\tau^2} \int_0^\tau s K_{c,\alpha}(s, \tau_1, \tau_2) ds \\ &\quad - \left\{ \frac{12\sigma}{\tau^3} \int_0^\tau s K_{c,\alpha}(s, \tau_1, \tau_2) ds - \frac{6\sigma}{\tau^2} \int_0^\tau K_{c,\alpha}(s, \tau_1, \tau_2) ds \right\} r \\ &= \sigma K_{c,\alpha}(r, \tau_1, \tau_2) - \frac{2\sigma}{\tau} \left(2 - \frac{3}{\tau} r \right) \int_0^\tau K_{c,\alpha}(s, \tau_1, \tau_2) ds \\ &\quad + \frac{6\sigma}{\tau^2} \left(1 - \frac{2}{\tau} r \right) \int_0^\tau s K_{c,\alpha}(s, \tau_1, \tau_2) ds \\ &\equiv \sigma K_{c,\alpha}^\beta(r, \tau, \tau_1, \tau_2) \end{aligned}$$

Now consider the recursive parameter estimate $\hat{\delta}_\tau$:

$$\begin{aligned} \hat{\delta}_\tau &= \frac{\sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \hat{u}_{\tau,t} \hat{u}_{\tau,t-1}}{\sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2} \\ T \hat{\delta}_\tau &= \frac{T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \hat{u}_{\tau,t} \hat{u}_{\tau,t-1}}{T^{-2} \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2} \end{aligned}$$

For the denominator we have

$$\begin{aligned} T^{-2} \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2 &= T^{-1} \sum_{\lfloor rT \rfloor=2}^{\lfloor \tau T \rfloor} (T^{-1/2} \hat{u}_{\tau, \lfloor rT \rfloor - 1})^2 \\ &\xrightarrow{d} \sigma^2 \int_0^\tau K_{c,\alpha}^\beta(r, \tau, \tau_1, \tau_2)^2 dr \end{aligned}$$

Now consider the numerator

$$\begin{aligned} T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \hat{u}_{\tau,t} \hat{u}_{\tau,t-1} &= \frac{1}{2} \left\{ (T^{-1/2} \hat{u}_{\tau, \lfloor \tau T \rfloor})^2 - (T^{-1/2} \hat{u}_{\tau, 1})^2 - T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta \hat{u}_{\tau,t})^2 \right\} \\ &\xrightarrow{d} \frac{1}{2} \left\{ \sigma^2 K_{c,\alpha}^\beta(\tau, \tau, \tau_1, \tau_2)^2 - \sigma^2 K_{c,\alpha}^\beta(0, \tau, \tau_1, \tau_2)^2 - \tau \sigma^2 \right\} \end{aligned}$$

The limit of the last term comes from

$$\begin{aligned}
T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta \hat{u}_{\tau,t})^2 &= T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\hat{u}_{\tau,t} - \hat{u}_{\tau,t-1})^2 \\
&= T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \left\{ \left(u_t - \hat{\mu}_\tau - \hat{\beta}_\tau t \right) - \left(u_{t-1} - \hat{\mu}_\tau - \hat{\beta}_\tau (t-1) \right) \right\}^2 \\
&= T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta u_t - \hat{\beta}_\tau)^2 \\
&= \tau(\tau T)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta u_t)^2 + T^{-1} \lfloor \tau T \rfloor \hat{\beta}_\tau^2 - 2\hat{\beta}_\tau T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \Delta u_t \\
&= \tau(\tau T)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta u_t)^2 + O_p(T^{-1}) - 2O_p(T^{-1/2})O_p(T^{-1/2}) \\
&= \tau(\tau T)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta u_t)^2 + o_p(1) \\
&\xrightarrow{p} \tau \sigma^2
\end{aligned}$$

So

$$\begin{aligned}
T\hat{\delta}_\tau &\xrightarrow{d} \frac{\frac{1}{2} \{ \sigma^2 K_{c,\alpha}^\beta(\tau, \tau, \tau_1, \tau_2)^2 - \sigma^2 K_{c,\alpha}^\beta(0, \tau, \tau_1, \tau_2)^2 - \tau \sigma^2 \}}{\sigma^2 \int_0^\tau K_{c,\alpha}^\beta(r, \tau, \tau_1, \tau_2)^2 dr} \\
&= \frac{K_{c,\alpha}^\beta(\tau, \tau, \tau_1, \tau_2)^2 - K_{c,\alpha}^\beta(0, \tau, \tau_1, \tau_2)^2 - \tau}{2 \int_0^\tau K_{c,\alpha}^\beta(r, \tau, \tau_1, \tau_2)^2 dr}
\end{aligned}$$

Next consider $\hat{\sigma}_\tau^2$ and then $s.e.(\hat{\delta}_\tau)$. We again obtain $\hat{\sigma}_\tau^2 \xrightarrow{p} \sigma^2$ and

$$\begin{aligned}
\{Ts.e.(\hat{\delta}_\tau)\}^2 &= \frac{\hat{\sigma}_\tau^2}{T^{-2} \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{u}_{\tau,t-1}^2} \\
&\xrightarrow{d} \frac{\sigma^2}{\sigma^2 \int_0^\tau K_{c,\alpha}^\beta(r, \tau, \tau_1, \tau_2)^2 dr} \\
&= \frac{1}{\int_0^\tau K_{c,\alpha}^\beta(r, \tau, \tau_1, \tau_2)^2 dr}
\end{aligned}$$

Putting it all together we find

$$\begin{aligned}
t_\tau &= \frac{T\hat{\delta}_\tau}{Ts.e.(\hat{\delta}_\tau)} \\
&\xrightarrow{d} \frac{\frac{K_{c,\alpha}^\beta(\tau, \tau, \tau_1, \tau_2)^2 - K_{c,\alpha}^\beta(0, \tau, \tau_1, \tau_2)^2 - \tau}{2 \int_0^\tau K_{c,\alpha}^\beta(r, \tau, \tau_1, \tau_2)^2 dr}}{\sqrt{\frac{1}{\int_0^\tau K_{c,\alpha}^\beta(r, \tau, \tau_1, \tau_2)^2 dr}}} \\
&= \frac{K_{c,\alpha}^\beta(\tau, \tau, \tau_1, \tau_2)^2 - K_{c,\alpha}^\beta(0, \tau, \tau_1, \tau_2)^2 - \tau}{2 \sqrt{\int_0^\tau K_{c,\alpha}^\beta(r, \tau, \tau_1, \tau_2)^2 dr}}
\end{aligned}$$

Asymptotic distribution of PWY_{GLS}^β

In this case the recursive test statistic calculated for the sub-sample $t = 1, \dots, \lfloor \tau T \rfloor$ is based on the t -ratio for $\tilde{\delta}_\tau$ from the estimated regression

$$\Delta \tilde{u}_{\tau,t} = \tilde{\delta}_\tau \tilde{u}_{\tau,t-1} + \tilde{v}_{\tau,t}, \quad t = 1, \dots, \lfloor \tau T \rfloor$$

where

$$\tilde{u}_{\tau,t} = y_t - \tilde{\mu}_\tau - \tilde{\beta}_\tau t$$

with $\tilde{\mu}_\tau$ and $\tilde{\beta}_\tau$ the recursively GLS detrending estimates. The t -ratio is then

$$t_\tau = \frac{\tilde{\delta}_\tau}{s.e.(\tilde{\delta}_\tau)}$$

with

$$\begin{aligned} \tilde{\delta}_\tau &= \frac{\sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \tilde{u}_{\tau,t} \tilde{u}_{\tau,t-1}}{\sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{u}_{\tau,t-1}^2} \\ s.e.(\tilde{\delta}_\tau)^2 &= \frac{\tilde{\sigma}_\tau^2}{\sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{u}_{\tau,t-1}^2} \\ \tilde{\sigma}_\tau^2 &= \lfloor \tau T \rfloor^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{v}_{\tau,t}^2 \end{aligned}$$

In what follows we can set $\mu = \beta = 0$, so that $y_t = u_t$ throughout. First consider the properties of the recursively detrending estimates:

$$\begin{bmatrix} \tilde{\mu}_\tau \\ \tilde{\beta}_\tau \end{bmatrix} = \begin{bmatrix} 1 + (\lfloor \tau T \rfloor - 1)(1 - \bar{\rho})^2 & 1 + (1 - \bar{\rho}) \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\} \\ 1 + (1 - \bar{\rho}) \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\} & 1 + \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\}^2 \end{bmatrix}^{-1} \begin{bmatrix} u_1 + (1 - \bar{\rho}) \sum_{t=2}^{\lfloor \tau T \rfloor} (u_t - \bar{\rho} u_{t-1}) \\ u_1 + \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\} \{u_t - \bar{\rho} u_{t-1}\} \end{bmatrix}$$

$$\begin{aligned} & T^{-1/2} \begin{bmatrix} 1 & 0 \\ 0 & T^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mu}_\tau \\ \tilde{\beta}_\tau \end{bmatrix} \\ &= \left(\begin{bmatrix} 1 + (\lfloor \tau T \rfloor - 1)(1 - \bar{\rho})^2 & 1 + (1 - \bar{\rho}) \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\} \\ 1 + (1 - \bar{\rho}) \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\} & 1 + \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\}^2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & T^{-1} \end{bmatrix} \right)^{-1} \\ & T^{-1/2} \begin{bmatrix} u_1 + (1 - \bar{\rho}) \sum_{t=2}^{\lfloor \tau T \rfloor} (u_t - \bar{\rho} u_{t-1}) \\ u_1 + \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\} \{u_t - \bar{\rho} u_{t-1}\} \end{bmatrix} \end{aligned}$$

$$\begin{aligned}
& \begin{bmatrix} T^{-1/2} \tilde{\mu}_\tau \\ T^{1/2} \tilde{\beta}_\tau \end{bmatrix} \\
&= \left(\begin{bmatrix} 1 + (\lfloor \tau T \rfloor - 1)(1 - \bar{\rho})^2 & T^{-1}[1 + (1 - \bar{\rho}) \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\}] \\ 1 + (1 - \bar{\rho}) \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\} & T^{-1}[1 + \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\}^2] \end{bmatrix} \right)^{-1} \\
&\quad \begin{bmatrix} T^{-1/2} u_1 + T^{-1/2} (1 - \bar{\rho}) \sum_{t=2}^{\lfloor \tau T \rfloor} (u_t - \bar{\rho} u_{t-1}) \\ T^{-1/2} u_1 + T^{-1/2} \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\} \{u_t - \bar{\rho} u_{t-1}\} \end{bmatrix} \\
&\equiv \begin{bmatrix} A_1 & T^{-1} A_2 \\ A_2 & A_3 \end{bmatrix}^{-1} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}
\end{aligned}$$

Taking each term separately:

$$\begin{aligned}
A_1 &= 1 + (\lfloor \tau T \rfloor - 1)(1 - \bar{\rho})^2 \\
&= 1 - \bar{c}^2 T^{-2} (\lfloor \tau T \rfloor - 1) \\
&\xrightarrow{d} 1
\end{aligned}$$

$$\begin{aligned}
A_2 &= 1 + (1 - \bar{\rho}) \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\} \\
&= 1 - \bar{c} T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \{1 - \bar{c} T^{-1}(t-1)\} \\
&\xrightarrow{d} 1 - \bar{c} \tau + \bar{c}^2 \tau^2 / 2
\end{aligned}$$

$$\begin{aligned}
A_3 &= T^{-1} [1 + \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\}^2] \\
&= T^{-1} [1 + \sum_{t=2}^{\lfloor \tau T \rfloor} \{1 - \bar{c} T^{-1}(t-1)\}^2] \\
&\xrightarrow{d} \tau - \bar{c} \tau^2 + \bar{c}^2 \tau^3 / 3
\end{aligned}$$

$$\begin{aligned}
B_1 &= T^{-1/2} u_1 + T^{-1/2} (1 - \bar{\rho}) \sum_{t=2}^{\lfloor \tau T \rfloor} (u_t - \bar{\rho} u_{t-1}) \\
&= T^{-1/2} u_1 - \bar{c} T^{-3/2} \sum_{t=2}^{\lfloor \tau T \rfloor} (u_t - \bar{\rho} u_{t-1}) \\
&\xrightarrow{p} \sigma \alpha
\end{aligned}$$

$$\begin{aligned}
B_2 &= T^{-1/2}u_1 + T^{-1/2} \sum_{t=2}^{\lfloor \tau T \rfloor} \{t - \bar{\rho}(t-1)\} \{u_t - \bar{\rho}u_{t-1}\} \\
&= T^{-1/2}u_1 + T^{-1/2} \sum_{t=2}^{\lfloor \tau T \rfloor} \{1 - \bar{c}T^{-1}(t-1)\} (\Delta u_t - \bar{c}T^{-1}u_{t-1}) \\
&= T^{-1/2}u_1 + T^{-1/2} \sum_{t=2}^{\lfloor \tau T \rfloor} \{\Delta u_t - \bar{c}T^{-1}u_{t-1}\} - \bar{c}T^{-3/2} \sum_{t=2}^{\lfloor \tau T \rfloor} t(\Delta u_t - \bar{c}T^{-1}u_{t-1}) + o_p(1) \\
&= T^{-1/2}u_1 + T^{-1/2}(u_{\lfloor \tau T \rfloor} - u_1) - \bar{c}T^{-3/2} \sum_{t=2}^{\lfloor \tau T \rfloor} u_{t-1} - \bar{c}T^{-3/2} \sum_{t=2}^{\lfloor \tau T \rfloor} t\Delta u_t \\
&\quad + \bar{c}^2T^{-5/2} \sum_{t=2}^{\lfloor \tau T \rfloor} tu_{t-1} + o_p(1) \\
&= T^{-1/2}u_{\lfloor \tau T \rfloor} - \bar{c}T^{-3/2} \sum_{t=2}^{\lfloor \tau T \rfloor} u_{t-1} - \bar{c}T^{-3/2} \sum_{t=2}^{\lfloor \tau T \rfloor} t\Delta u_t + \bar{c}^2T^{-5/2} \sum_{t=2}^{\lfloor \tau T \rfloor} tu_{t-1} + o_p(1) \\
&\xrightarrow{d} \sigma K_{c,\alpha}(\tau, \tau_1, \tau_2) - \bar{c}\sigma \int_0^\tau K_{c,\alpha}(s, \tau_1, \tau_2)ds - \bar{c}\sigma \int_0^\tau s dK_{c,\alpha}(s, \tau_1, \tau_2) \\
&\quad + \bar{c}^2\sigma \int_0^\tau s K_{c,\alpha}(s, \tau_1, \tau_2)ds \\
&= \sigma K_{c,\alpha}(\tau, \tau_1, \tau_2) - \bar{c}\sigma \int_0^\tau K_{c,\alpha}(s, \tau_1, \tau_2)ds \\
&\quad - \bar{c}\sigma \left\{ \tau K_{c,\alpha}(\tau, \tau_1, \tau_2) - \int_0^\tau K_{c,\alpha}(s, \tau_1, \tau_2)ds \right\} + \bar{c}^2\sigma \int_0^\tau s K_{c,\alpha}(s, \tau_1, \tau_2)ds \\
&= \sigma(1 - \bar{c}\tau)K_{c,\alpha}(\tau, \tau_1, \tau_2) + \bar{c}^2\sigma \int_0^\tau s K_{c,\alpha}(s, \tau_1, \tau_2)ds
\end{aligned}$$

Substituting back in gives

$$\begin{aligned}
\begin{bmatrix} T^{-1/2}\tilde{\mu}_\tau \\ T^{1/2}\tilde{\beta}_\tau \end{bmatrix} &\xrightarrow{d} \begin{bmatrix} 1 & 0 \\ 1 - \bar{c}\tau + \bar{c}^2\tau^2/2 & \tau - \bar{c}\tau^2 + \bar{c}^2\tau^3/3 \end{bmatrix}^{-1} \\
&\begin{bmatrix} \sigma\alpha \\ \sigma(1 - \bar{c}\tau)K_{c,\alpha}(\tau, \tau_1, \tau_2) + \bar{c}^2\sigma \int_0^\tau sK_{c,\alpha}(s, \tau_1, \tau_2)ds \end{bmatrix} \\
&= \sigma \begin{bmatrix} 1 & 0 \\ -\frac{1-\bar{c}\tau+\bar{c}^2\tau^2/2}{\tau-\bar{c}\tau^2+\bar{c}^2\tau^3/3} & (\tau - \bar{c}\tau^2 + \bar{c}^2\tau^3/3)^{-1} \end{bmatrix} \\
&\begin{bmatrix} \alpha \\ (1 - \bar{c}\tau)K_{c,\alpha}(\tau, \tau_1, \tau_2) + \bar{c}^2 \int_0^\tau sK_{c,\alpha}(s, \tau_1, \tau_2)ds \end{bmatrix} \\
&= \sigma \begin{bmatrix} \alpha \\ (\tau - \bar{c}\tau^2 + \bar{c}^2\tau^3/3)^{-1}\{(1 - \bar{c}\tau)K_{c,\alpha}(\tau, \tau_1, \tau_2) \\ + \bar{c}^2 \int_0^\tau sK_{c,\alpha}(s, \tau_1, \tau_2)ds\} - \frac{1-\bar{c}\tau+\bar{c}^2\tau^2/2}{\tau-\bar{c}\tau^2+\bar{c}^2\tau^3/3}\alpha \end{bmatrix}
\end{aligned}$$

Next consider the recursively detrended y_t , i.e. $\tilde{u}_{\tau,t}$. Since $y_t = u_t$ we have

$$\begin{aligned}
\tilde{u}_{\tau,t} &= u_t - \tilde{\mu}_\tau - \tilde{\beta}_\tau t \\
T^{-1/2}\tilde{u}_{\tau,[rT]} &= T^{-1/2}u_{[rT]} - T^{-1/2}\tilde{\mu}_\tau - T^{-1/2}\tilde{\beta}_\tau t \\
&= T^{-1/2}u_{[rT]} - T^{-1/2}\tilde{\mu}_\tau - T^{1/2}\tilde{\beta}_\tau r \\
&\xrightarrow{d} \sigma K_{c,\alpha}(r, \tau_1, \tau_2) - \sigma\alpha - \\
&\quad \sigma \begin{bmatrix} (\tau - \bar{c}\tau^2 + \bar{c}^2\tau^3/3)^{-1}\{(1 - \bar{c}\tau)K_{c,\alpha}(\tau, \tau_1, \tau_2) \\ + \bar{c}^2 \int_0^\tau sK_{c,\alpha}(s, \tau_1, \tau_2)ds\} - \frac{1-\bar{c}\tau+\bar{c}^2\tau^2/2}{\tau-\bar{c}\tau^2+\bar{c}^2\tau^3/3}\alpha \end{bmatrix} r \\
&\equiv \sigma K_{c,\bar{c},\alpha}^{\beta,G}(r, \tau, \tau_1, \tau_2)
\end{aligned}$$

Now consider the recursive parameter estimate $\tilde{\delta}_\tau$:

$$\begin{aligned}
\tilde{\delta}_\tau &= \frac{\sum_{t=2}^{[rT]} \Delta \tilde{u}_{\tau,t} \tilde{u}_{\tau,t-1}}{\sum_{t=2}^{[rT]} \tilde{u}_{\tau,t-1}^2} \\
T\tilde{\delta}_\tau &= \frac{T^{-1} \sum_{t=2}^{[rT]} \Delta \tilde{u}_{\tau,t} \tilde{u}_{\tau,t-1}}{T^{-2} \sum_{t=2}^{[rT]} \tilde{u}_{\tau,t-1}^2}
\end{aligned}$$

For the denominator we have

$$\begin{aligned}
T^{-2} \sum_{t=2}^{[rT]} \tilde{u}_{\tau,t-1}^2 &= T^{-1} \sum_{[rT]=2}^{[rT]} (T^{-1/2}\tilde{u}_{\tau,[rT]-1})^2 \\
&\xrightarrow{d} \sigma^2 \int_0^\tau K_{c,\bar{c},\alpha}^{\beta,G}(r, \tau, \tau_1, \tau_2)^2 dr
\end{aligned}$$

Now consider the numerator

$$\begin{aligned}
T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \Delta \tilde{u}_{\tau,t} \tilde{u}_{\tau,t-1} &= \frac{1}{2} \left\{ (T^{-1/2} \tilde{u}_{\lfloor \tau T \rfloor})^2 - (T^{-1/2} \tilde{u}_1)^2 - T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta \tilde{u}_{\tau,t})^2 \right\} \\
&\xrightarrow{d} \frac{1}{2} \left\{ \sigma^2 K_{c,\bar{c},\alpha}^{\beta,G}(\tau, \tau, \tau_1, \tau_2)^2 - \sigma^2 K_{c,\bar{c},\alpha}^{\beta,G}(0, \tau, \tau_1, \tau_2)^2 - \tau \sigma^2 \right\} \\
&= \frac{1}{2} \left\{ \sigma^2 K_{c,\bar{c},\alpha}^{\beta,G}(\tau, \tau, \tau_1, \tau_2)^2 - \tau \sigma^2 \right\}
\end{aligned}$$

using $T^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} (\Delta \tilde{u}_{\tau,t})^2 \xrightarrow{p} \tau \sigma^2$ as before, and noting that $K_{c,\bar{c},\alpha}^{\beta,G}(0, \tau, \tau_1, \tau_2) = 0$.

So

$$\begin{aligned}
T \hat{\delta}_\tau &\xrightarrow{d} \frac{\frac{1}{2} \left\{ \sigma^2 K_{c,\bar{c},\alpha}^{\beta,G}(\tau, \tau, \tau_1, \tau_2)^2 - \sigma^2 K_{c,\bar{c},\alpha}^{\beta,G}(0, \tau, \tau_1, \tau_2)^2 - \tau \sigma^2 \right\}}{\sigma^2 \int_0^\tau K_{c,\bar{c},\alpha}^{\beta,G}(r, \tau, \tau_1, \tau_2)^2 dr} \\
&= \frac{K_{c,\bar{c},\alpha}^{\beta,G}(\tau, \tau, \tau_1, \tau_2)^2 - \tau}{2 \int_0^\tau K_{c,\bar{c},\alpha}^{\beta,G}(r, \tau, \tau_1, \tau_2)^2 dr}
\end{aligned}$$

Next consider $\tilde{\sigma}_\tau^2$ and then $s.e.(\tilde{\delta}_\tau)$. We again obtain $\tilde{\sigma}_\tau^2 \xrightarrow{p} \sigma^2$ and

$$\begin{aligned}
\{Ts.e.(\tilde{\delta}_\tau)\}^2 &= \frac{\tilde{\sigma}_\tau^2}{T^{-2} \sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{u}_{\tau,t-1}^2} \\
&\xrightarrow{d} \frac{\sigma^2}{\sigma^2 \int_0^\tau K_{c,\bar{c},\alpha}^{\beta,G}(r, \tau, \tau_1, \tau_2)^2 dr} \\
&= \frac{1}{\int_0^\tau K_{c,\bar{c},\alpha}^{\beta,G}(r, \tau, \tau_1, \tau_2)^2 dr}
\end{aligned}$$

Putting it all together we find

$$\begin{aligned}
t_\tau &= \frac{T \tilde{\delta}_\tau}{Ts.e.(\tilde{\delta}_\tau)} \\
&\xrightarrow{d} \frac{K_{c,\bar{c},\alpha}^{\beta,G}(\tau, \tau, \tau_1, \tau_2)^2 - \tau}{2 \sqrt{\int_0^\tau K_{c,\bar{c},\alpha}^{\beta,G}(r, \tau, \tau_1, \tau_2)^2 dr}}
\end{aligned}$$

Table 2.1: Asymptotic critical values for PWY_{OLS}^μ , PWY_{GLS}^μ , PWY_{OLS}^β , PWY_{GLS}^β , and λ_ζ^μ and λ_ζ^β values

ζ	PWY_{OLS}^μ	PWY_{GLS}^μ	PWY_{OLS}^β	PWY_{GLS}^β	λ_ζ^μ	λ_ζ^β
0.10	1.166	2.319	0.298	4.592	1.111	1.238
0.05	1.433	2.626	0.529	5.215	1.096	1.168
0.01	1.923	3.223	1.007	6.415	1.081	1.089

Asymptotic null critical values are generated by direct simulation of the limit distributions $L_{c,\tau_1,\tau_2}^{OLS^\mu}$, $L_{c,\tau_1,\tau_2}^{GLS^\mu}$, $L_{c,\tau_1,\tau_2}^{OLS^\beta}$ and $L_{c,\tau_1,\tau_2}^{GLS^\beta}$ using $IID N(0, 1)$ random variates, with the integrals approximated by normalised sums of 1,000 steps.

Table 2.2: Finite sample critical values, $T = 150$

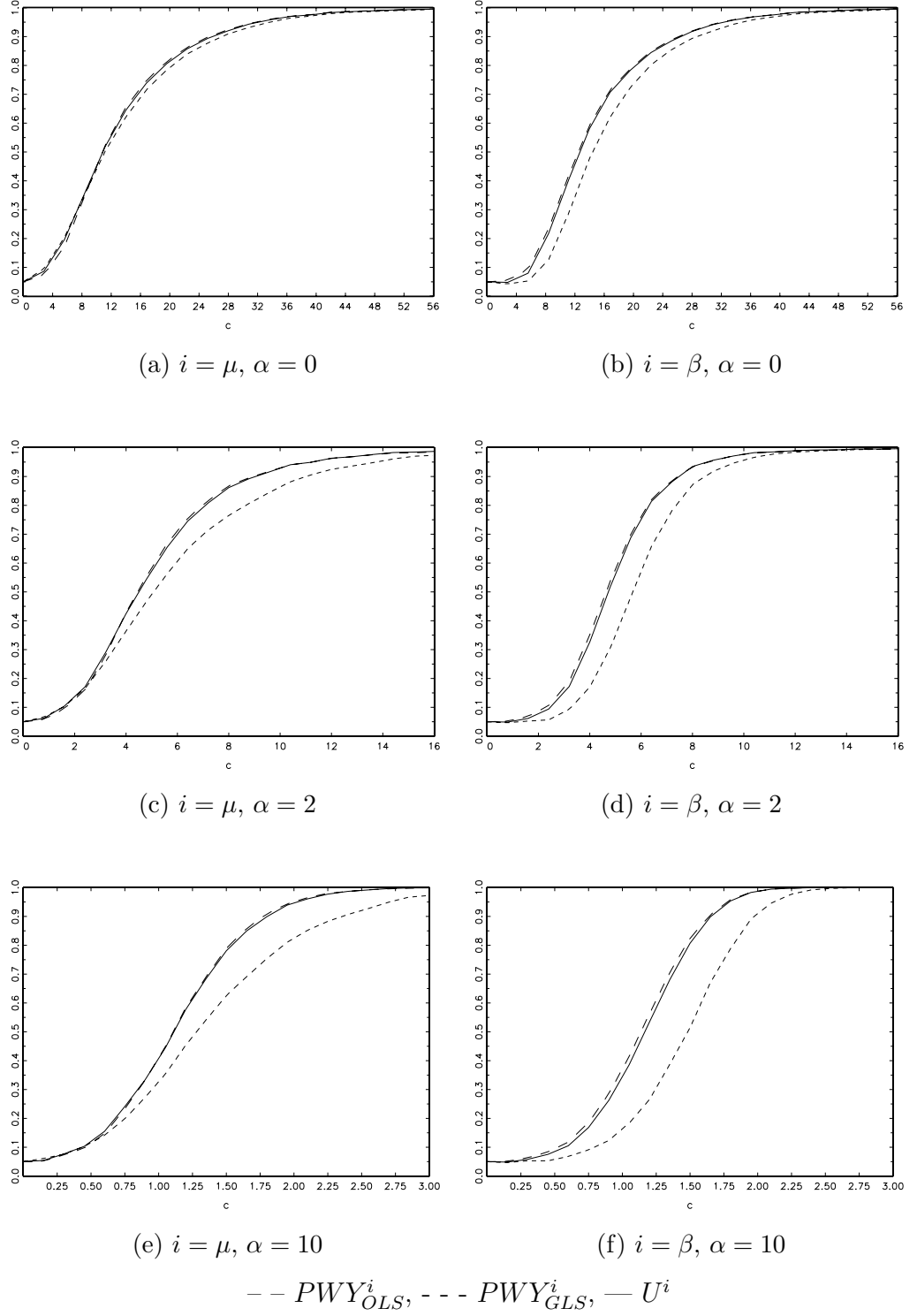
	PWY_{OLS}^μ	PWY_{GLS}^μ	PWY_{OLS}^β	PWY_{GLS}^β
0.10	1.174	2.498	0.308	5.950
0.05	1.467	2.906	0.572	6.633
0.01	2.137	3.634	1.137	7.980

Table 2.3: Application of PWY_{OLS}^μ , PWY_{GLS}^μ , PWY_{OLS}^β , PWY_{GLS}^β , U^μ , U^β to NASDAQ prices

End date	PWY_{OLS}^μ	PWY_{GLS}^μ	PWY_{OLS}^β	PWY_{GLS}^β	U^μ	U^β
31 Dec 2015	1.147	3.394***	1.398***	2.923	**	***
31 Dec 2005	1.147	3.394***	1.398***	5.543**	**	***

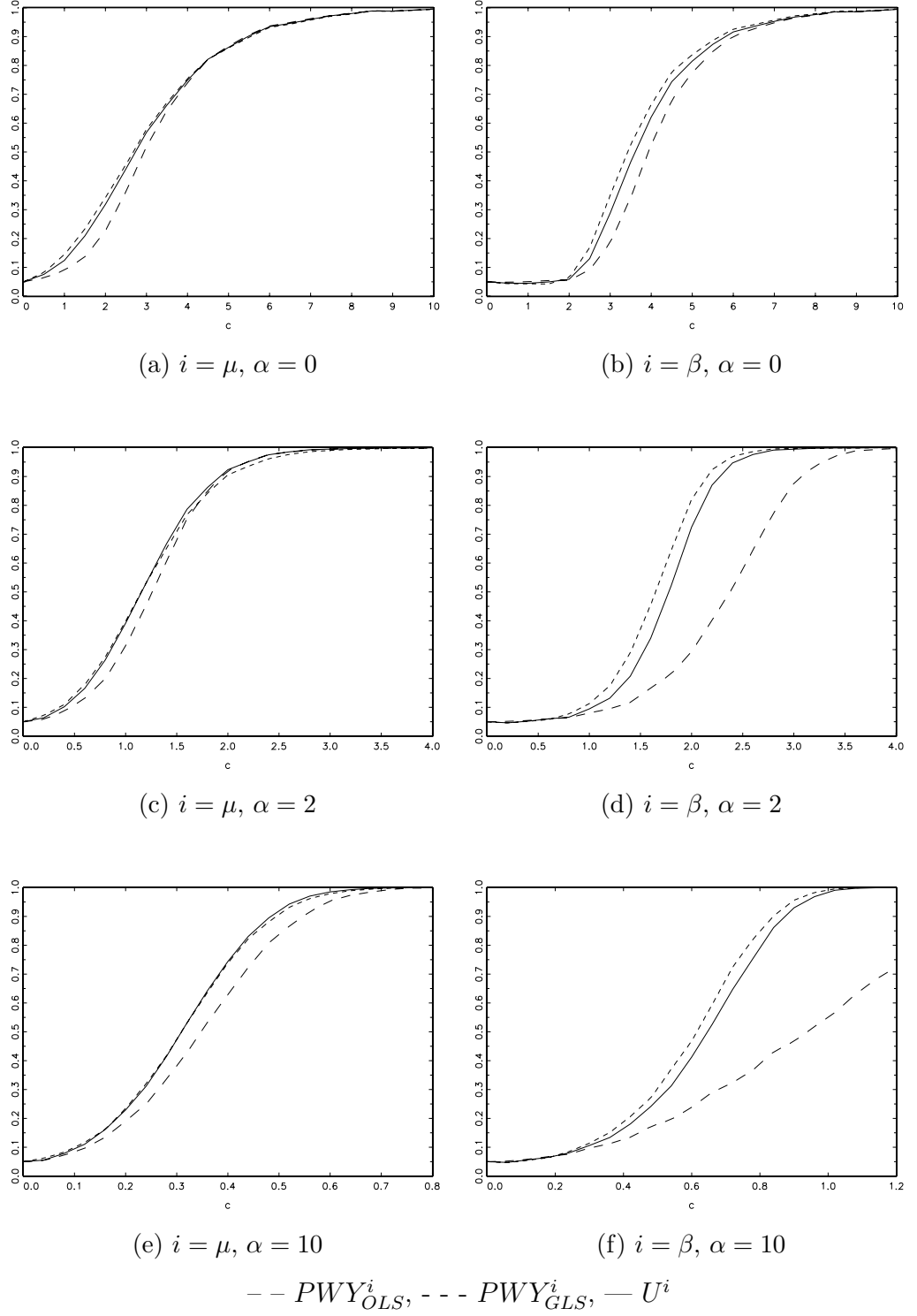
** and *** indicate rejections at a 0.05 and 0.01 level respectively

Figure 2.1: Asymptotic power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.45, 0.55]$, where $i = \{\mu, \beta\}$



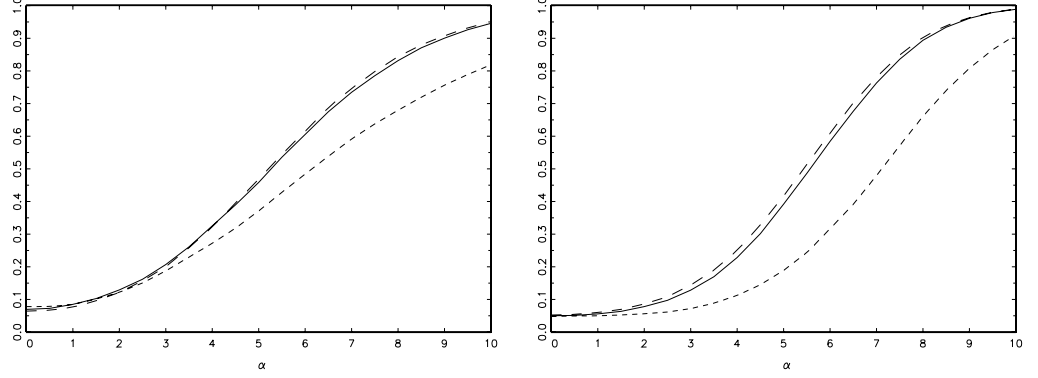
Power results are displayed for the DGP given by equations (2.3)-(2.4) where $\delta = c/T$

Figure 2.2: Asymptotic power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.2, 0.8]$, where $i = \{\mu, \beta\}$



Power results are displayed for the DGP given by equations (2.3)-(2.4) where $\delta = c/T$

Figure 2.3: Asymptotic power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.45, 0.55]$, where $i = \{\mu, \beta\}$



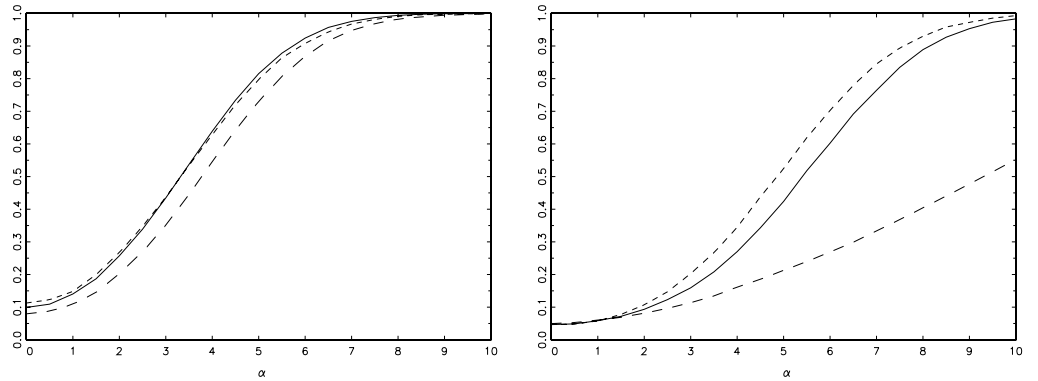
(a) $i = \mu, c = 2$

(b) $i = \beta, c = 2$

-- PWY_{OLS}^i , --- PWY_{GLS}^i , — U^i

Power results are displayed for the DGP given by equations (2.3)-(2.4) where $\delta = c/T$

Figure 2.4: Asymptotic power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.2, 0.8]$, where $i = \{\mu, \beta\}$



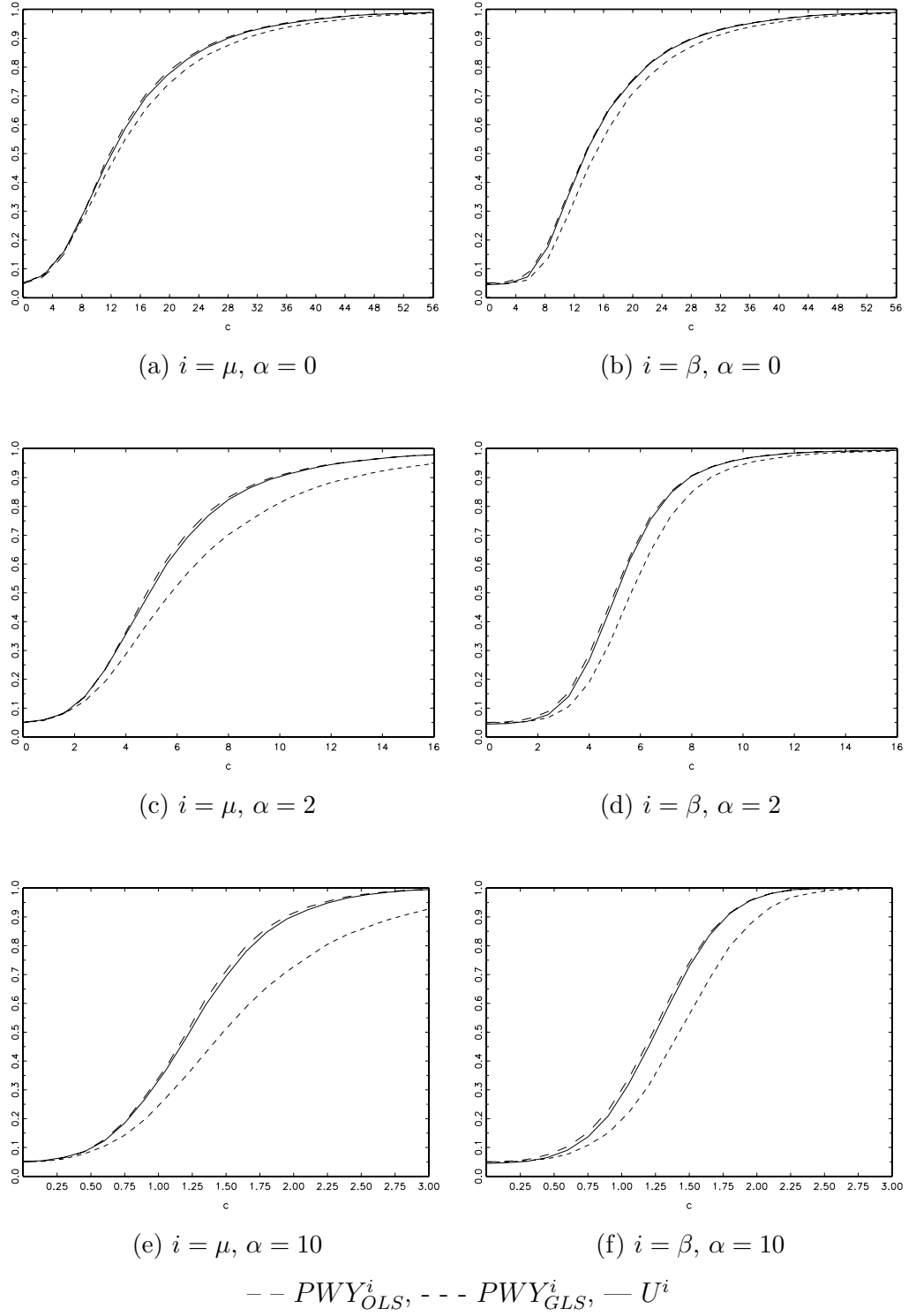
(a) $i = \mu, c = 0.8$

(b) $i = \beta, c = 1$

-- PWY_{OLS}^i , --- PWY_{GLS}^i , — U^i

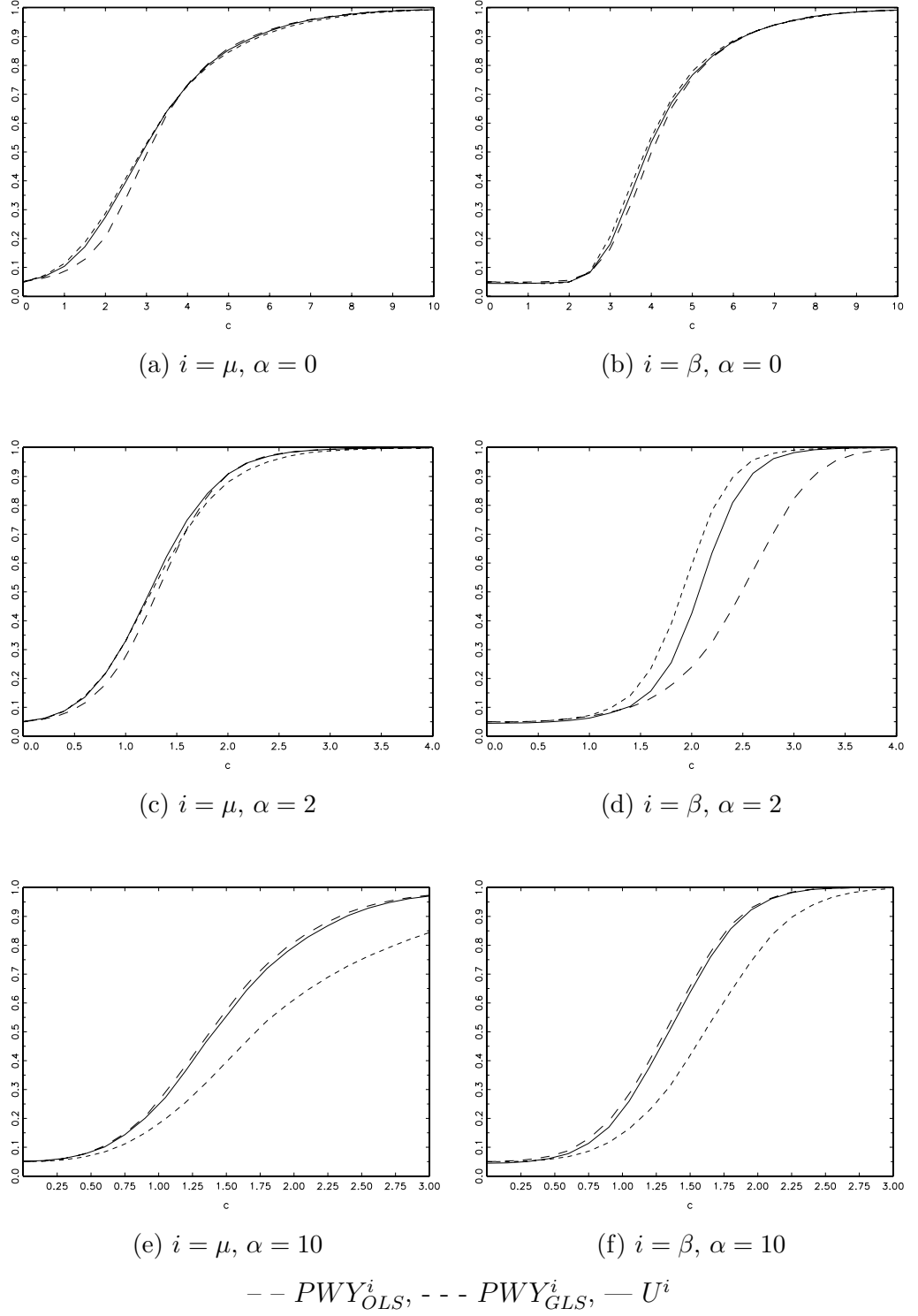
Power results are displayed for the DGP given by equations (2.3)-(2.4) where $\delta = c/T$

Figure 2.5: Finite sample power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.45, 0.55]$, where $i = \{\mu, \beta\}$



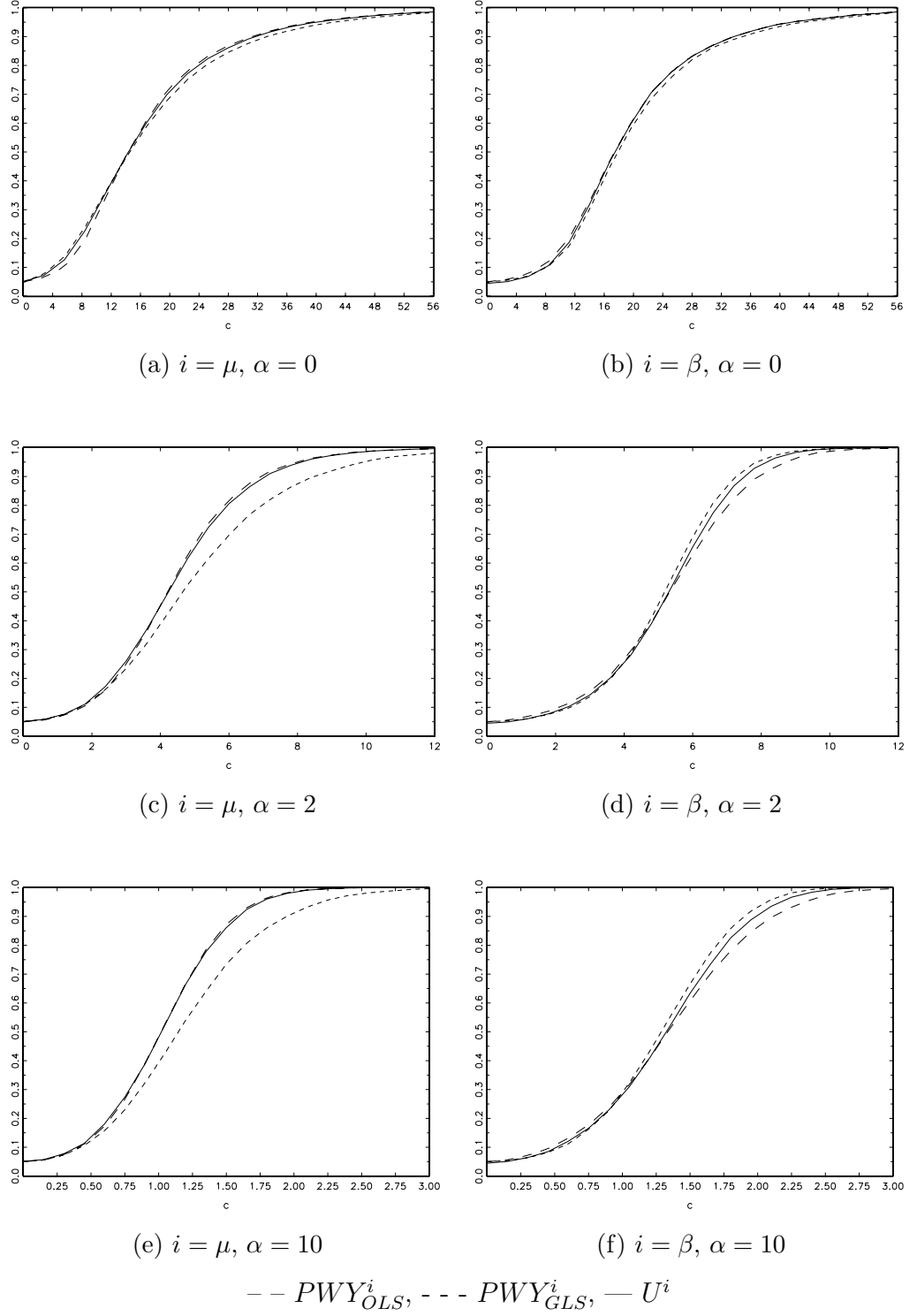
Power results are displayed for the DGP given by equations (2.3)-(2.4) where $\delta = c/T$

Figure 2.6: Finite sample power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.2, 0.8]$, where $i = \{\mu, \beta\}$



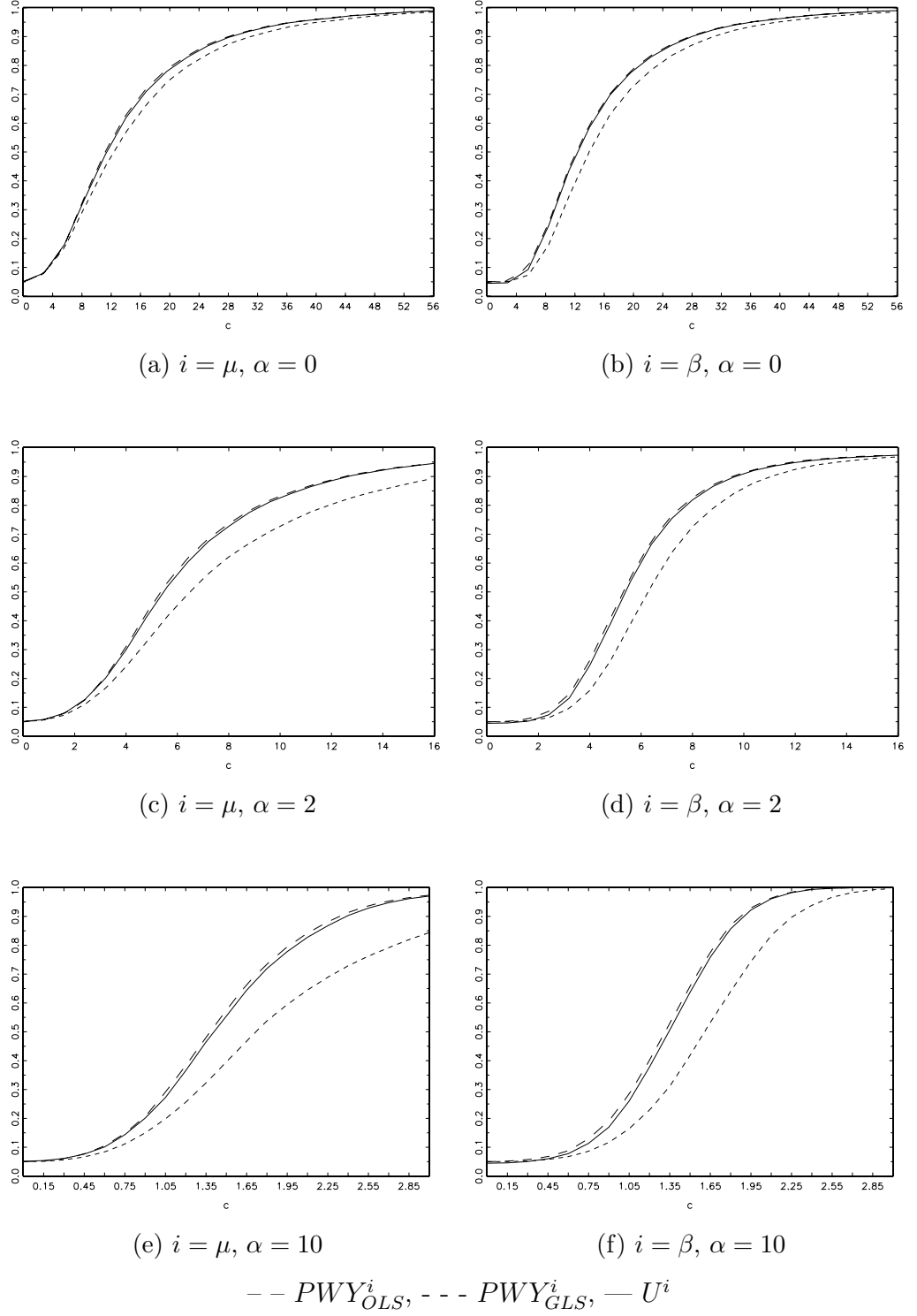
Power results are displayed for the DGP given by equations (2.3)-(2.4) where $\delta = c/T$

Figure 2.7: Finite sample power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.15, 0.25]$, where $i = \{\mu, \beta\}$



Power results are displayed for the DGP given by equations (2.3)-(2.4) where $\delta = c/T$

Figure 2.8: Finite sample power of PWY_{OLS}^i , PWY_{GLS}^i and U^i for $[\tau_1, \tau_2] = [0.75, 0.85]$, where $i = \{\mu, \beta\}$



Power results are displayed for the DGP given by equations (2.3)-(2.4) where $\delta = c/T$

Figure 2.9: Logarithms of NASDAQ closing prices



(a) 2 Jan 1996 - 31 Dec 2015



(b) 2 Jan 1996 - 31 Dec 2005

Chapter 3

Testing for a unit root against ESTAR stationarity

In this chapter we examine the local power of unit root tests against globally stationary exponential smooth transition autoregressive [ESTAR] alternatives under two sources of uncertainty: the degree of nonlinearity in the ESTAR model, and the presence of a linear deterministic trend. First, we show that the Kapetanios, Shin and Snell (2003, *Journal of Econometrics* 112, 359–379) [KSS] test for nonlinear stationarity has local asymptotic power gains over standard Dickey-Fuller [DF] tests for certain degrees of nonlinearity in the ESTAR model, but that for other degrees of nonlinearity, the linear DF test has superior power. Second, we derive limiting distributions of demeaned, and demeaned and detrended KSS and DF tests under a local ESTAR alternative when a local trend is present in the DGP. We show that the power of the demeaned tests outperforms that of the detrended tests when no trend is present in the DGP, but deteriorates as the magnitude of the trend increases. We propose a union of rejections testing procedure that combines all four individual tests and show that this captures most of the power available from the individual tests across different degrees of nonlinearity and trend magnitudes. We also show that incorporating a trend detection procedure into this union testing strategy can result in higher power when a large trend is present in the DGP. An empirical application of our proposed union of rejections procedures to energy consumption data in 180 countries shows the value of these procedures in practical cases.

3.1 Introduction

The application of standard DF unit root tests to certain economic time series, such as real exchange rates or real interest rates, has often led to surprising results. Whilst economic theory suggests that these series should be stationary processes, many authors find that they are unable to reject the unit root null hypothesis. This has led to concern that the DF unit root test has low power against some classes of stationary alternatives as the linear ARMA framework it utilises is unable to capture the underlying dynamics of the economic time series. To address this problem, Kapetanios et al. (2003) [KSS] propose a test of the null hypothesis of a unit root against the alternative of a nonlinear, but globally stationary, exponential smooth transition autoregressive (ESTAR) process. They demonstrate that this procedure has power gains over standard DF tests under some ESTAR alternatives, and the test has subsequently been frequently employed in the applied econometric literature.

Demetrescu and Kruse (2012) [DK] demonstrate the sensitivity of the relative power performance of DF and KSS unit root tests to the degree of nonlinearity under the ESTAR alternative, such that when the degree of nonlinearity is small relative to the error variance, the DF test has a power advantage over the KSS test. They therefore suggest that a union of rejections approach, along the lines of Harvey et al. (2009, 2012) would provide a successful strategy for unit root testing that is robust to the degree of nonlinearity under the ESTAR alternative. We formalise this union testing strategy and demonstrate that such a procedure is able to capitalize on the respective power advantages exhibited by the DF and KSS tests under different DGPs, and can offer power close to the effective envelope across all degrees of nonlinearity.

Unit root testing is further complicated in practice by the issue of whether or not to account for a linear deterministic trend. If a trend is present in the DGP, Harvey et al. (2009) show that a DF unit root test which is demeaned but not detrended will have decreasing power as the magnitude of this trend increases. However, when no trend is present, the demeaned DF test is shown to have superior power to a DF test which is both demeaned and detrended. Motivated by these concerns, we derive limiting distributions of DF and KSS tests under a local ESTAR alternative when a linear deterministic trend is present in the DGP. We show that, as in the case of the DF test, the relative power of demeaned, and demeaned and detrended, KSS tests is sensitive to the presence or otherwise of a linear deterministic trend. Therefore, in addition to considering a union of rejections strategy based on demeaned DF and KSS tests, and another based

on demeaned and detrended DF and KSS tests, we propose a union procedure which combines all four individual tests. This union procedure is shown to achieve attractive power levels across all settings of the trend coefficient and nonlinearity parameter.

Harvey et al. (2012) note that the power of a union of rejections procedure which includes both detrended and non-detrended unit root tests can be improved by utilising information about the presence of a trend. We therefore incorporate a BIC-based trend detection method into our union test, such that if strong evidence of a trend is found only the detrended tests enter into the union procedure. Finite sample simulations show that this approach is indeed able to improve on the power offered by the simple four-test union of rejections when the magnitude of the trend coefficient is large.

Finally, we apply the union of rejections procedures considered in this chapter to energy consumption data across the period 1980-2011 for 180 countries. We find that the union of rejections procedures are able to capitalize on the differing rejections offered by the four individual tests, and at both a 0.10 and 0.05 significance level, the BIC-based union of rejections procedure achieves rejections for more countries than any individual test or other union of rejections procedure considered in this chapter.

The outline of this chapter is as follows. Section 3.2 and 3.3 review the literature on testing for nonlinear stationarity and dealing with uncertainty over the presence of a trend respectively. In Section 3.4 we briefly outline the ESTAR model and the KSS and DF tests for a unit root. Section 3.5 presents the limit distributions of the four individual tests under a local ESTAR alternative for both local to zero and fixed magnitude trends. In Section 3.6 we present asymptotic power simulations and outline the different union of rejections procedures that we propose. Section 3.8 considers the finite sample performance of the proposed test procedures, and the empirical application to energy consumption data is given in Section 3.9. Section 3.10 concludes.

3.2 Nonlinear stationarity

Standard unit root tests, such as the ordinary least squares (OLS) DF test of Dickey and Fuller (1979) or the generalized least squares (GLS) version proposed by Elliott et al. (1996), consider the null hypothesis of a unit root against the alternative of linear stationarity. There exist several areas of economics where economic theory suggests a variable should be stationary, but these DF tests fail to reject the null hypothesis of a unit root. Perhaps the most obvious example of

this phenomenon occurring is with exchange rates. If variations in real exchange rates represent deviations from Purchasing Power Parity (PPP), then the finding of nonstationary behaviour in real exchange rates undermines the existence of a long-run tendency to PPP (Taylor et al., 2001).

Faced with conflicts of this nature between economic theory and empirical results, the literature has approached the problem in two ways. Some researchers have chosen to accept the empirical results of nonstationarity and reformulate the economic theory accordingly. Edison and Klovland (1987), for example, argue that the PPP hypothesis holds in the long run when both short run cyclical variables that cause temporary deviations from PPP and long run structural factors such as relative levels of productivity are accounted for. Others, however, have suggested that the finding of nonstationarity should not necessarily be interpreted as a rejection of the PPP hypothesis but instead as the result of flawed econometric methodology. It is this view that we focus on throughout the remainder of the chapter.

In the context of exchange rates, a large number of studies have focused on the post-Bretton Woods era, i.e. from 1973 onwards, yielding a small number of observations with which to test the PPP hypothesis. Taylor et al. (2001) note that conventional DF tests may have low statistical power to reject a unit root null in small sample sizes such as these. A number of papers have therefore attempted to increase the power of unit root tests by using a longer sample period of data. Lothian and Taylor (1996), for example, employ a DF test on annual dollar-sterling exchange rate data between 1791 and 1900, as well as the franc-sterling exchange rate between 1803 and 1990, and are able to reject the unit root null in both series. Other papers attempt to increase the power of standard unit root tests by using panel data, see Abuaf and Jorion (1990) for example, who are able to reject the null hypothesis of a unit root using DF tests in a multilateral framework.

The failure of standard tests to reject a unit root null has led to additional concerns that, for the purposes of unit root testing, a linear time series framework might provide an insufficiently rich description of the underlying dynamics of many series. Indeed, a number of papers have found theoretical justification for modelling certain economic time series as nonlinear processes. For example, *inter alios*, Sercu et al. (1995) develop equilibrium models of real exchange rates in the presence of transaction costs, and show that the introduction of transaction costs results in a nonlinear, rather than linear, adjustment process towards PPP. A particular type of nonlinear model, the smooth threshold autoregressive (STAR) model, has received much attention in the literature (for a review, see van Dijk

et al. (2002)). The STAR framework allows for a smooth transition between two autoregressive regimes such that the process exhibits a sharper mean reversion when it suffers from larger deviations away from its equilibrium but, for smaller deviations, displays more persistent, unit root type behaviour. Michael et al. (1997) argue that not accounting for STAR nonlinearity in unit root testing may explain the failure of previous studies to reject a unit root null in favour of the PPP hypothesis. Evidence therefore suggests that this form of nonlinearity may provide a more appropriate framework within which to test for unit roots for various economic series.

Kapetanios et al. (2003) [KSS] propose a test of the null hypothesis of a unit root against the alternative of a nonlinear, but globally stationary, exponential STAR (ESTAR) process. Under this alternative hypothesis it has been shown that KSS has power gains over the standard DF test, particularly where the process is highly persistent. The KSS test has subsequently been widely used in empirical applications on a variety of economic series such as, *inter alia*, real interest rates (Baharumshah et al., 2009; Tsong and Lee, 2013; Guney and Hasanov, 2014), GDP (Beechey and Osterholm, 2008; Cook, 2008) and current account balances (Akdogan, 2014; Chen, 2014).

Demetrescu and Kruse (2012) [DK] compare the KSS test to the linear DF test and show that the size of the error variance has an impact on their relative power performance against local ESTAR alternatives. In situations where the degree of nonlinearity is small relative to the error variance, the DF test can have superior local asymptotic power. In contrast, when the nonlinearity is large relative to the error variance, the KSS test maintains a power advantage over DF. DK therefore argue that a union of rejections approach, following Harvey et al. (2009, 2012), could provide a successful strategy for unit root testing against ESTAR alternatives. The first major contribution of this chapter is therefore to formally consider such a union of rejections approach, whereby the null hypothesis is rejected if either of the individual DF or KSS tests reject, subject to a critical value scaling that ensures the empirical size of the overall procedure is equal to nominal size asymptotically. Unit root testing in practice is further complicated by uncertainty over the trend, an issue which DK abstract from. The second major contribution of this chapter is therefore to consider the effect that trend uncertainty can have on the size and power performance of unit root tests, as outlined in the next section.

3.3 Trend uncertainty

Most economic time series are not characterized by zero-mean processes, but instead contain non-zero means or possibly non-zero means plus linear deterministic trends. KSS suggest accommodating these deterministic features by prior OLS demeaning, or demeaning and detrending of the data. In empirical applications, inference made from KSS unit root tests will often depend on the deterministic specification assumed. For example, Christidou et al. (2013) consider the persistence of carbon dioxide emissions. They employ the no deterministic, demeaned, and demeaned and detrended KSS tests on a set of 36 countries. At least one of these three tests is able to reject the null hypothesis for the vast majority of countries. However, inference depends on whether or not a mean and trend are accommodated in the deterministic specification of the unit root tests. Kisswani and Nusair (2013) investigate the dynamics of real oil prices in seven Asian currencies plus the US dollar, using the same three KSS specifications. Again, they find that they are able to reject the null hypothesis of a unit root in most cases using at least one test, but that the three different KSS specifications provide conflicting results.

In situations such as these, where inference depends on the choice of deterministic specification, a practitioner uncertain about the presence or otherwise of a trend must choose which test to employ. A risk-averse strategy might be to always employ trend-invariant unit root tests. However, Marsh (2007) shows that, in the case of standard DF tests where data is generated by a linear AR process, the Fisher information for a test statistic invariant to a linear trend is zero at the unit root. Consequently, when a trend is absent, the power of a unit root test that is invariant to a trend will be compromised relative to the power of an appropriate demeaned but not detrended test statistic. Harvey et al. (2009) show that these power losses can be substantial, therefore opting to always use the trend-invariant test is a costly strategy. Conversely when a trend is present, the power of a DF test that is demeaned but not detrended is shown to decrease as the magnitude of the trend increases. Motivated by these considerations, in this chapter we not only examine the power performance of a union of rejections based on demeaned tests and a union of rejections based on demeaned and detrended tests, but also consider a union of rejections based on all four possible tests (i.e. demeaned DF and KSS tests, and demeaned and detrended DF and KSS tests).

In a related setting, Harvey et al. (2012) note that use of information regarding the presence of a trend can be applied to reduce the number of tests entering the

union of rejections when it is clear that a trend exists, allowing more power to be achieved. We therefore consider a variant procedure where the Bayesian Information Criterion (BIC) is used to determine whether there is evidence of a trend. When the BIC suggests a trend is present, the procedure employs a union of rejections of only the two demeaned and detrended DF and KSS tests, while otherwise all four tests are used in the union of rejections.

3.4 Unit root tests and the ESTAR model

A univariate STAR model of order 1, i.e. a STAR(1) model, is given by

$$y_t = \rho y_{t-1} + \gamma y_{t-1} G(\theta, y_{t-d}) + \varepsilon_t, \quad t = d+1, \dots, T$$

where $\varepsilon_t \sim iid(0, \sigma^2)$ and $y_1 = O_p(1)$, and ρ and γ are unknown parameters. KSS adopt an exponential transition function of the form

$$G(\theta, y_{t-d}) = 1 - \exp(-\theta y_{t-d}^2)$$

where $\theta \geq 0$ determines the degree of nonlinearity. The delay parameter, d , measures the time taken for the transition process to begin reverting an economic variable back to its long run equilibrium. To simplify, we set $d = 1$ throughout this chapter, as KSS do.¹ The exponential transition function is bounded between 0 and 1, i.e. $G : \mathbb{R} \rightarrow [0, 1]$ has the properties

$$G(\theta, 0) = 0, \quad \lim_{y_{t-d} \rightarrow \pm\infty} G(\theta, y_{t-d}) = 1$$

and is symmetrical around zero. The function also implies that for $\theta = 0$ and $\theta \rightarrow +\infty$, y_t reduces to a linear AR(1) process, while nonlinearity is present for $0 < \theta < \infty$.

In what follows, we make the distinction between the *magnitude of the nonlinearity parameter* θ and the *degree of nonlinearity* manifest in the y_t process. Since the model for y_t reduces to a linear process when $\theta = 0$ or $\theta \rightarrow +\infty$, an increase in the magnitude of the nonlinearity parameter does not translate to a monotonically increasing degree of nonlinearity in the process. For very small and very large magnitudes of θ , the degree of nonlinearity will be relatively small, while relatively large degrees of nonlinearity will be associated with intermediate

¹In practical applications, d can be chosen to maximise goodness of fit over $d = \{d_1, \dots, d_{\max}\}$. Norman (2009) shows that the asymptotic distribution of the KSS test is unchanged for $d > 1$, but that using the true value of d , when it is known with certainty, offers almost uniformly higher power than setting $d = 1$.

values of θ .

The exponential STAR (ESTAR) model can then be written as

$$\Delta y_t = \phi y_{t-1} + \gamma y_{t-1} (1 - \exp(-\theta y_{t-1}^2)) + \varepsilon_t$$

where $\phi = \rho - 1$. Imposing $\phi = 0$, i.e.

$$\Delta y_t = \gamma y_{t-1} (1 - \exp(-\theta y_{t-1}^2)) + \varepsilon_t \quad (3.1)$$

implies that y_t follows either a unit root or globally stationary process, and we consider testing the null hypothesis that y_t follows a unit root process, given by $\gamma = 0$ or $\theta = 0$, against the alternative that y_t is nonlinear and globally stationary, i.e. $\theta > 0$, with $-2 < \gamma < 0$ by assumption.²

The hypotheses can therefore be written in terms of the parameter θ

$$H_0 : \theta = 0 \quad H_1 : \theta > 0$$

where γ is a nuisance parameter that appears only under the alternative. As γ is not identified under the null, the hypothesis can't be tested directly (see Davies (1977, 1987) for example). KSS therefore take the first order approximation of the Taylor expansion of the ESTAR model under the null. The Taylor expansion about $\theta = 0$ is given by

$$1 - \exp(-\theta y_{t-1}^2) = \theta y_{t-1}^2 - \frac{1}{2} \theta^2 y_{t-1}^4 + \dots$$

Taking the first order approximation

$$1 - \exp(-\theta y_{t-1}^2) \simeq \theta y_{t-1}^2$$

yields the auxiliary regression

$$\Delta y_t = \delta y_{t-1}^3 + e_t \quad (3.2)$$

KSS then test the hypotheses

$$H_0 : \delta = 0, \quad H_1 : \delta < 0$$

²KSS also consider an extension to the ESTAR model where $\phi > 0$ and y_t is locally explosive and globally stationary under the alternative hypothesis.

using the test statistic

$$KSS = \frac{\hat{\delta}}{s.e.(\hat{\delta})}$$

where $\delta = \gamma\theta$, $\hat{\delta}$ is the OLS estimate of δ from (3.2) and $s.e.(\hat{\delta})$ is the corresponding standard error. The KSS test for a unit root is clearly similar in form to the DF test, with the DF test statistic the t -statistic for testing $\delta = 0$ in the regression

$$\Delta y_t = \delta y_{t-1} + e_t \quad (3.3)$$

as opposed to (3.2).

KSS suggest that a non-zero mean can be accommodated by using the demeaned data $y_{\mu,t} = y_t - \bar{y}$, where $\bar{y} = T^{-1} \sum_{t=1}^T y_t$. We refer to a test using this demeaned data as KSS_{μ} . Both a non-zero mean and a deterministic (linear) trend can be accommodated by using the demeaned and detrended data $y_{\tau,t} = y_t - \hat{\mu} - \hat{\beta}t$ where $\hat{\mu}$ and $\hat{\beta}$ are the OLS estimators from the regression $y_t = \mu + \beta t + \eta_t$, $t = 1, \dots, T$. We refer to this test as KSS_{τ} . Similarly mean-invariant and trend-invariant DF tests are denoted DF_{μ} and DF_{τ} . We focus on these four tests throughout this chapter. Note that to illustrate the results, our analysis assumes that ε_t is not serially correlated. This assumption can be relaxed along the lines of KSS, where additional serial correlation is permitted to enter the model in a linear fashion, provided the usual lag augmentation is applied to the DF and KSS tests, i.e. including lags of Δy_t (or $\Delta y_{\tau,t}$ in the demeaned and detrended case) as additional regressors in (3.2) and (3.3).

3.5 Asymptotic properties of tests

To consider the behaviour of the unit root tests discussed in Section 3.4 when data is generated by an ESTAR process, and to examine the effect that a linear deterministic trend has on the power of these tests, we consider a local to unity alternative form of the ESTAR DGP given in (3.1). DK consider a local alternative by setting $\gamma = \gamma_T = -\frac{c}{T}$ ($c > 0$) in (3.1). This is the natural framework to evaluate local power, since under such a setting, u_t is a local to unit root process, and it is well known that standard DF tests have non-trivial asymptotic power against this type of local alternative, with T^{-1} being the appropriate Pitman drift. However, in the nonlinear model, the alternative DGP is characterized in an additional dimension due to the unknown parameter θ . DK consider three mutually exclusive possibilities for this parameter. First, the limiting case $\theta \rightarrow \infty$ implies that $1 - \exp(-\theta y_{t-1}^2) \rightarrow 1$ and the ESTAR model in 3.1 defaults to a lin-

ear model. The second case considers a fixed $\theta > 0$. It is shown that the limiting behaviour of y_t (normalised by $\sigma\sqrt{T}$) under a fixed nonlinear DGP of this form is identical to that obtained under the linear case. Indeed, although nonlinearity still enters the KSS and DF test statistics through Δy_t , DK demonstrate that the asymptotic distributions of both test statistics will be unaffected by the fixed nonlinearity in the DGP. That is, the limit distributions of the test statistics under a fixed nonlinear alternative are the same as those obtained from under a linear alternative. Motivated by these findings, DK consider a third possibility such that θ converges to zero at an appropriate rate as $T \rightarrow \infty$. In order to prevent the nonlinear ESTAR component from becoming asymptotically negligible, it is shown that this rate is $\theta = O(1/T)$. We adopt this ‘persistently nonlinear approach’ in what follows, with the additional introduction of a scaling by σ^2 , so that the degree of nonlinearity is measured relative to the variance of ε_t , ensuring that σ^2 does not appear in the limit distributions that follow; specifically, we set $\theta = \theta_T = \frac{g^2}{\sigma^2 T}$.

Allowing for a non-zero mean and linear trend, our ESTAR DGP is

$$y_t = \mu + \beta t + u_t, \quad t = 2, \dots, T \quad (3.4)$$

$$\Delta u_t = -\frac{c}{T} u_{t-1} \left(1 - \exp \left(-\frac{g^2}{\sigma^2 T} u_{t-1}^2 \right) \right) + \varepsilon_t \quad (3.5)$$

where $\varepsilon_t \sim iid(0, \sigma^2)$ with finite fourth order moment. We set $u_1 = O_p(1)$, such that the initial condition is assumed to be asymptotically negligible. The exponential transition function can be written as

$$G(\theta_T, u_{t-1}) = 1 - \exp \left(-\frac{g^2}{\sigma^2 T} u_{t-1}^2 \right)$$

so that

$$\Delta u_t = -\frac{c}{T} u_{t-1} G(\theta_T, u_{t-1}) + \varepsilon_t.$$

Following DK, it can be shown that the partial sums of ε_t and u_t have the following limit distributions

$$\left(\frac{1}{\sigma\sqrt{T}} \sum_{j=1}^{\lfloor rT \rfloor} \varepsilon_j ; \frac{1}{\sigma\sqrt{T}} u_{\lfloor rT \rfloor} \right) \xrightarrow{d} (W(r) ; X(r))$$

where $X(r)$ is the diffusion given by the stochastic differential equation

$$dX(r) = -cX(r)\mathcal{G}(g^2 X(r)^2) dr + dW(r) \quad (3.6)$$

with $X(0) = 0$, $W(r)$ a standard Brownian motion process, and

$$\mathcal{G}(g^2 X(r)^2) = 1 - \exp(-g^2 X(r)^2).$$

3.5.1 Asymptotic behaviour under a local trend

We now consider the effect of a local linear trend on the unit root tests discussed in Section 3.4 under both the null hypothesis $H_0 : \gamma = 0$ and the local alternative hypothesis $H_1 : \gamma = -c/T < 0$. We set $\beta = \kappa\sigma T^{-1/2}$, with κ a finite constant, using the appropriate Pitman drift on the trend coefficient. The scaling by σ means that the trend magnitude is measured relative to the standard deviation of the innovations, and following Vogelsang (1998) ensures that this quantity does not appear in the limit expressions. The asymptotic distributions of the four unit root tests: DF_μ , KSS_μ , DF_τ and KSS_τ are given by the following lemma.

Lemma 3.1 *Let y_t be generated according to (3.4)-(3.5) with $\beta = \kappa\sigma T^{-1/2}$. For $c \geq 0$,*

$$\begin{aligned} DF_\mu &\xrightarrow{d} \frac{\int_0^1 (\kappa(r - \frac{1}{2}) + X_\mu(r)) dW(r) - c \int_0^1 (\kappa(r - \frac{1}{2}) + X_\mu(r)) X(r) \mathcal{G}(g^2 X(r)^2) dr}{\sqrt{\int_0^1 (\kappa(r - \frac{1}{2}) + X_\mu(r))^2 dr}} \\ KSS_\mu &\xrightarrow{d} \frac{\kappa \int_0^1 (\kappa(r - \frac{1}{2}) + X_\mu(r))^3 dr + \int_0^1 (\kappa(r - \frac{1}{2}) + X_\mu(r))^3 dW(r) - c \int_0^1 (\kappa(r - \frac{1}{2}) + X_\mu(r))^3 X(r) \mathcal{G}(g^2 X(r)^2) dr}{\sqrt{\int_0^1 (\kappa(r - \frac{1}{2}) + X_\mu(r))^6 dr}} \\ DF_\tau &\xrightarrow{d} \frac{\int_0^1 X_\tau(r) dW(r) - c \int_0^1 X_\tau(r) X(r) \mathcal{G}(g^2 X(r)^2) dr}{\sqrt{\int_0^1 X_\tau(r)^2 dr}} \\ KSS_\tau &\xrightarrow{d} \frac{\int_0^1 X_\tau(r)^3 dW(r) - c \int_0^1 X_\tau(r)^3 X(r) \mathcal{G}(g^2 X(r)^2) dr - 12 \int_0^1 (r - \frac{1}{2}) X(r) dr \int_0^1 X_\tau(r)^3 dr}{\sqrt{\int_0^1 X_\tau(r)^6 dr}} \end{aligned}$$

where $X_\mu(r) = X(r) - \int_0^1 X(s) ds$ and $X_\tau(r) = X(r) + (6r - 4) \int_0^1 X(s) ds + (6 - 12r) \int_0^1 sX(s) ds$.

Proof: See Appendix.

The null limit distributions of the DF_μ , KSS_μ , DF_τ and KSS_τ statistics are

obtained as a special case of Lemma 3.1, on setting $c = 0$, so that $X(r) = W(r)$. Note that when $g^2 = 0$, $\mathcal{G}(g^2 X(s)^2) = 0$, and the limit distributions again reduce to those under the $c = 0$ null. Hence, the null can be viewed as either $c = 0$ or $g^2 = 0$.

The limiting distribution of KSS_τ under the null corrects the corresponding result in Hanck (2012), which in turn corrected the limits given in KSS. Defining $W_\tau(r) = W(r) + (6r - 4) \int_0^1 W(s)ds + (6 - 12r) \int_0^1 sW(s)ds$, Hanck (2012) writes the limit of the numerator as $\int_0^1 W_\tau(r)^3 dW(r)$, whereas the correct expression is $\int_0^1 W_\tau(r)^3 dW_\tau(r)$. Writing the test statistic as

$$KSS_\tau = \frac{\sum_{t=2}^T y_{\tau,t-1}^3 \Delta y_{\tau,t}}{\sqrt{\hat{\sigma}^2 \sum_{t=2}^T y_{\tau,t-1}^6}}$$

we can show that $\Delta y_{\tau,t} = \Delta y_t - \hat{\beta}$ and therefore $dW_\tau(r) \neq dW(r)$. This is the source of discrepancy between the two limits, with Lemma 3.1 providing the correct result.

We observe that the limiting distribution of the demeaned statistics DF_μ and KSS_μ depend on the trend term via κ , whilst the invariant detrended statistics DF_τ and KSS_τ do not. These results coincide with the behaviour of DF_μ and DF_τ reported in Harvey et al. (2009) under a linear AR DGP.

As $g^2 \rightarrow \infty$, $\mathcal{G}(g^2 X(s)^2) \rightarrow 1$, and (3.6) can be rewritten as

$$dX(r) = -cX(r)dr + dW(r)$$

with the numerator of DF_μ becoming

$$\begin{aligned} \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right) dW(r) - c \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right) X(r) dr \\ = \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right) dX(r) \end{aligned}$$

and the numerator of DF_τ becoming

$$\int_0^1 X_\tau(r) dW(r) - c \int_0^1 X_\tau(r) X(r) dr = \int_0^1 X_\tau(r) dX(r)$$

such that, as $g^2 \rightarrow \infty$, the limiting distributions of the two DF tests under an ESTAR alternative will converge to those obtained in Harvey et al. (2009) under a linear local AR alternative.

3.5.2 Asymptotic behaviour under a fixed trend

For completeness, we also consider the behaviour of DF_μ and KSS_μ under a fixed magnitude trend $\beta = \kappa\sigma$ where κ is a finite constant. The limiting distribution of the trend-invariant DF_τ and KSS_τ statistics do not change under a fixed trend. The limiting distributions of DF_μ and KSS_μ under the *iid* assumption for ε_t are given in the following Lemma.

Lemma 3.2 *Let y_t be generated according to (3.4)-(3.5) with $\beta = \kappa\sigma \neq 0$. For $c \geq 0$,*

$$DF_\mu \xrightarrow{d} \frac{\int_0^1 (r - \frac{1}{2}) dW(r) - c \int_0^1 (r - \frac{1}{2}) X(r) \mathcal{G}(g^2 X(r)^2) dr}{\sqrt{(\kappa^2 + 1)/12}}$$

$$KSS_\mu \xrightarrow{d} \frac{3 \int_0^1 (r - \frac{1}{2})^2 X_\mu(r) dr + \int_0^1 (r - \frac{1}{2})^3 dW(r) - c \int_0^1 (r - \frac{1}{2})^3 X(r) \mathcal{G}(g^2 X(r)^2) dr}{\sqrt{(\kappa^2 + 1)/448}}.$$

where the definition of $X_\mu(r)$ is given in 3.1.

Proof: See Appendix.

We note that both DF_μ and KSS_μ possess well-defined limiting distributions that are dependent on the local alternative parameter c . These results correspond with those found by Harvey et al. (2009) in the case of the DF test under a linear AR alternative; indeed, as $g^2 \rightarrow \infty$, the limit of DF_μ reduces to the Harvey et al. (2009) limit. Under the null hypothesis $c = 0$, the DF_μ limit again coincides with the null limit given in Harvey et al. (2009), and from that paper we can state that DF_μ follows an asymptotic normal distribution with mean zero and variance $1/(\kappa^2 + 1)$. This implies that for any non-zero fixed trend, DF_μ will have a null limiting normal distribution with variance less than one, resulting in an asymptotically under-sized test (maximum size can be calculated to be 0.002 as in Harvey et al. (2009)). Similarly, KSS_μ will follow an asymptotic normal distribution under the null, with mean zero and variance approximately $2.32/(\kappa^2 + 1)$ (obtained by simulation). As with DF_μ , this implies an asymptotically under-sized test in the presence of a fixed trend (here, maximum size can be found to be 0.027).

3.6 A union of rejections strategy

3.6.1 Union of DF and KSS

To begin our analysis, we temporarily abstract from uncertainty regarding the presence of a trend, and evaluate the local asymptotic powers of the DF_μ and KSS_μ tests when no trend is present in the DGP ($\kappa = 0$), and subsequently the DF_τ and KSS_τ tests which are invariant to any trend. Asymptotic critical values at a nominal 0.05 significance level were obtained by direct simulation of the limiting distributions given in Lemma 3.1. The $W(r)$ processes were approximated using *iid* standard normal random variables and with the integrals approximated by normalized sums of 2,000 steps using 50,000 Monte Carlo replications. These critical values (and those for 0.10 and 0.01 significance levels) are reported in Table 3.1.

Figure 3.1 shows the local asymptotic powers of both DF_μ and KSS_μ unit root tests for six different settings of the nonlinearity parameter g^2 , when $\kappa = 0$. For each of $g^2 = \{2, 5, 10, 20, 50, 100\}$ we consider values of c such that the power of the tests approaches one. Here and throughout the chapter, asymptotic power is simulated using normalized sums of 1,000 steps and 20,000 Monte Carlo replications. Figure 3.1(a) shows the case for $g^2 = 2$. Here, KSS_μ achieves higher power than DF_μ across all values of c . This highlights the superiority of KSS tests when the degree of nonlinearity in the process is relatively large via the nonlinearity parameter in the ESTAR model being of modest magnitude. For example, when $c = 99$, KSS_μ has power of 0.64 whilst DF_μ has power of 0.49. It is clear in these circumstances that a practitioner would want to employ KSS_μ to test for a unit root. In Figure 3.1(b) where $g^2 = 5$, KSS_μ maintains a power advantage over DF_μ for almost all settings of c , but these power gains are smaller than those seen in Figure 1(a). For example, when $c = 99$, KSS_μ now has power of 0.67 whilst DF_μ has power of 0.56. Figure 3.1(c) shows the power of both tests when $g^2 = 10$. Here, matters are less clear cut, with the power curves intersecting at (approximately) $c = 44$. In Figure 3.1(d), g^2 is increased to 20; the power of both tests is now very similar up until $c = 20$, at which point the power curves diverge with DF_μ outperforming KSS_μ for these higher values of c . Figure 3.1(e) shows the case for $g^2 = 50$; the DF_μ test now displays higher power across almost all values of c . Finally, in Figure 3.1(f), $g^2 = 100$, and DF_μ displays substantially higher power than KSS_μ . For example, when $c = 22$, the power of DF_μ is 0.87 whilst KSS_μ has power of 0.66. These latter figures draw attention to the power gains DF tests can have against KSS tests for large magnitudes of g^2 ; recall that as $g^2 \rightarrow \infty$, the ESTAR process reduces to a linear process, hence, compared to

modest values of g^2 , the degree of nonlinearity is reduced for large g^2 , and it is not surprising that DF power gains become apparent in these cases.

Figure 3.2 shows the asymptotic powers of both DF_τ and KSS_τ unit root tests for the same six settings of the nonlinearity parameter g^2 and the same range of values for the local alternative parameter c . Given that both tests are trend-invariant, the setting of κ used in these simulations is irrelevant. Figure 3.2(a), where $g^2 = 2$, shows that, for modest magnitudes of the nonlinearity parameter, KSS_τ has power gains over DF_τ , due to a high degree of nonlinearity in the process. For example, for $c = 171$, KSS_τ has power of 0.61 and DF_τ has power of 0.51. We note that these power gains are smaller than those seen in the demeaned case in Figure 3.1. As we would expect, for the same values of c , both detrended tests have less power than their demeaned counterparts. In Figure 3.2(b), where $g^2 = 5$, KSS_τ still maintains a power advantage over DF_τ , but these gains are diminished compared to the previous figure. Figures 3.2(c) to 3.2(f) show that as the value of g^2 increases, the relative power performance of DF_τ to KSS_τ improves, to the point where DF_τ outperforms KSS_τ , in line with the degree of nonlinearity reducing. For example, for $g^2 = 100$ and $c = 30$, the power of DF_τ is 0.90 and the power of KSS_τ is 0.67. This mirrors the results of the demeaned tests.

The results from Figures 1 and 2 show that the relative powers of the DF and KSS tests are very sensitive to the degree of nonlinearity under the ESTAR alternative, with the KSS test offering greater power for modest magnitudes of the nonlinearity parameter, and the DF test displaying relatively higher power for larger magnitudes. Given that the degree of nonlinearity is almost certainly unknown in practical applications, these results highlight the potential benefits of a testing strategy that combines inference from the DF and KSS tests. We therefore want to employ a testing strategy that can capitalize on the respective power advantages of the DF and KSS tests, regardless of the degree of nonlinearity under the ESTAR alternative. In line with the suggestion of DK, we now consider two union of rejections strategies, cf. Harvey et al. (2009), one that combines DF_μ and KSS_μ (for the case of no trend), and one that combines DF_τ and KSS_τ (for the trend case).

The union of rejections approach is a simple decision rule where the null hypothesis of a unit root is rejected if either of the individual tests reject. We can write our proposed union of rejections strategies as

$$U_i : \text{Reject } H_0 \text{ if } DF_i < \lambda_i^\zeta cv_{DF_i}^\zeta \text{ or } KSS_i < \lambda_i^\zeta cv_{KSS_i}^\zeta, \quad i = \mu, \tau$$

where $cv_{DF_i}^\zeta$ and $cv_{KSS_i}^\zeta$ denote the asymptotic null critical values of DF_i and KSS_i respectively for a significance level ζ . Note that if we simply rejected the null hypothesis whenever either DF_i or KSS_i is smaller than its respective (unscaled) critical value, this would result in an over-sized procedure, hence we incorporate the scaling constant λ_i^ζ , calculated such that the asymptotic empirical size of U_i equals the nominal size ζ . The decision rule can also be written as

$$U_i : \text{Reject } H_0 \text{ if } t_{U_i} = \min \left(DF_i, \frac{cv_{DF_i}^\zeta}{cv_{KSS_i}^\zeta} KSS_i \right) < \lambda_i^\zeta cv_{DF_i}^\zeta, \quad i = \mu, \tau.$$

The limiting distribution of t_{U_i} , under a local trend, is then given by a function of the limiting distributions derived in Lemma 3.1

$$t_{U_i} \xrightarrow{d} \min \left(L_{DF_i}, \frac{cv_{DF_i}^\gamma}{cv_{KSS_i}^\gamma} L_{KSS_i} \right), \quad i = \mu, \tau$$

where L_{DF_i} and L_{KSS_i} denote the asymptotic distributions of DF_i and KSS_i , respectively, given in Lemma 3.1.

To obtain the appropriate value for the scaling constant λ_i^ζ , we can simulate the limit distribution of t_{U_i} and calculate the ζ level critical value $cv_{U_i}^\zeta$. Computing $\lambda_i^\zeta = cv_{U_i}^\zeta / cv_{DF_i}^\zeta$ will then give the value for the scaling constant that ensures U_i is asymptotically correctly sized. Asymptotic scaling constants for nominal 0.10, 0.05 and 0.01 significance levels are given in Table 3.1.

Figure 3.1 also displays the power performance of U_μ for $g^2 = \{2, 5, 10, 20, 50, 100\}$. Figure 3.1(a) shows asymptotic power of U_μ for $g^2 = 2$. We see that the power of U_μ dominates that of DF_μ for all values of c , with its power curve lying just underneath that of KSS_μ for approximately $c \leq 108$, and with power rising above both KSS_μ and DF_μ beyond this point. U_μ is therefore able to capture the power advantage offered by KSS_μ in this small nonlinearity parameter case. For $g^2 = 5$ in Figure 3.1(b), we see that, again, U_μ outperforms DF_μ across all values of c , and offers power that is very slightly below that of KSS_μ for lower values of c , and slightly above at higher values of c . In Figure 3.1(c), where $g^2 = 10$, the power curve of U_μ closely tracks that of whichever test, KSS_μ or DF_μ , is superior for a given c . In Figures 3.1(d) - 3.1(f), as the value of g^2 increases, the power of U_μ now dominates that of KSS_μ and closely tracks that of DF_μ for all values of c . U_μ therefore provides power near the effective envelope offered by KSS_μ and DF_μ across all values of the nonlinearity parameter g^2 .

Figure 3.2 shows the power performance of U_τ for the same six settings of g^2 . We observe a very similar pattern of behaviour, relative to the DF and KSS tests, to that seen in Figure 3.1. In Figures 3.2(a) and 3.2(b), the power of U_τ

dominates that of DF_τ and closely tracks that of KSS_τ . In Figure 3.2(c), as the power curves of the two individual tests intersect, we see that the power of U_τ is marginally higher than both of these tests for most values of c . As the nonlinearity parameter, g^2 , increases in Figures 3.2(d) - 3.2(f), the power of U_τ now dominates that of KSS_τ for all values of c , whilst closely tracking the power curve of DF_τ .

3.6.2 Union of demeaned, and demeaned and detrended tests

The above results clearly illustrate the benefit of employing a union of rejections procedure in practice where the degree of nonlinearity is unknown. However, we have so far assumed that it is known whether or not the DGP contains a linear deterministic trend, so that a practitioner would know whether to apply U_μ or U_τ . In practice, there will be uncertainty about the presence or otherwise of a trend in many economic time series. In Section 3.5, we showed that the limiting distributions of DF_μ and KSS_μ , under both local and fixed trends, are not invariant to the trend, suggesting that the power of these tests may be compromised if a linear trend is present in the DGP. Whilst this might suggest that a conservative strategy of always using U_τ should be employed, Harvey et al. (2009) show that, in the context of a linear AR DGP, DF_μ has a substantial power advantage over DF_τ when no trend is present in the DGP, and might also be expected to offer some power gains for very small trend magnitudes. We therefore investigate the power performance of U_μ and U_τ when a local trend is present in the DGP.

Figures 3.3 - 3.5 show the asymptotic power of U_μ and U_τ for six different settings of the local trend coefficient, $\kappa = \{0, 0.25, 0.5, 1, 2, 4\}$ and three different settings of the nonlinearity parameter $g^2 = \{2, 10, 50\}$. In Figure 3.3, we set $g^2 = 2$; Figure 3.3(a) shows the power of U_μ and U_τ when $\kappa = 0$ such that no trend is present in the DGP. We see that U_μ significantly outperforms U_τ across all values of c . For example, when $c = 135$, U_μ has power of 0.83 whilst U_τ has power of 0.50. This highlights the potential losses involved in always employing the trend-invariant U_τ when there is uncertainty about the presence or otherwise of a linear trend. In Figure 3.3(b), $\kappa = 0.25$; the presence of this linear trend has decreased the power of U_μ such that, for $c = 135$, power has dropped to 0.56, only marginally higher than that of U_τ . When the local trend coefficient increases to $\kappa = 0.5$ in Figure 3.3(c), the power of U_τ now dominates that of U_μ . Figures 3.3(d) - 3.3(f) show that as the trend magnitude continues to increase, the power

of U_μ deteriorates dramatically, such that for $\kappa = 1$ in Figure 3.3(d), U_μ has power lower than size for most c , and for $\kappa = 2$ and $\kappa = 4$ in Figures 3.3(e) - 3.3(f), the size and power of U_μ approach zero. Almost identical patterns of behaviour are observed for $g^2 = 10$ and $g^2 = 50$ in Figures 3.4 and 3.5 respectively. These results demonstrate the potential losses that a practitioner can make if they fail to choose the appropriate deterministic specification for unit root testing.

Given that, in practice, the presence or otherwise of a linear deterministic trend will often be unknown, we propose a further union of rejections procedure that combines information from U_μ and U_τ by taking a union of rejections of all four individual tests DF_μ , KSS_μ , DF_τ and KSS_τ . This proposed testing strategy can then be written as

$$U_4 : \quad \text{Reject } H_0 \text{ if } DF_\mu < \lambda_4^\zeta cv_{DF_\mu}^\zeta \text{ or } KSS_\mu < \lambda_4^\zeta cv_{KSS_\mu}^\zeta \text{ or } DF_\tau < \lambda_4^\zeta cv_{DF_\tau}^\zeta \\ \text{or } KSS_\tau < \lambda_4^\zeta cv_{KSS_\tau}^\zeta$$

or

$$U_4 : \text{Reject } H_0 \text{ if } t_{U_4} = \min \left(DF_\mu, \frac{cv_{DF_\mu}^\zeta}{cv_{KSS_\mu}^\zeta} KSS_\mu, \frac{cv_{DF_\mu}^\zeta}{cv_{DF_\tau}^\zeta} DF_\tau, \frac{cv_{DF_\mu}^\zeta}{cv_{KSS_\tau}^\zeta} KSS_\tau \right) < \lambda_4^\zeta cv_{DF_\mu}^\zeta$$

The limiting distribution of t_{U_4} , under a local trend, is then given by

$$t_{U_4} \xrightarrow{d} \min \left(L_{DF_\mu}, \frac{cv_{DF_\mu}^\zeta}{cv_{KSS_\mu}^\zeta} L_{KSS_\mu}, \frac{cv_{DF_\mu}^\zeta}{cv_{DF_\tau}^\zeta} L_{DF_\tau}, \frac{cv_{DF_\mu}^\zeta}{cv_{KSS_\tau}^\zeta} L_{KSS_\tau} \right).$$

Values of the scaling constant λ_4^ζ required to control the asymptotic size of U_4 are given in Table 3.1 for 0.10, 0.05 and 0.01 significance levels. We note that the value of λ_4^ζ is larger than that of λ_μ^ζ and λ_τ^ζ as the Bonferroni bound of the combination of four individual tests will be twice as high as that of the combination of two individual tests. Consequently a larger scaling of the critical values is required to ensure correct asymptotic size of U_4 .

Figures 3.3 - 3.5 display asymptotic power results for our proposed union, U_4 , for the same three settings of g^2 and six settings of κ used previously. Figure 3.3(a) displays power in the case where $g^2 = 2$ and no trend is present in the DGP ($\kappa = 0$). The power of U_4 dominates that of U_τ across all values of c . For example, when $c = 135$, the power of U_4 is 0.72, whilst the power of U_τ is 0.50. U_4 tracks the power of U_μ , with power losses of, at most, 0.13 when $c = 117$. In Figure 3.3(b), $\kappa = 0.25$, and the power curve of U_4 sits slightly above U_τ and slightly below U_μ for all values of c . As the trend coefficient increases in Figures

3.3(c) - 3.3(f), U_4 outperforms U_μ for all values of c , whilst tracking underneath the power curve of U_τ . For example, when $\kappa = 4$, in Figure 3.3(f), U_4 has a maximum power loss of 0.21 relative to U_τ , but a power gain of 0.49 relative to U_μ . A similar pattern of results emerges when considering $g^2 = 10$ in Figure 3.4 and $g^2 = 50$ in Figure 3.5. Across all settings of κ , U_4 has superior power performance to the worst-performing test of either U_μ and U_τ whilst tracking underneath the power of the best-performing test. These results demonstrate that U_4 provides a strategy for unit root testing against ESTAR alternatives which offers decent power levels across different trend magnitudes, including the case of no trend, as well as across the degree of nonlinearity under the ESTAR alternative.

3.6.3 Asymptotic power under fixed c

Our analysis so far has characterised the local alternative in terms of c , and considered a number of fixed settings of the nonlinearity parameter g^2 . This highlighted the power advantage that KSS tests could have over DF tests for certain degrees of nonlinearity. Of course, it is also possible to hold c fixed and evaluate power across g^2 . Figure 3.6 shows the local asymptotic powers of DF_μ , KSS_μ and U_μ unit root tests for six different settings of c when $\kappa = 0$. For each of $c = \{20, 40, 60, 80, 100, 150\}$ we consider values of g^2 such that the power of the tests approaches one (or the maximum power available for each c setting). For all six c settings, we note that the KSS_μ and DF_μ power curves intersect. For lower values of g^2 , KSS_μ dominates DF_μ whilst DF_μ has a power advantage over KSS_μ for higher values of g^2 . The specific value of g^2 at which this intersection occurs differs depending on the value of c . As c increases, the process becomes less persistent and both DF_μ and KSS_μ will have high power detecting a stationary alternative for smaller values of $g^2 > 0$. We see that the power curve of U_μ is either above that of the two individual tests, or tracking just below that of the better-performing test across all values of g^2 .

Local asymptotic powers of DF_τ , KSS_τ and U_τ are given in Figure 3.7 for the same settings of c and g^2 . Again, we note the intersection of the DF_τ and KSS_τ power curves as g^2 increases (although this is less evident in figures 3.7(c) and 3.7(f) due to the curves intersecting only at the largest values of g^2 we consider for those settings of c). Again, the union of rejections procedure U_τ is shown to capitalise on the respective power advantages of both individual tests across all g^2 and c .

Figures 3.8 - 3.10 display the local asymptotic power of U_μ , U_τ and U_4 for

$c = \{20, 60, 100\}$ respectively. For each value of c we again consider six different settings of the local trend coefficient, $\kappa = \{0, 0.25, 0.5, 1, 2, 4\}$. These graphs confirm our earlier findings that U_μ outperforms U_τ for lower settings of the trend coefficient, but that as the size of the coefficient increases, the power of U_μ deteriorates such that U_τ begins to dominate. U_4 is again shown to always outperform the worst-performing of either U_μ or U_τ whilst closely tracking the power of the better-performing test. These results confirm the findings of the previous two subsections regarding the role of the nonlinearity and trend parameters in affecting the power performance of DF and KSS tests. The same inference is made regardless of whether we evaluate power over fixed values of c or g^2 .

3.7 Trend detection

A potential drawback of combining all four individual tests is that as the number of tests within the union of rejections procedure increases, the asymptotic scaling value, λ_ζ also increases to ensure the overall procedure is asymptotically correctly sized. As λ_ζ increases, so too does the chance that the union of rejections procedure will fail to reject the null hypothesis of a unit root even if one or more of the individual tests is able to reject at the same level of significance. In a related setting, Harvey et al. (2012) note that use of information regarding the presence of a trend can be applied to reduce the number of tests entering the union of rejections in certain cases, reducing the extent of the required critical value scaling. Given that the asymptotic size and power of U_μ decreases as the trend magnitude increases, it follows that the DF_μ and KSS_μ tests are redundant in U_4 for large trends. This implies that U_4 is correctly sized for $\beta = 0$ but under-sized elsewhere, resulting in a loss of power. Therefore, for large enough values of the trend, we want to eliminate U_μ from our testing strategy and use U_τ rather than U_4 .

Following in the spirit of Harvey et al. (2012), we propose using a hybrid testing strategy where a BIC procedure is used to detect the presence of a trend. If evidence of a trend is apparent, we can be confident that a trend of reasonably large magnitude is present in the data, and hence that use of U_τ is appropriate, with little to be gained by also including DF_μ and KSS_μ in the union of rejections. However, if no trend is detected by the BIC procedure, it would be unwise to confidently infer that no trend exists and use only U_μ , since a small magnitude trend could still be present in the data (which could have a substantial impact on the power of U_μ), but its magnitude not sufficient to have triggered detection by the BIC procedure. Consequently, if no trend is detected by the BIC, we default to

the risk-averse approach of using U_4 , retaining the DF_τ and KSS_τ tests as well as the DF_μ and KSS_μ tests in the union of rejections. This approach therefore uses the information from the BIC in an indicative manner, rather than as a standard model selection procedure, consistent with the approach of Harvey et al. (2012).

Our procedure is devised as follows. Consider the null DGP

$$\begin{aligned} y_t &= \zeta + \beta t + u_t \\ \Delta u_t &= \varepsilon_t. \end{aligned}$$

If $\beta = 0$ such that no trend is present, then $\Delta y_t = \varepsilon_t$, while if $\beta \neq 0$, then $\Delta y_t = \beta + \varepsilon_t$. Treating these two possibilities as regression models, we can select between them on the basis of the BIC. Specifically, consider the two corresponding residual series

$$\begin{aligned} \hat{\varepsilon}_{1t} &= \Delta y_t \\ \hat{\varepsilon}_{2t} &= \Delta y_t - \frac{\sum_{s=2}^T \Delta y_s}{T-1} = \Delta y_t - \frac{y_T - y_1}{T-1} \end{aligned}$$

The BIC for each model is then

$$\begin{aligned} BIC_1 &= \ln \left(\frac{\sum_{t=1}^{T-1} \hat{\varepsilon}_{1t}^2}{T-1} \right) \\ BIC_2 &= \ln \left(\frac{\sum_{t=1}^{T-1} \hat{\varepsilon}_{2t}^2}{T-1} \right) + \frac{\ln(T-1)}{T-1} \end{aligned}$$

If $BIC_2 < BIC_1$ then there is evidence that a trend is present in y_t , and therefore U_τ should be applied in place of U_4 . We can write this testing procedure formally as

$$U^* = \begin{cases} U_\tau & \text{if } BIC_2 < BIC_1 \\ U_4 & \text{otherwise} \end{cases}$$

When no trend is present in the DGP, U^* will select U_4 with probability one in the limit. When a fixed magnitude trend is present, it is U_τ that will be selected asymptotically with probability one. Hence U^* will be asymptotically correctly sized under no trend or a fixed trend, and would be expected to achieve greater power than simply using U_4 whenever a trend of large magnitude is present in the data. In the next section we consider the finite sample performance of U^* to quantify the potential gains that this approach can offer.

3.8 Finite sample simulations

In this section we consider the extent to which asymptotic power is a good indicator of finite sample performance by simulating the four individual test statistics DF_μ , KSS_μ , DF_τ and KSS_τ as well as the three union of rejections procedures U_μ , U_τ and U_4 using a sample of $T = 100$ observations. In addition, we examine the finite sample power of our trend detection-based U^* procedure. We use the same values of g^2 and c used in our asymptotic simulations. As before, we use 20,000 Monte Carlo simulations and the asymptotic critical values and scaling constants given in Table 3.1. Finite sample sizes of the four individual tests and four union of rejection procedures are given in Table 3.2 for $\kappa = 0$. None of the procedures display much in the way of size distortions, with empirical size close to nominal size for all significance levels.

Figures 3.11 - 3.15 report finite sample size and power at a nominal 0.05 significance level. First, we can observe the desirable size-control of the tests, with the exception of under-sizing for U_μ in the presence of large trends (due to under-sizing in DF_μ and KSS_μ). Next, we find that the power results obtained are qualitatively very similar from those observed in the asymptotic case. Figure 3.11 shows results for DF_μ , KSS_μ and U_μ . We see that for smaller degrees of nonlinearity, i.e. smaller values of g^2 , KSS_μ has a power advantage over DF_μ ; see, for example, Figure 3.11(a). However, for larger values of g^2 , DF_μ is shown to have superior power to KSS_μ , for example in Figures 3.11(d) - 3.11(f). As in the asymptotic case, the power of U_μ dominates that of the worst-performing test for all settings of g^2 and all values of c . It either closely tracks or has slightly higher power than the best-performing test across all values of g^2 and c . These results are mirrored in Figure 3.12 when finite sample power results for DF_τ , KSS_τ and U_τ are considered. Again, we see that KSS_τ has higher power relative to DF_τ for lower settings of g^2 , but that the relative power ranking is reversed for higher settings of g^2 . We show that U_τ has power either close to or above the power offered by the best-performing of either DF_τ or KSS_τ for all values of c and g^2 .

Figures 3.13 - 3.15 report finite sample power results of U_μ , U_τ and U_4 for three settings of the mean reversion parameter, $g^2 = \{2, 10, 50\}$. We set $\beta = \{0, 0.025, 0.05, 0.1, 0.2, 0.4\}$ such that for $T = 100$ these values coincide with the local trend values considered in the previous asymptotic analysis. The behaviour of U_μ , U_τ and U_4 is very similar to that observed in the asymptotic simulations. We see that the power of U_μ decreases as β increases, such that in Figure 3.13(a), where $g^2 = 2$ and $\beta = 0$, U_μ has a substantial power advantage over U_τ , but for higher values of β the power of U_μ tends to zero. U_4 is able to capitalize on

the differing power profiles of U_μ and U_τ and captures much of the power gains associated with U_μ for small trends, and U_τ for large trends. An almost identical pattern of behaviour is seen for $g^2 = 10$ in Figure 3.14 and $g^2 = 50$ in Figure 3.15.

Figures 3.13 - 3.15 also report results for the U^* procedure which selects between U_τ and U_4 on the basis of the BIC trend detection procedure. The power of U^* equals that of U_4 for smaller trend magnitudes, i.e. $\beta \leq 0.1$, as here the trend is not detected by the BIC. In Figure 3.13(e), we see that when $\beta = 0.2$, U^* is selecting U_τ in some replications due to detection of a trend by the BIC, and correspondingly, the power of U^* is marginally higher than that of U_4 . In Figure 3.13(f), when $\beta = 0.4$, the trend is now detectable in all replications, and U^* always selects U_τ , so that the power of U^* equals that of U_τ , and therefore exceeds that of U_4 by a substantial margin. The same pattern of results is seen in Figures 3.14 and 3.15, where higher values of g^2 are considered. These results show that the overall power performance of U_4 can be improved by incorporating a trend detection procedure. In our simulations, U^* never has lower power than U_4 , but when a large trend is present in the data, U^* achieves higher power than U_4 by defaulting to U_τ .

3.9 Empirical example

To demonstrate the practical use of our proposed union of rejection procedures, we undertake an empirical application of the unit root tests considered in this chapter to primary energy consumption data. Economists are keen to understand whether shocks to energy consumption have permanent or transitory effects. If energy consumption follows a unit root process, then shocks to world energy markets have a permanent effect. This is likely to have implications for policymakers who wish to set government targets relating to energy consumption. Additionally, an understanding of the statistical properties of energy consumption is required in order to reliably model the relationship between energy consumption and other macroeconomic variables. In recent years, a growing number of studies have been dedicated to understanding the integration properties of energy consumption. For a review of this literature see Smyth (2013).

The concept of nonlinearities in both energy consumption and production dates back to the peak oil literature, beginning with Hubbert (1956), who successfully predicted that oil production in the United States would reach a peak in the mid-1960s to 1970, and decline thereafter. This nonlinear behaviour in oil markets has been seen to hold in other countries, for example in both the United

Kingdom and Russia, crude oil production has exhibited several peaks. Maslyuk and Smyth (2009) suggest that this nonlinearity is exacerbated by uncertainty in oil markets affected by factors such as temperature, wind speed and technology. Additionally, as new sources of energy are found (for example, the discovery of new oil fields) energy prices will drop having an effect on demand which is reflected by an increase in production. Thus energy markets in general are subject to peaks of production and consumption, inciting nonlinearities.

In a recent study, Hasanov and Telatar (2011) apply the four individual unit root tests DF_μ , DF_τ , KSS_μ and KSS_τ to annual total primary energy consumption per capita data obtained for 178 countries across the time period 1980-2006.³ At a 0.10 significance level they are able to reject the null hypothesis of a unit root for 55 out of the 178 countries using DF_μ and DF_τ tests and for 71 countries using KSS_μ and KSS_τ tests. Combining results from these different test procedures, they suggest that the null hypothesis of a unit root can be rejected in many of the countries considered. Implicitly, this strategy of testing for a unit root is equivalent to a union of rejections procedure that does not apply a scaling constant to the critical values of the individual tests. As discussed in Section 3.6, without this scaling constant the union of rejections procedure will be over-sized. Caution must therefore be exercised over any inference made from the combination of results provided by these individual tests, as the high number of rejections could be due to over-sizing.

We obtain total primary energy consumption per capita data from the Energy Information Administration as Hasanov and Telatar (2011) do. Our sample covers 180 countries and a slightly longer time series from 1980-2011. We apply the four individual unit root tests, DF_μ , DF_τ , KSS_μ and KSS_τ , as well as the four union of rejections procedures considered in this chapter, U_μ , U_τ , U_4 and U^* to this data.

Serial correlation in the data is accounted for by augmenting the test regressions with lagged differences of the dependent variable. Given the small number of observations available per country, we want to minimise any size concerns that might arise with using asymptotic critical values. Su et al. (2013) show that in finite samples the KSS_μ and KSS_τ unit root tests suffer from smaller size distortions when the MAIC procedure of Ng and Perron (2001) is used to select the optimal number of lags rather than the “General to Specific” [GS] sequential downwards testing method of Ng and Perron (1995). To confirm that these size properties hold for our union procedure, we simulate the size of U_4 under MAIC

³Hasanov and Telatar (2011) also apply the ST-TAR unit root test of Sollis (2004) that allows for an asymmetric transition function and a gradual trend break in the data.

and GS lag selection by generating the unit root null

$$u_t = u_{t-1} + \varepsilon_t, \quad t = 2, \dots, T$$

where the errors ε_t follow either an AR(1) process

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t$$

or an MA(1) process

$$\varepsilon_t = \eta v_{t-1} + v_t$$

where $v_t \sim iidN(0, 1)$. We set $\rho = \{-0.5, -0.9, 0, 0.5, 0.9\}$ in the AR(1) model and $\eta = \{-0.5, -0.9, 0.5, 0.9\}$ in the MA(1) model and consider samples sizes of $T = \{30, 50, 100\}$. Size simulations are obtained for both the GS and MAIC lag selection methods with a maximum possible lag length of 6 at a 0.05 significance level. As before, we use 20,000 Monte Carlo simulations and the asymptotic critical values and scaling constants given in Table 3.1. Results displayed in Table 3.3 show that for $T = 30$, the most relevant sample size in the context of this empirical example, when data is generated by an AR(1) process, the size of U_4 ranges from 0.019 to 0.076 when using MAIC, and from 0.048 to 0.158 when GS is employed. As expected, both lag selection methods result in some degree of size distortion when using asymptotic critical values. Given the greater degree of over-sizing seen when employing GS lag selection, we prefer the more conservative approach of MAIC. When data is generated by an MA process, a similar pattern of over-sizing vs under-sizing occurs for positive values of the MA coefficient. Of course, for large, negative values of the MA coefficient, we have near-cancelling factors, and severe over-sizing is exhibited using either method of lag selection at $T = 30$. These results hold for $T = 50$ and $T = 100$ with MAIC leaning towards under-sizing and GS to over-sizing across most values of the AR and MA coefficient, although the size distortions in both directions decrease as T increases. Given this broad pattern of results, we select the optimal number of lagged difference terms using MAIC in our empirical example, with a maximum lag order of 6.

Results from our unit root testing exercise are given in Table 3.4 and 3.5. For the 60 countries where at least one of the four individual tests rejects at a 0.10 significance level, the countries are listed in Table 3.4, detailing the significance levels at which rejections are obtained by the different tests. Table 3.5 lists the remaining countries for which we failed to reject the null hypothesis of a unit root using any of the four individual tests. We note that of the four individual test

procedures, KSS_τ is able to reject the null hypothesis of a unit root more often than the other procedures. Indeed, at a 0.10 level of significance the difference between the number of rejections found by KSS_τ (40) is substantially higher than that found by the next most frequently rejecting test, KSS_μ (29). The pattern of rejections found by these four individual tests allows us to demonstrate the advantages of our union of rejections procedures. For example, a rejection of the null is found for Kenya (see Figure 3.16(c)) for a plot of this series) at a 0.05 significance level using DF_μ . No other individual test rejects, but U_μ picks up this rejection at a 0.05 level, and U_4 and U^* are able to reject at a 0.10 level. For Panama (see Figure 3.16(d)), DF_τ is the only individual test to reject, but U_τ , U_4 and U^* also pick up this rejection, all at a 0.01 significance level. A 0.01 level rejection for the Cook Islands (see Figure 3.16(b)) by KSS_μ results in a 0.01 level rejection by U_μ and 0.05 level rejections by U_4 and U^* . Finally, a 0.01 level rejection found for Portugal (see Figure 3.16(e)) by only KSS_τ is picked up by the three relevant union procedures at the same significance level. We can also highlight countries where rejections are found only in the two demeaned tests (Suriname, see Figure 3.16(f)), or only in the two detrended tests (Bangladesh, see Figure 3.16(a)). Again, these rejections are then picked up by the relevant union procedures.

It is also to be expected from the simulation results that there will be cases where the union of rejections procedures fail to pick up some rejections found by individual tests, and we can see examples of this in the results, for example there are a number of countries for which one of the individual tests rejects at the 0.10 level, but U^* does not reject at conventional significance levels. There is also one case (Bhutan), where U^* fails to reject when U_4 rejects at the 0.01 level; here, the BIC is suggesting the presence of a trend, but the rejection is lost since it only arises from application of KSS_μ . Overall, however, the results clearly demonstrate the ability of the union of rejections procedures to capitalize on the differing rejections offered by the four individual unit root tests. In total, the union procedure U_μ is able to reject the null for a higher proportion of countries than the individual test DF_μ at all conventional significance levels, and rejects a higher proportion of countries than KSS_μ at a 0.05 significance level. Similarly, U_τ rejects the null more frequently than either DF_τ or KSS_τ at all conventional levels of significance. Our proposed U_4 procedure is able to further capitalize on the differing rejections offered by demeaned and detrended unit root tests, and rejects for a higher proportion of countries than U_μ at all conventional levels of significance and U_τ at a 0.10 level. Finally, our BIC-based strategy, U^* , rejects the same number of times as U_4 and U_τ at a 0.01 level of significance and a

higher proportion of the time at both 0.05 and 0.10 levels. Therefore we conclude that the highest number of rejections across all 8 unit root testing procedures considered in this exercise was achieved by U^* at 0.05 and 0.10 levels and by U^* , U_4 and U_τ at a 0.01 level. This provides clear justification for the use of our proposed union of rejections procedures in empirical applications. We can also conclude that, compared with Hasanov and Telatar (2011), at a 0.10 significance level we find evidence of stationarity in, at most, 45 countries out of 180. This suggests that nonstationarity in energy consumption may be less prevalent than previously thought.

3.10 Conclusion

In this chapter we have examined the power performance of DF and KSS unit root tests against an alternative hypothesis of ESTAR stationarity. Our analysis focused on prior demeaned, and demeaned and detrended test statistics. We highlighted the sensitivity of these tests to both the degree of nonlinearity under the ESTAR alternative and to the presence or otherwise of a linear deterministic trend. In practice, these are likely to be sources of uncertainty in unit root testing. We therefore proposed four union of rejections strategies that attempt to mitigate against this uncertainty while capitalizing on the power available from each individual DF and KSS test under different degrees of nonlinearity and trend magnitudes. Both asymptotic and finite sample power simulations were undertaken, and we found that a union of rejections procedure that combines all four individual tests offers attractive levels of power across different nonlinearity and trend settings. Also, a hybrid approach that uses information from a BIC approach to trend detection to select between this four-way union of rejections and a union of rejections based on just the trend based DF and KSS variants, was found to improve power still further when the magnitude of the trend is large. An empirical example using energy consumption data highlighted the potential usefulness of this testing strategy in practice.

This chapter's focus is on determining the order of integration of a time series, where we have no prior knowledge of whether the process would be best characterised by linear or nonlinear stationarity under the alternative. Of course, if the series is found to be stationary, a practitioner might want to know whether that mean reverting behaviour is linear or nonlinear. There exist several tests in the literature, for example that of Terasvirta (1994), which are able to distinguish between linearity and ESTAR nonlinearity, conditional on the order of integration being known. Should these tests suggest a process is defined by ESTAR

stationarity, the magnitude of the nonlinearity parameter could be estimated by fitting an ESTAR process to the data. The findings of this paper suggest that rejections of the unit root null will arise from the two KSS tests for moderate values of the nonlinearity parameter, and from the two DF tests for large values of the nonlinearity parameter. However, the union of rejections testing strategy enables us to abstract from any concern over the size of this parameter when our main objective is simply to determine the integration properties of the data.

Table 3.1: Asymptotic critical values of KSS_μ , DF_μ , KSS_τ , DF_τ and union of rejection scaling values λ_μ^ζ , λ_τ^ζ and λ_4^ζ for significance level ζ

ζ	KSS_μ	DF_μ	KSS_τ	DF_τ	λ_μ^ζ	λ_τ^ζ	λ_4^ζ
0.10	-2.655	-2.564	-3.118	-3.122	1.063	1.056	1.140
0.05	-2.935	-2.864	-3.396	-3.409	1.059	1.049	1.118
0.01	-3.471	-3.424	-3.939	-3.959	1.053	1.041	1.085

Table 3.2: Finite sample sizes of KSS_μ , DF_μ , KSS_τ , DF_τ and union of rejection procedures U_μ , U_τ , U_4 and U^* for significance level ζ and $T = 100$

ζ	KSS_μ	DF_μ	KSS_τ	DF_τ	U_μ	U_τ	U_4	U^*
0.10	0.092	0.104	0.087	0.102	0.103	0.098	0.100	0.101
0.05	0.044	0.055	0.043	0.053	0.051	0.051	0.052	0.053
0.01	0.009	0.011	0.009	0.013	0.010	0.012	0.012	0.012

Table 3.3: Finite sample sizes of U_4 with GS and $MAIC$ lag selection methods for $\zeta = 0.05$

	AR			MA		
T	30	50	100	30	50	100
	$\rho = 0$					
GS	0.106	0.095	0.079			
$MAIC$	0.047	0.039	0.040			
	$\rho = 0.5$			$\eta = 0.5$		
GS	0.110	0.094	0.074	0.116	0.111	0.089
$MAIC$	0.019	0.020	0.034	0.017	0.016	0.030
	$\rho = 0.9$			$\eta = 0.9$		
GS	0.116	0.094	0.069	0.096	0.076	0.055
$MAIC$	0.045	0.042	0.037	0.014	0.014	0.017
	$\rho = -0.5$			$\eta = -0.5$		
GS	0.158	0.127	0.070	0.280	0.328	0.194
$MAIC$	0.076	0.040	0.036	0.219	0.116	0.061
	$\rho = -0.9$			$\eta = -0.9$		
GS	0.048	0.045	0.047	0.498	0.662	0.750
$MAIC$	0.028	0.025	0.030	0.629	0.613	0.549

Table 3.4: Total primary energy consumption per capita: DF_μ , DF_τ , KSS_μ , KSS_τ , U_μ , U_τ , U_4 and U^* unit root test results

	DF_μ	DF_τ	KSS_μ	KSS_τ	U_μ	U_τ	U_4	U^*
Algeria			***	***	***	***	***	***
American Samoa		**	***	***	**	***	***	***
Antigua and Barbuda		***				***	***	***
Argentina				*		*		
Bahamas	*							
Bahrain				*				
Bangladesh		***		**		***	**	***
Bermuda	*		*		*			
Bhutan			***		***		***	
Burkina Faso		*						
Cambodia				**		**	**	**
Cayman Islands			*					
Central African Republic	**	***		*	**	***	***	***
Chad			*		*			
Comoros		*						
Cook Islands			***		***		**	**
Djibouti			*	*				
Dominica		*	*	**		**	*	**
Egypt		**	*	*	*	**	*	**
El Salvador				**		*		*
Faroe Islands	*		***	*	***	*	***	***
Finland			*		*			
French Polynesia	***	**	***	**	***	**	***	***
Greenland			***	***	***	***	***	***
Grenada				**		**	*	*
Guadeloupe		***	***		***	***	***	***
Guinea			**	***	*	***	**	**
Honduras		*		***		***	**	**
Hong Kong		**				**	*	**
India				**		*		*
Jordan			**	*	**		*	*
Kenya	**				**		*	*
Kuwait		**				**	**	**
Lebanon				*				
Lesotho				**		**	*	*
Libya		***	***	***	***	***	***	***

	DF_μ	DF_τ	KSS_μ	KSS_τ	U_μ	U_τ	U_4	U^*
Maldives				**		**	*	**
Mali	***	*	*		**		**	**
Mozambique			*	**	*	**	**	**
Nauru			**	*	**	*	**	**
Netherlands Antilles	***	**	**	***	***	***	***	***
New Caledonia	**	*	**		**	*	*	*
Niger				*				
Panama		***				***	***	***
Portugal				***		***	***	***
Puerto Rico			**	**	*	*	*	*
Saint Lucia		**		***		***	***	***
Samoa		***		***		***	***	***
Sao Tome and Principe				*				
Senegal			**	**	*	**	*	*
Seychelles				**		**	*	*
Sierra Leone				***		***	***	***
Solomon Islands		***		***		***	***	***
Suriname	**		*		**		**	**
Swaziland	***	**	***	*	**	*	**	**
Tonga			***	***	***	**	**	**
Trinidad and Tobago				**		*		*
Uganda		***				***	***	***
United Arab Emirates	***	**	***	***	***	***	***	***
Uruguay		*						
Total Rejections at 0.10	12	25	29	40	28	41	43	45
Total Rejections at 0.05	9	18	19	28	20	32	31	34
Total Rejections at 0.01	5	9	12	14	11	19	19	19

*, **, *** indicate rejections at a 0.10, 0.05, 0.01 significance level respectively

Table 3.5: Total primary energy consumption per capita: countries for which no unit root test rejections found

Afghanistan	Gibraltar	Pakistan
Albania	Greece	Papua New Guinea
Angola	Guam	Paraguay
Australia	Guatemala	Peru
Austria	Guinea-Bissau	Philippines
Barbados	Haiti	Poland
Belgium	Hungary	Qatar
Belize	Iceland	Reunion
Benin	Indonesia	Romania
Bolivia	Iran	Rwanda
Botswana	Iraq	Saint Kitts and Nevis
Brazil	Ireland	Saint Pierre and Miquelon
British Virgin Islands	Israel	Saint Vincent/Grenadines
Brunei	Italy	Saudi Arabia
Bulgaria	Jamaica	Singapore
Burma (Myanmar)	Japan	Somalia
Burundi	Kiribati	South Africa
Cameroon	Laos	South Korea
Canada	Liberia	Spain
Cape Verde	Luxembourg	Sri Lanka
Chile	Macau	Sudan and South Sudan
China	Madagascar	Sweden
Colombia	Malawi	Switzerland
Congo (Brazzaville)	Malaysia	Syria
Congo (Kinshasa)	Malta	Taiwan
Costa Rica	Martinique	Tanzania
Cote d'Ivoire	Mauritania	Thailand
Cuba	Mauritius	Togo
Cyprus	Mexico	Tunisia
Denmark	Mongolia	Turkey
Dominican Republic	Montserrat	United Kingdom
Ecuador	Morocco	United States
Equatorial Guinea	Nepal	United States Virgin Islands
Ethiopia	Netherlands	Vanuatu
Fiji	New Zealand	Venezuela
France	Nicaragua	Vietnam
French Guiana	Nigeria	Western Sahara
Gabon	North Korea	Yemen
Gambia	Norway	Zambia
Ghana	Oman	Zimbabwe

Figure 3.1: Asymptotic size and local power of KSS_μ , DF_μ and U_μ for fixed g^2 ($\kappa = 0$)

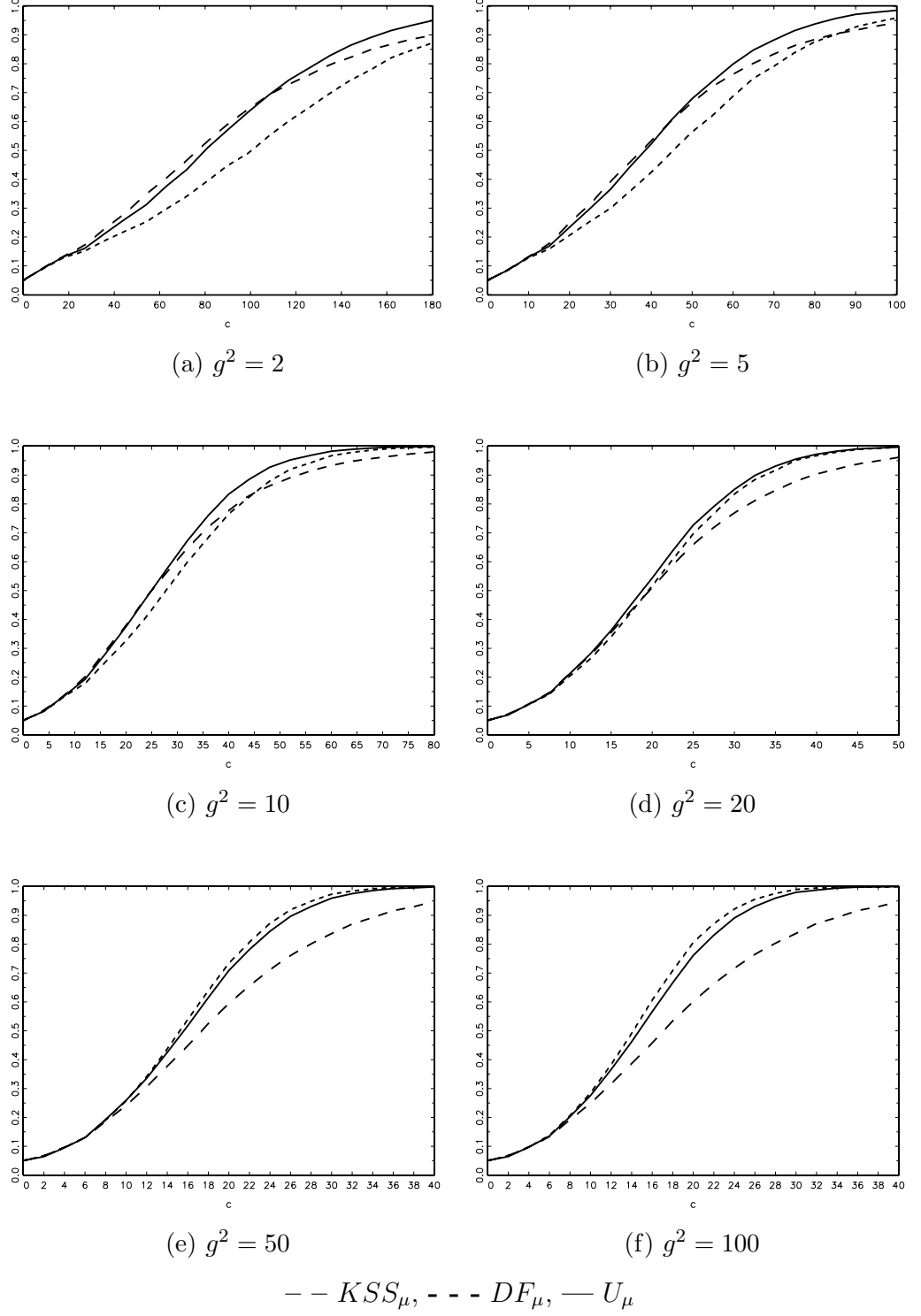
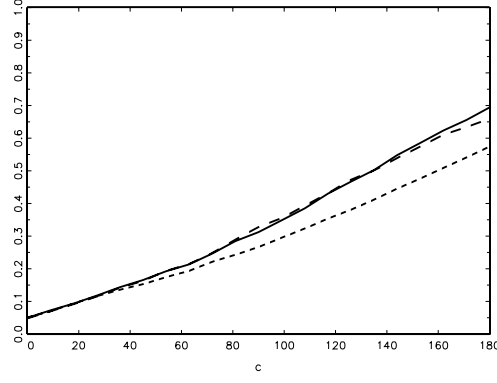
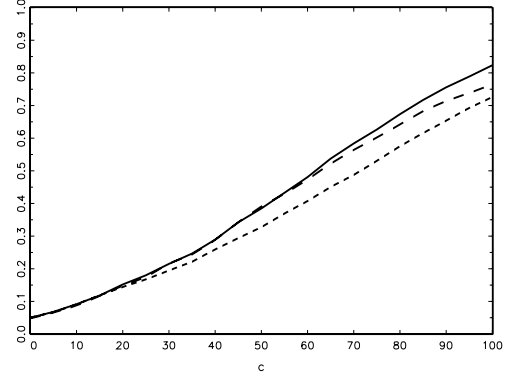


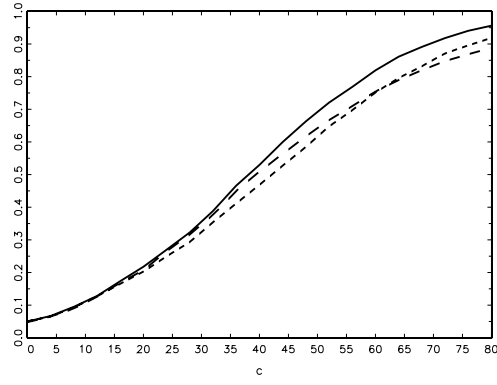
Figure 3.2: Asymptotic size and local power of KSS_τ , DF_τ and U_τ for fixed g^2



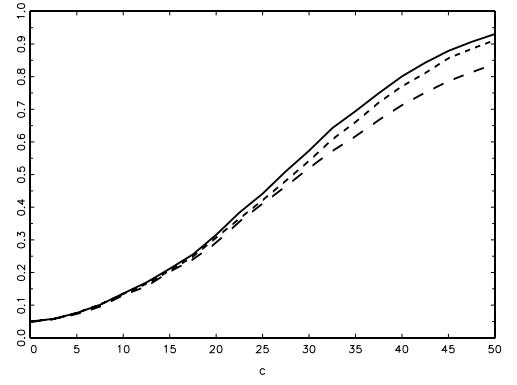
(a) $g^2 = 2$



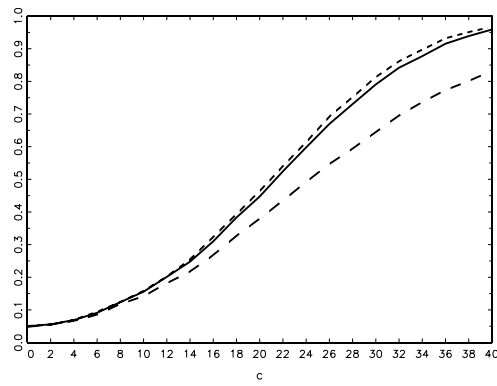
(b) $g^2 = 5$



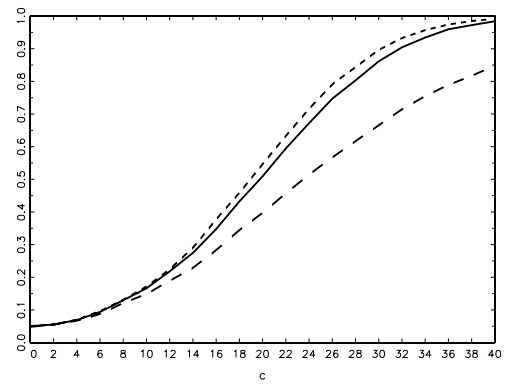
(c) $g^2 = 10$



(d) $g^2 = 20$



(e) $g^2 = 50$



(f) $g^2 = 100$

— KSS_τ , - - - DF_τ , — U_τ

Figure 3.3: Asymptotic size and local power of U_μ , U_τ and U_4 for fixed $g^2 = 2$

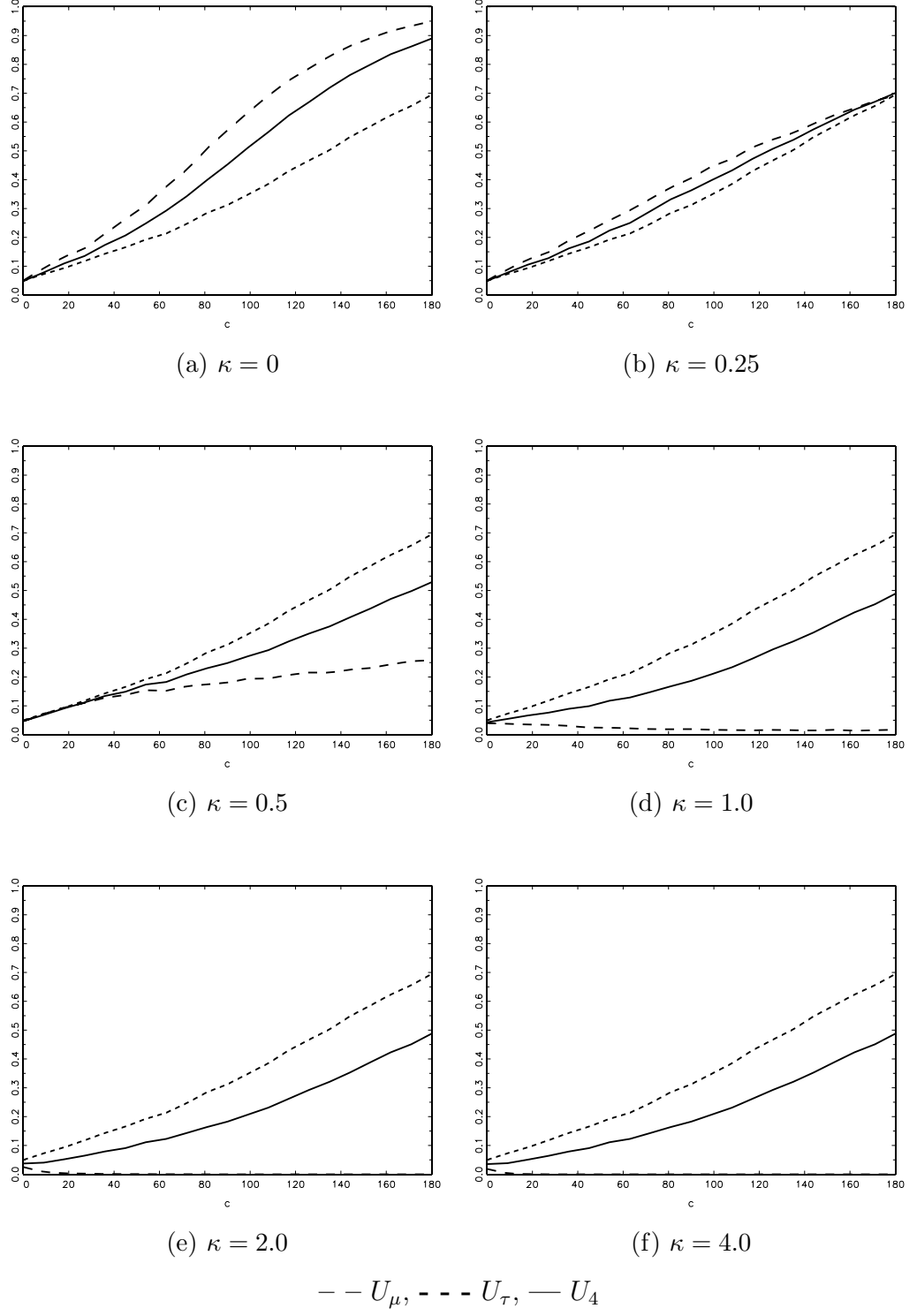


Figure 3.4: Asymptotic size and local power of U_μ , U_τ and U_4 for fixed $g^2 = 10$

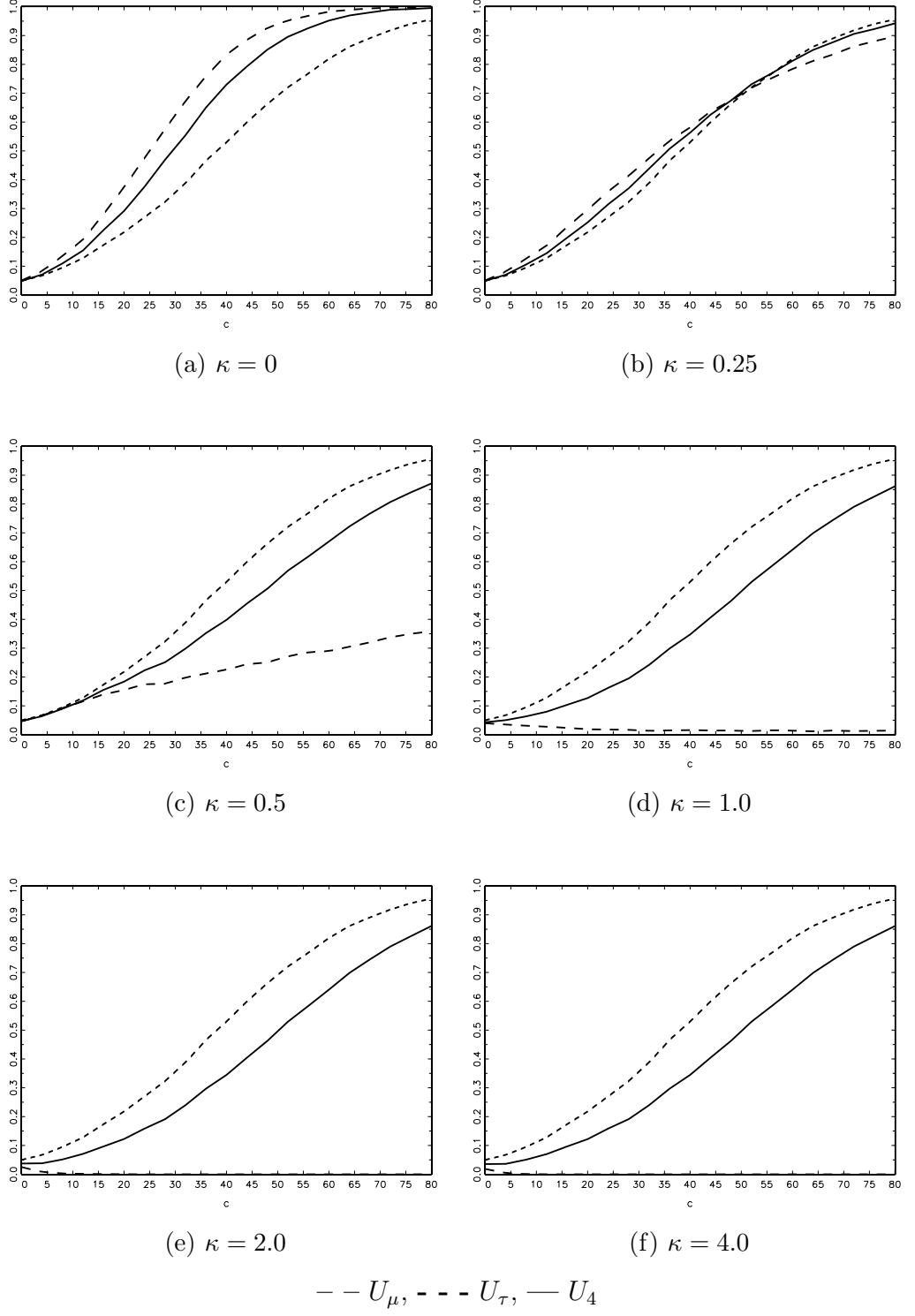
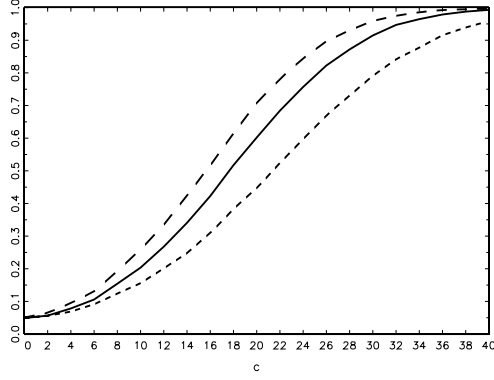
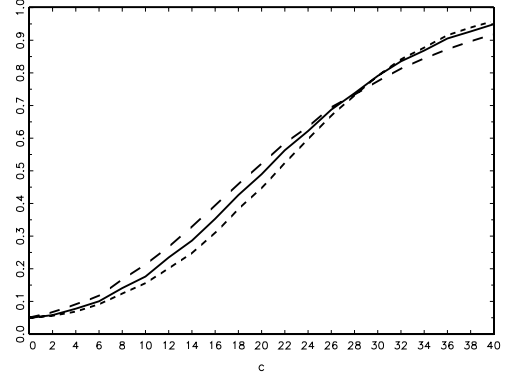


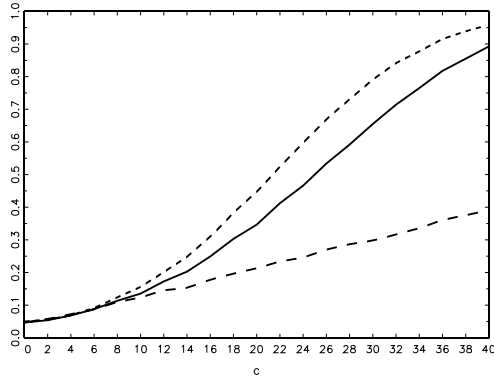
Figure 3.5: Asymptotic size and local power of U_μ , U_τ and U_4 for fixed $g^2 = 50$



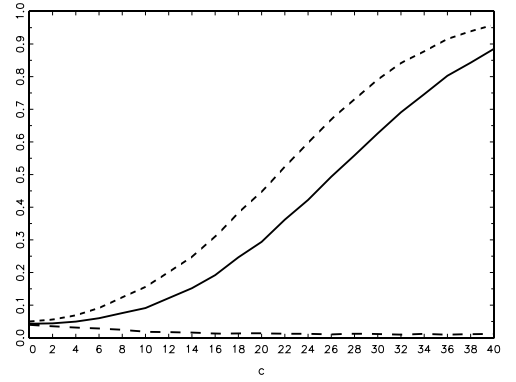
(a) $\kappa = 0$



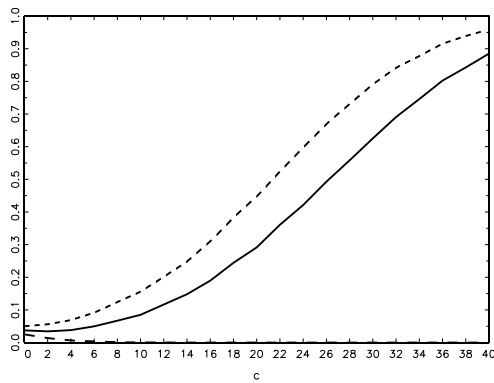
(b) $\kappa = 0.25$



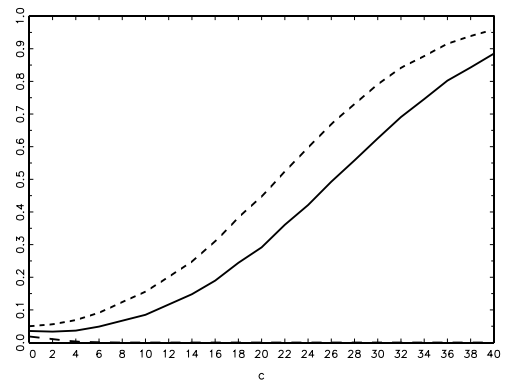
(c) $\kappa = 0.5$



(d) $\kappa = 1.0$



(e) $\kappa = 2.0$



(f) $\kappa = 4.0$

— U_μ , - - - U_τ , — U_4

Figure 3.6: Asymptotic size and local power of KSS_μ , DF_μ and U_μ for fixed c ($\kappa = 0$)

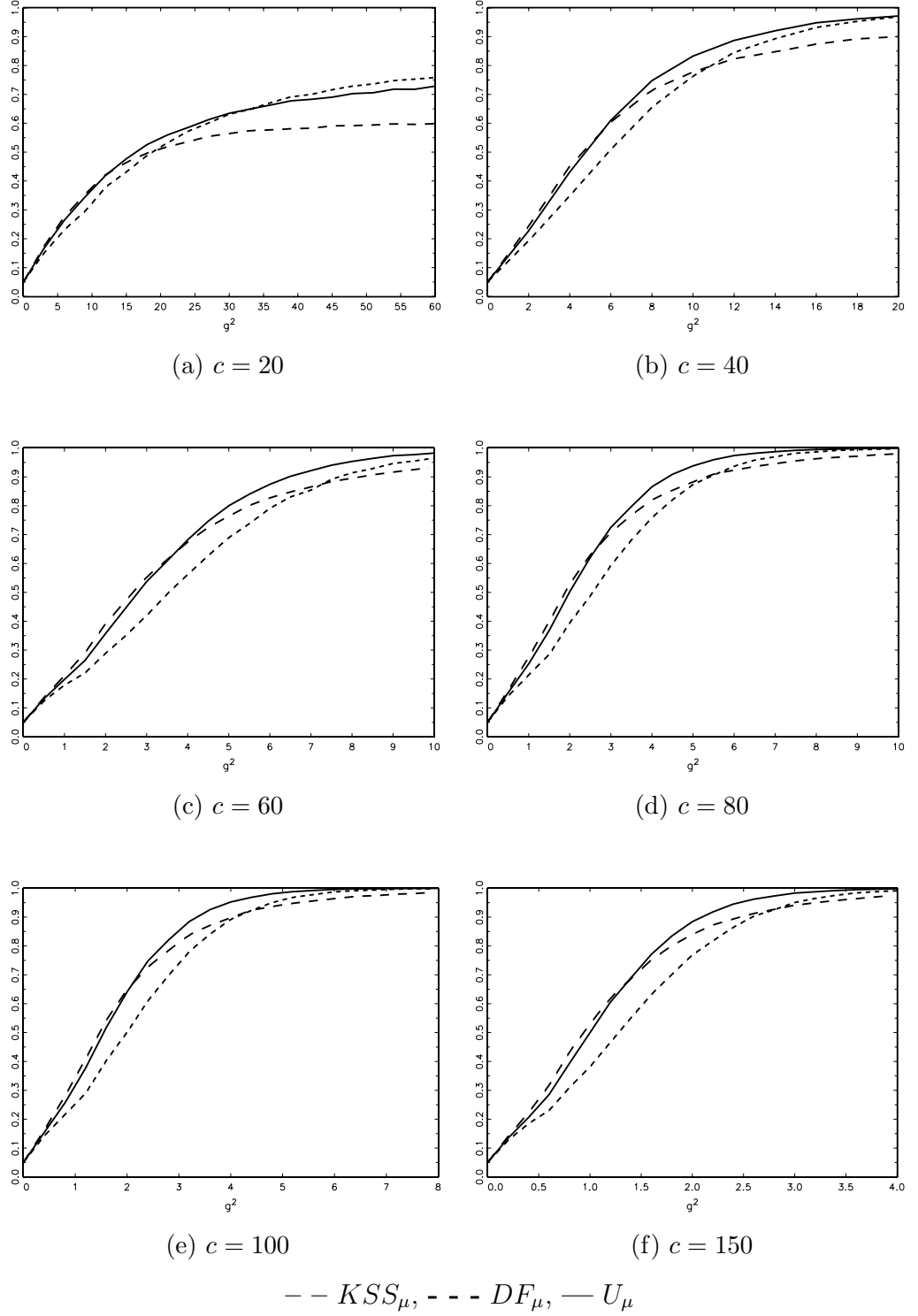
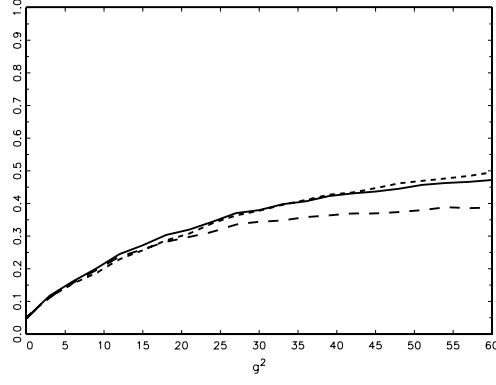
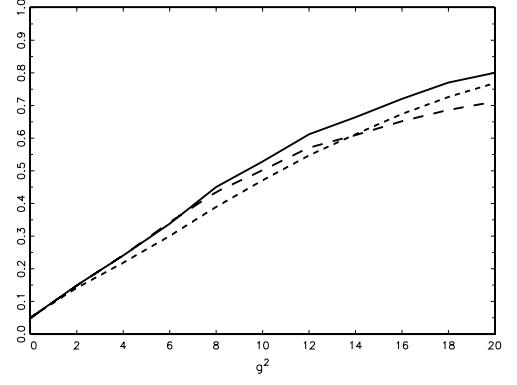


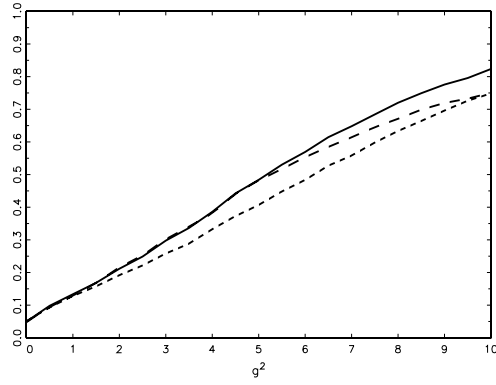
Figure 3.7: Asymptotic size and local power of KSS_τ , DF_τ and U_τ for fixed c



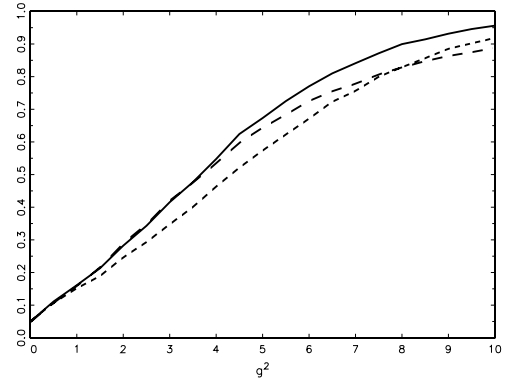
(a) $c = 20$



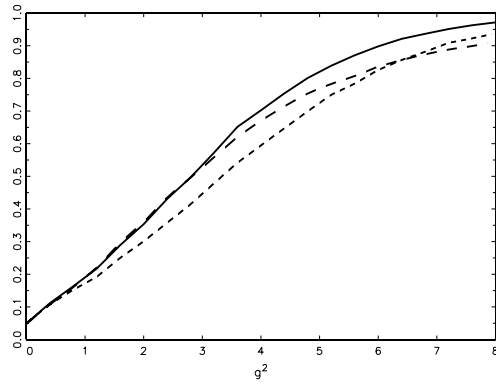
(b) $c = 40$



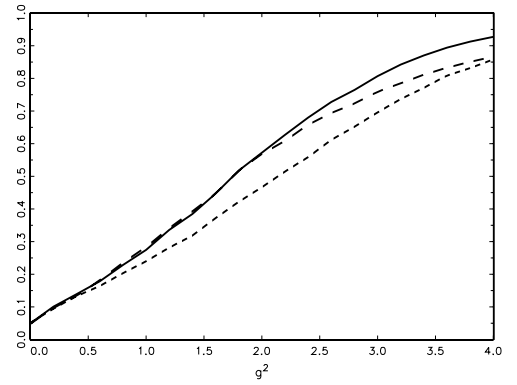
(c) $c = 60$



(d) $c = 80$



(e) $c = 100$



(f) $c = 150$

— KSS_τ , - - - DF_τ , — U_τ

Figure 3.8: Asymptotic size and local power of U_μ , U_τ and U_4 for fixed $c = 20$

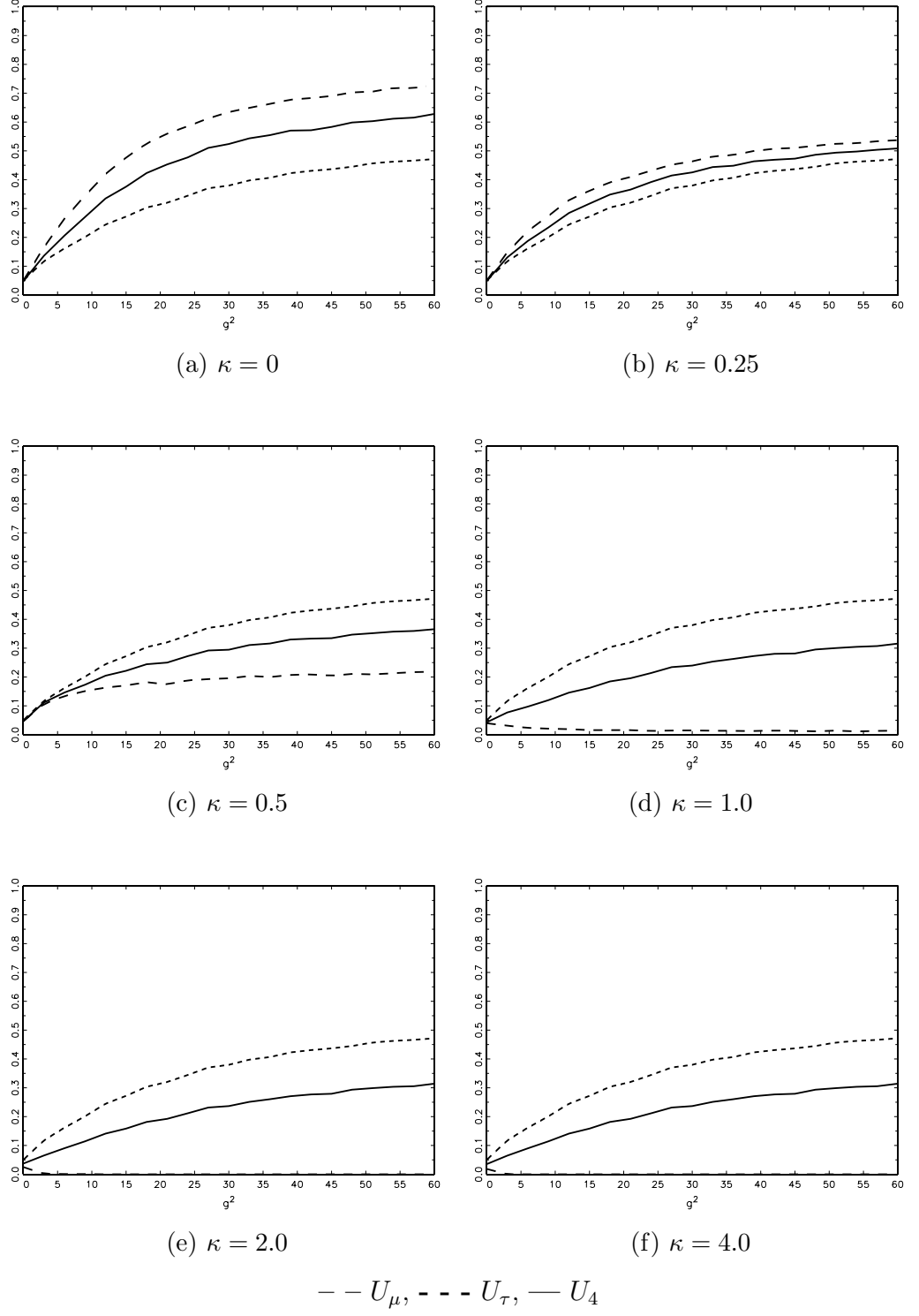


Figure 3.9: Asymptotic size and local power of U_μ , U_τ and U_4 for fixed $c = 60$

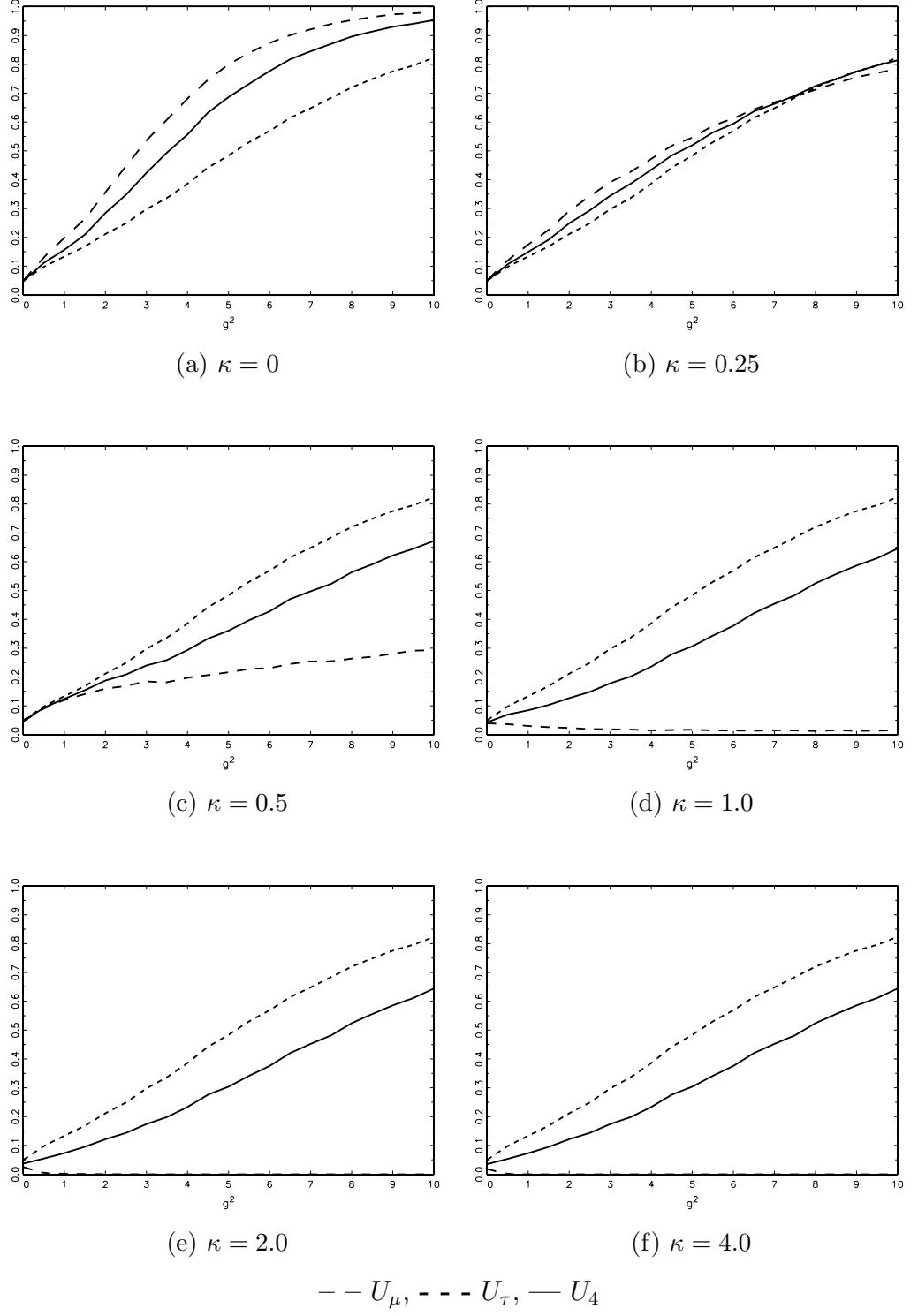


Figure 3.10: Asymptotic size and local power of U_μ , U_τ and U_4 for fixed $c = 100$

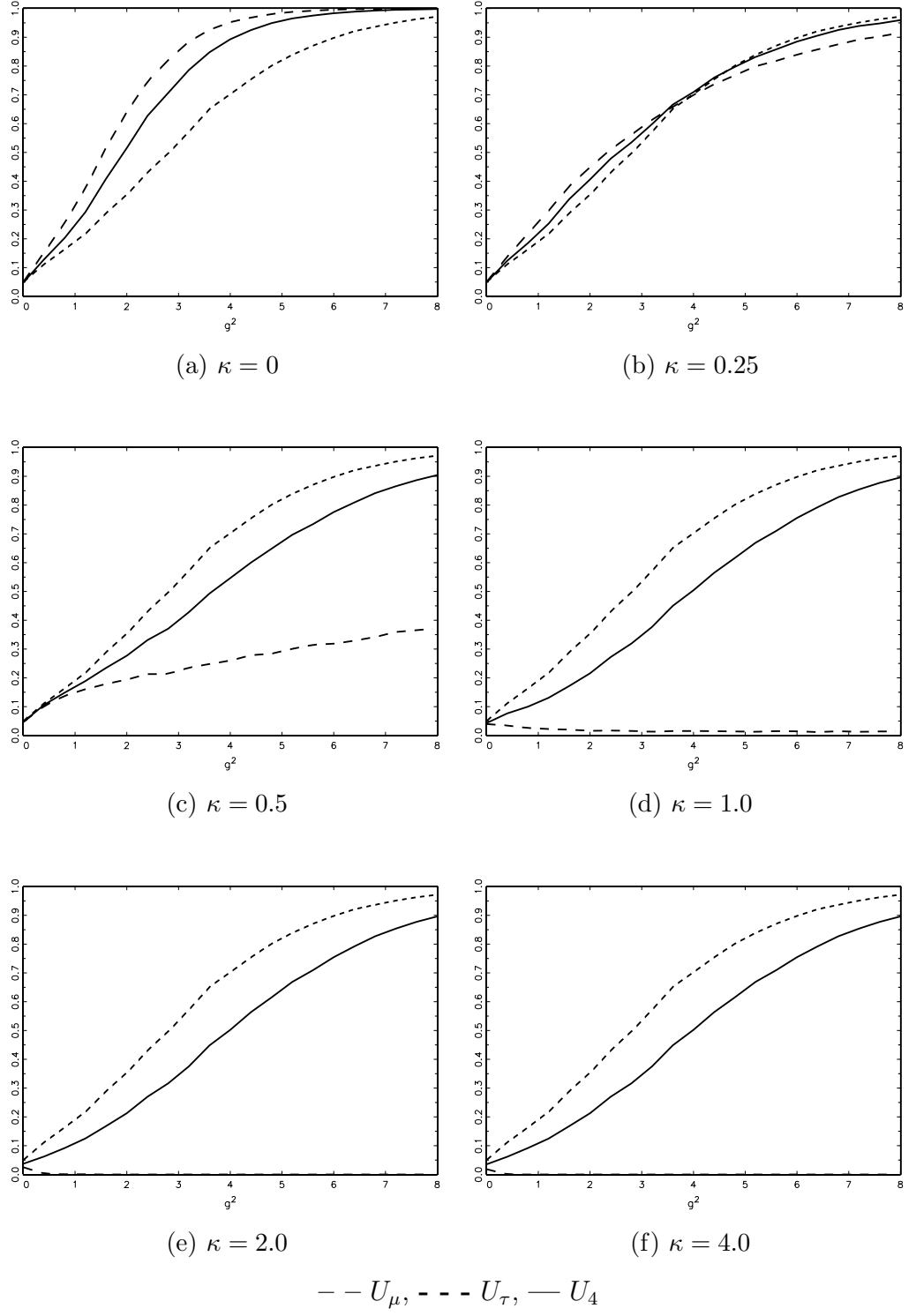


Figure 3.11: Finite sample size and local power of KSS_μ , DF_μ and U_μ for fixed g^2 ($\beta = 0$)

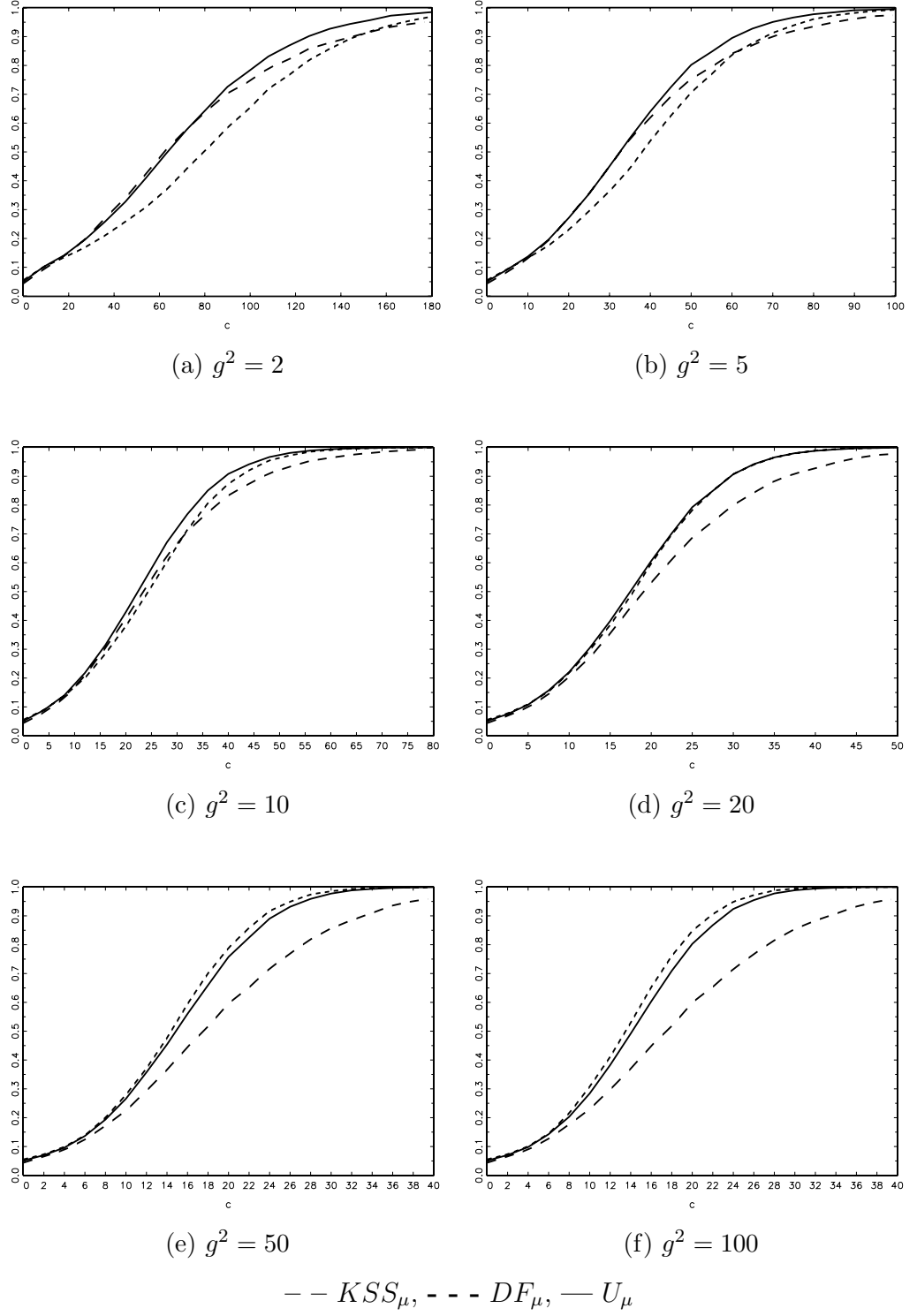


Figure 3.12: Finite sample size and local power of KSS_τ , DF_τ and U_τ for fixed g^2

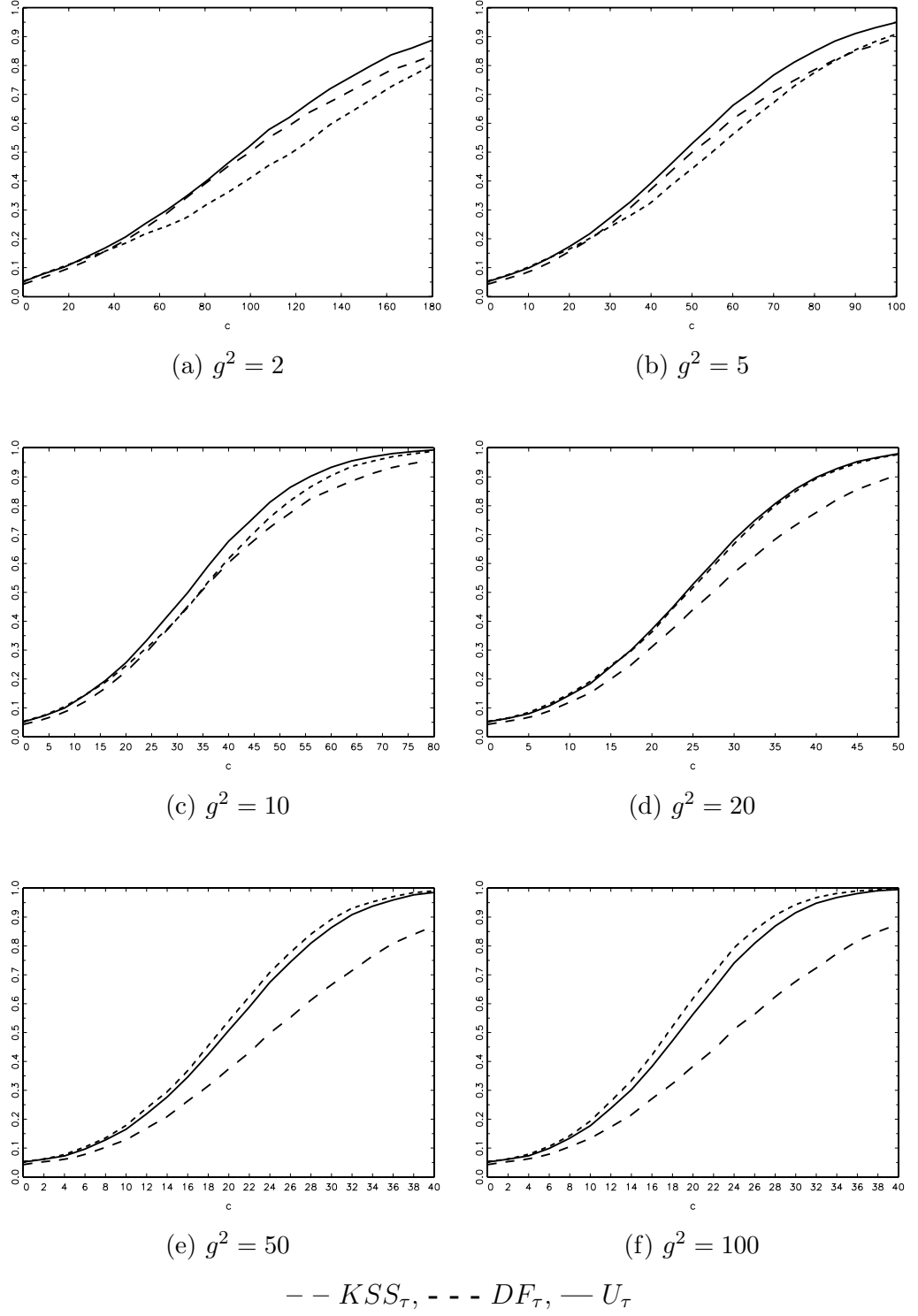


Figure 3.13: Finite sample size and local power of U_μ , U_τ , U_4 and U^* for fixed $g^2 = 2$

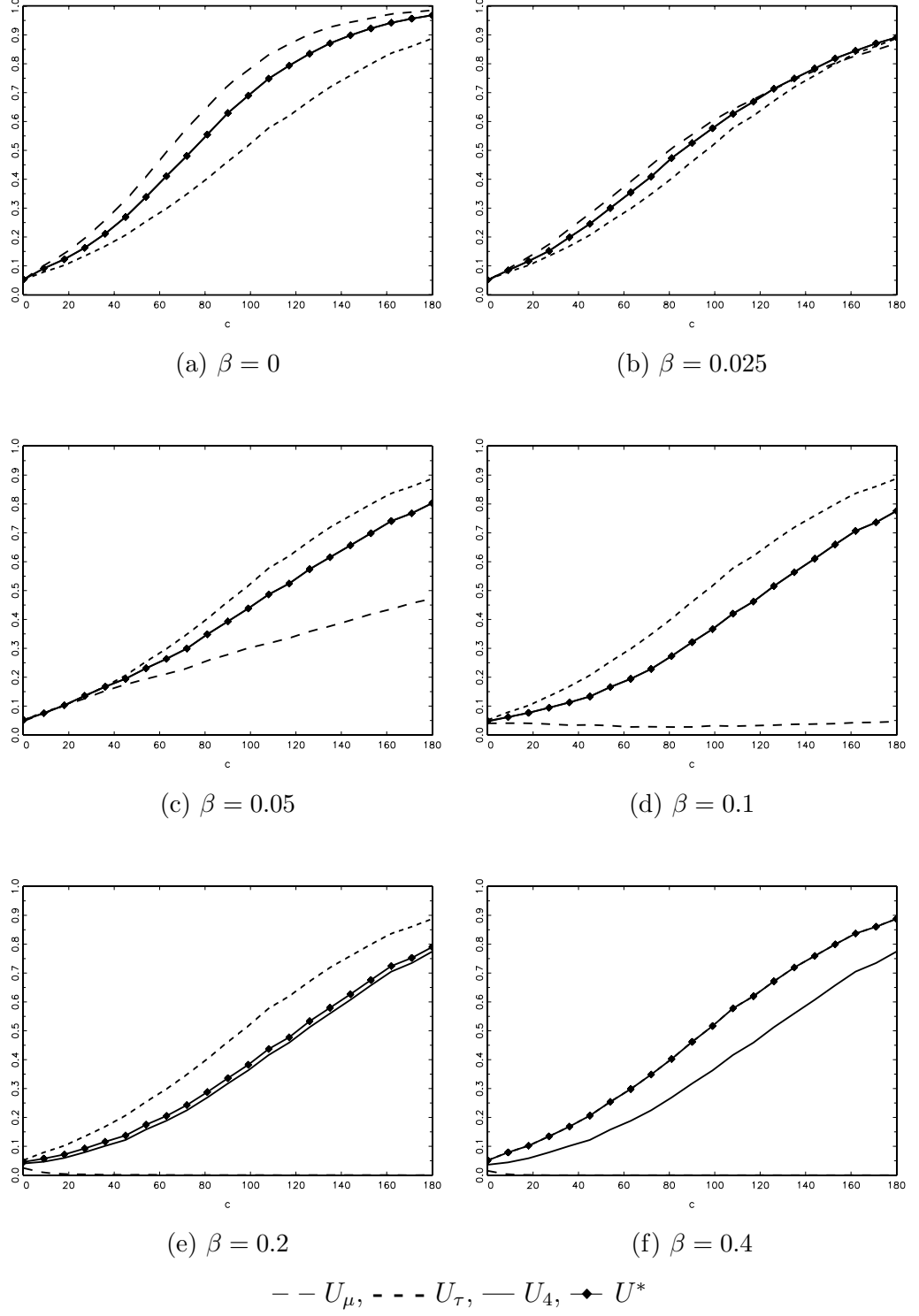


Figure 3.14: Finite sample size and local power of U_μ , U_τ , U_4 and U^* for fixed $g^2 = 10$

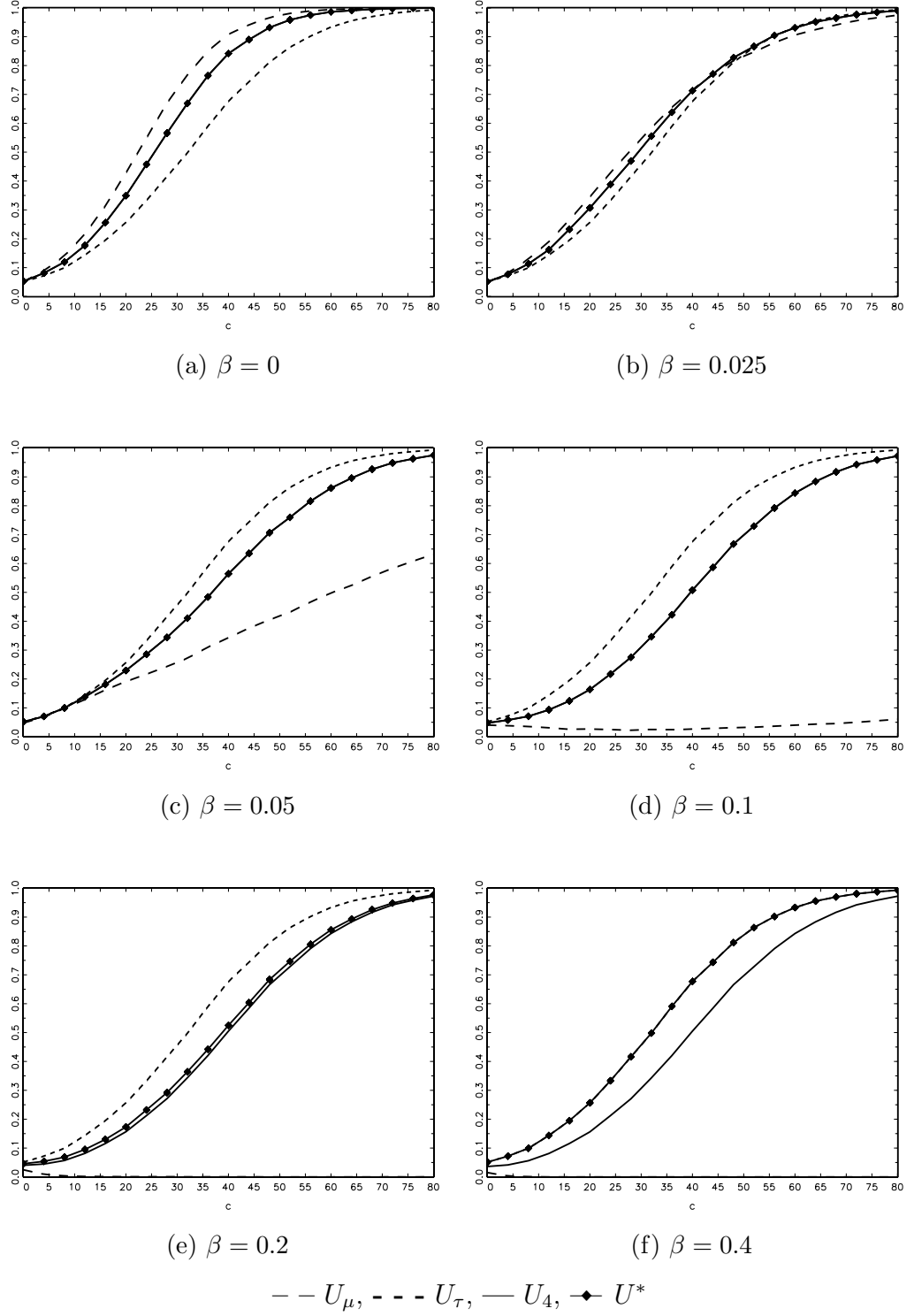


Figure 3.15: Finite sample size and local power of U_μ , U_τ , U_4 and U^* for fixed $g^2 = 50$

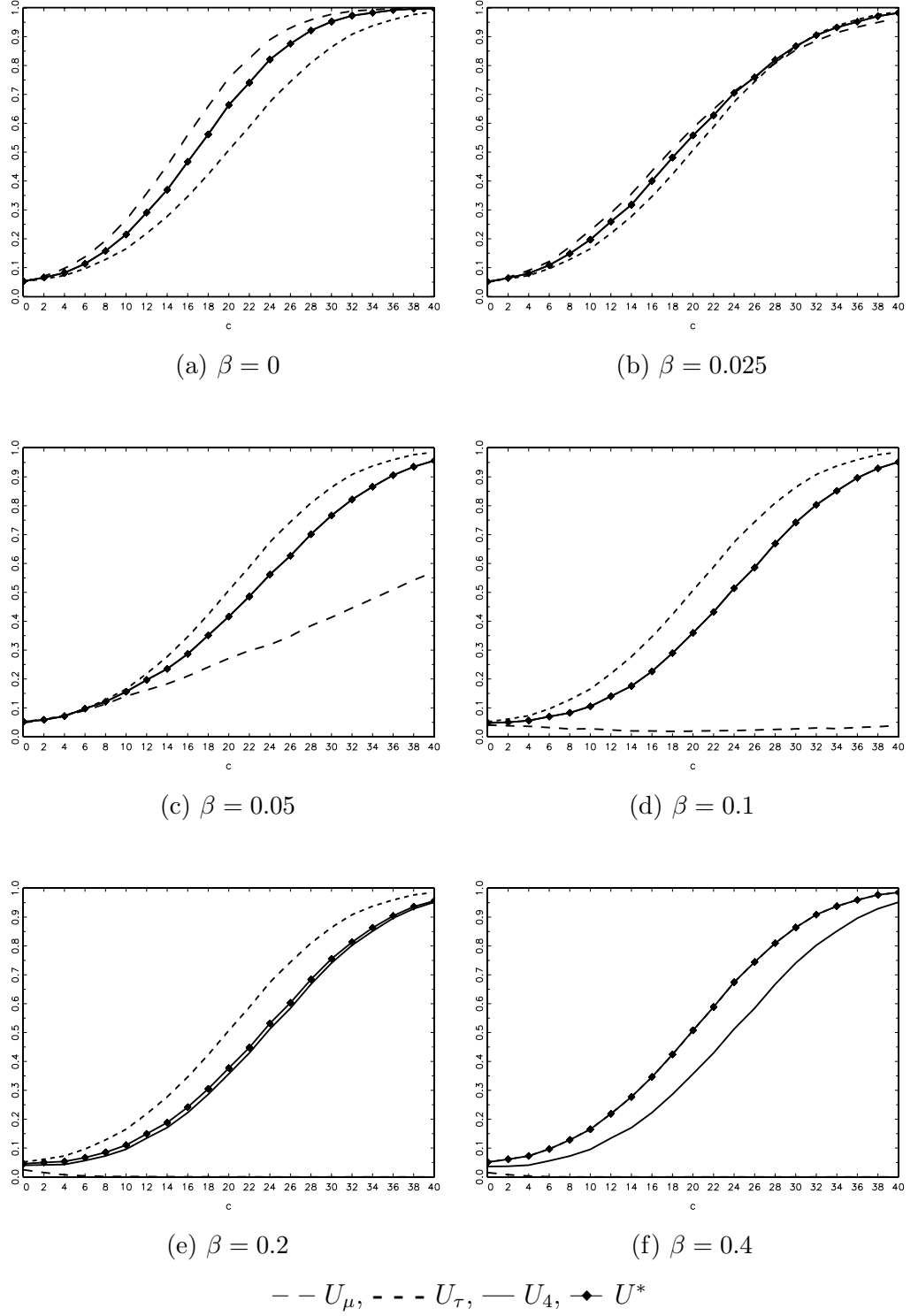
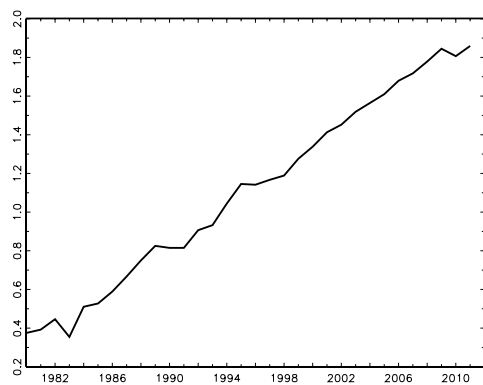
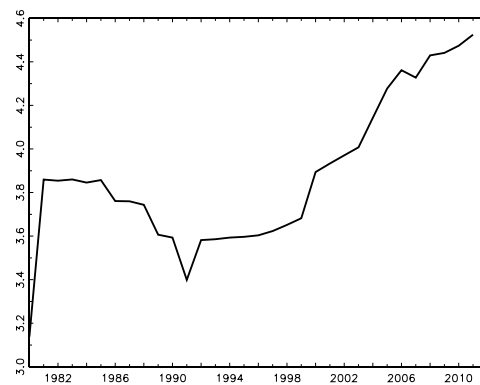


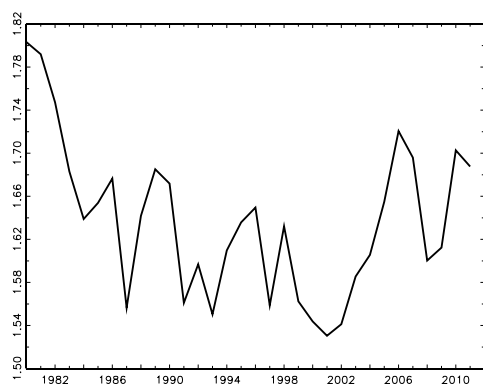
Figure 3.16: Total energy consumption per capita (logs of million BTU per person)



(a) Bangladesh



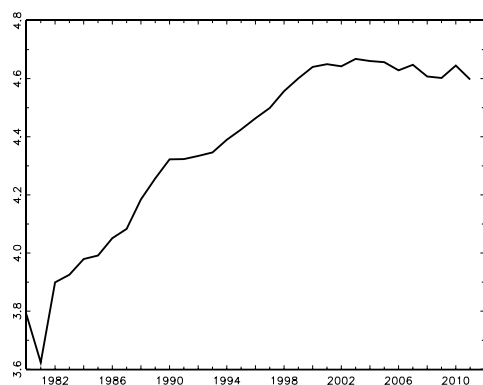
(b) Cook Islands



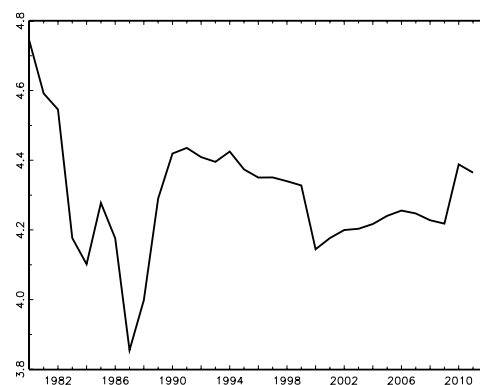
(c) Kenya



(d) Panama



(e) Portugal



(f) Suriname

Appendix

Consider the DGP in (3.5)

$$\Delta u_t = -\frac{c}{T} u_{t-1} \left(1 - \exp \left(-\frac{g^2}{\sigma^2 T} u_{t-1}^2 \right) \right) + \varepsilon_t$$

Standardising this equation by $\frac{1}{\sigma\sqrt{T}}$ gives

$$\frac{\Delta u_t}{\sigma\sqrt{T}} = -\frac{c}{T} \frac{u_{t-1}}{\sigma\sqrt{T}} \left(1 - \exp \left(-g^2 \frac{u_{t-1}^2}{\sigma^2 T} \right) \right) + \frac{\varepsilon_t}{\sigma\sqrt{T}}$$

Weak convergence of the standardised u_t to $X(r)$ follows from Gikhman and Skorokhod (1969), whilst weak convergence of standardised partial sums of ε_t to the standard Brownian motion process $W(r)$ is a standard result. We can therefore write

$$\left(\frac{1}{\sigma\sqrt{T}} \sum_{j=1}^{\lfloor rT \rfloor} \varepsilon_j ; \frac{1}{\sigma\sqrt{T}} u_{\lfloor rT \rfloor} \right) \xrightarrow{d} (W(r) ; X(r))$$

where, following Kurtz and Protter (1991), $X(r)$ is the diffusion given by the stochastic differential equation

$$dX(r) = -cX(r)\mathcal{G}(g^2 X(r)^2) dr + dW(r)$$

with $X(0) = 0$.

Due to the invariance of all statistics to μ in (3.4), we set $\mu = 0$ in what follows, without loss of generality.

Proof of Lemma 3.1

First, we define

$$\begin{aligned} T^{-1/2} u_{\mu, t-1} &= T^{-1/2} \left(u_{t-1} - T^{-1} \sum_{t=2}^T u_{t-1} \right) \\ &= T^{-1/2} u_{t-1} - T^{-3/2} \sum_{t=2}^T u_{t-1} \\ &\xrightarrow{d} X(r) - \int_0^1 X(s) ds \\ &\equiv X_\mu(r) \end{aligned}$$

The DF_μ test statistic is given by

$$DF_\mu = \frac{\sum_{t=2}^T y_{\mu,t-1} \Delta y_{\mu,t}}{\sqrt{\hat{\sigma}^2 \sum_{t=2}^T y_{\mu,t-1}^2}}$$

where $\hat{\sigma}^2$ denotes the error variance estimate from the OLS regression of $\Delta y_{\mu,t}$ on $y_{\mu,t-1}$ and

$$\begin{aligned} \Delta y_{\mu,t} &= \Delta y_t \\ &= \mu + \beta_T t + u_t - (\mu + \beta_T(t-1) + u_{t-1}) \\ &= \beta_T + \Delta u_t \\ &= \beta_T - \frac{c}{T} u_{t-1} G(\theta_T; u_{t-1}^2) + \varepsilon_t \end{aligned}$$

Considering first the numerator of DF_μ , we can write

$$\begin{aligned} T^{-1} \sum_{t=2}^T y_{\mu,t-1} \Delta y_t &= T^{-1} \sum_{t=2}^T (y_{t-1} - \bar{y}) \left(\beta_T - \frac{c}{T} u_{t-1} G(\theta_T, u_{t-1}) + \varepsilon_t \right) \\ &= \beta_T T^{-1} \sum_{t=2}^T (y_{t-1} - \bar{y}) - c T^{-2} \sum_{t=2}^T (y_{t-1} - \bar{y}) u_{t-1} G(\theta_T, u_{t-1}) \\ &\quad + T^{-1} \sum_{t=2}^T (y_{t-1} - \bar{y}) \varepsilon_t. \end{aligned}$$

Evaluating each term separately, and defining $\bar{t} = T^{-1} \sum_{t=1}^T t$, $\bar{u} = T^{-1} \sum_{t=1}^T u_t$,

$$\beta_T T^{-1} \sum_{t=2}^T (y_{t-1} - \bar{y}) = 0$$

and

$$\begin{aligned} cT^{-2} \sum_{t=2}^T (y_{t-1} - \bar{y}) u_{t-1} G(\theta_T, u_{t-1}) \\ &= cT^{-2} \sum_{t=2}^T \{ \beta_T (t-1 - \bar{t}) + u_{t-1} - \bar{u} \} u_{t-1} G(\theta_T, u_{t-1}) \\ &= cT^{-2} \sum_{t=2}^T \{ T^{-1/2} \kappa \sigma (t-1 - \bar{t}) + u_{t-1} - \bar{u} \} u_{t-1} G(\theta_T, u_{t-1}) \\ &= cT^{-1} \sum_{t=2}^T \{ \kappa \sigma T^{-1} (t - \bar{t}) + T^{-1/2} (u_{t-1} - \bar{u}) \} T^{-1/2} u_{t-1} G(\theta_T, u_{t-1}) + o_p(1) \\ &\xrightarrow{d} \sigma^2 c \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right) X(r) \mathcal{G}(g^2 X(r)^2) dr \end{aligned}$$

and

$$\begin{aligned} T^{-1} \sum_{t=2}^T (y_{t-1} - \bar{y}) \varepsilon_t &= T^{-1} \sum_{t=2}^T \{ \beta_T (t-1 - \bar{t}) + u_{t-1} - \bar{u} \} \varepsilon_t \\ &= T^{-1/2} \sum_{t=2}^T \{ \kappa \sigma T^{-1} (t - \bar{t}) + T^{-1/2} (u_{t-1} - \bar{u}) \} \varepsilon_t + o_p(1) \\ &\xrightarrow{d} \sigma^2 \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right) dW(r). \end{aligned}$$

In the denominator of DF_μ , we can write

$$\begin{aligned} T^{-2} \sum_{t=2}^T (y_{t-1} - \bar{y})^2 &= T^{-1} \sum_{t=2}^T \left(\kappa \sigma T^{-1} (t - \bar{t}) + T^{-1/2} (u_{t-1} - \bar{u}) \right)^2 \\ &\xrightarrow{d} \sigma^2 \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right)^2 dr. \end{aligned}$$

and we can show that

$$\begin{aligned}
\hat{\sigma}^2 &= T^{-1} \sum_{t=2}^T \left(\Delta y_t - \frac{T^{-1} \sum_{s=2}^T (y_{s-1} - \bar{y}) \Delta y_s}{T^{-2} \sum_{s=2}^T (y_{s-1} - \bar{y})^2} T^{-1} (y_{t-1} - \bar{y}) \right)^2 \\
&= T^{-1} \sum_{t=2}^T (\Delta y_t)^2 + o_p(1) \\
&= T^{-1} \sum_{t=2}^T \left(\beta_T - \frac{c}{T} u_{t-1} G(\theta_T; u_{t-1}^2) + \varepsilon_t \right)^2 + o_p(1) \\
&= \beta_T^2 - c^2 T^{-3} \sum_{t=2}^T u_{t-1}^2 G(\theta_T; u_{t-1}^2)^2 + T^{-1} \sum_{t=2}^T \varepsilon_t^2 - c \beta_T T^{-2} \sum_{t=2}^T u_{t-1} G(\theta_T; u_{t-1}^2) \\
&\quad + \beta_T T^{-1} \sum_{t=2}^T \varepsilon_t - c T^{-2} \sum_{t=2}^T u_{t-1} G(\theta_T; u_{t-1}^2) \varepsilon_t + o_p(1) \\
&= T^{-1} \sum_{t=2}^T \varepsilon_t^2 + o_p(1) \\
&\xrightarrow{p} \sigma^2
\end{aligned}$$

The DF_μ test statistic therefore has the limiting distribution

$$DF_\mu \xrightarrow{d} \frac{\int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right) dW(r) - c \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right) X(r) \mathcal{G}(g^2 X(r)^2) dr}{\sqrt{\int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right)^2 dr}}$$

The KSS_μ test statistic is given by

$$KSS_\mu = \frac{\sum_{t=2}^T y_{\mu,t-1}^3 \Delta y_{\mu,t}}{\sqrt{\hat{\sigma}^2 \sum_{t=2}^T y_{\mu,t-1}^6}}$$

where $\hat{\sigma}^2$ denotes the error variance estimate from the OLS regression of $\Delta y_{\mu,t}$ on $y_{\mu,t-1}^3$. In the numerator,

$$\begin{aligned} T^{-2} \sum_{t=2}^T y_{\mu,t-1}^3 \Delta y_t &= T^{-2} \sum_{t=2}^T (y_{t-1} - \bar{y})^3 \left(\beta_T - \frac{c}{T} u_{t-1} G(\theta_T, u_{t-1}) + \varepsilon_t \right) \\ &= \kappa \sigma T^{-5/2} \sum_{t=2}^T (y_{t-1} - \bar{y})^3 - c T^{-3} \sum_{t=2}^T (y_{t-1} - \bar{y})^3 u_{t-1} G(\theta_T, u_{t-1}) \\ &\quad + T^{-2} \sum_{t=2}^T (y_{t-1} - \bar{y})^3 \varepsilon_t \\ &= \kappa \sigma T^{-5/2} \sum_{t=2}^T \{ \beta_T (t-1-\bar{t}) + u_{t-1} - \bar{u} \}^3 \\ &\quad - c T^{-3} \sum_{t=2}^T \{ \beta_T (t-1-\bar{t}) + u_{t-1} - \bar{u} \}^3 u_{t-1} G(\theta_T, u_{t-1}) \\ &\quad + T^{-2} \sum_{t=2}^T \{ \beta_T (t-1-\bar{t}) + u_{t-1} - \bar{u} \}^3 \varepsilon_t \\ &\xrightarrow{d} \kappa \sigma^4 \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right)^3 dr \\ &\quad - c \sigma^4 \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right)^3 X(r) \mathcal{G}(g^2 X(r)^2) dr \\ &\quad + \sigma^4 \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right)^3 dW(r) \end{aligned}$$

while in the denominator, it is easy to show that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ and

$$T^{-4} \sum_{t=2}^T y_{\mu,t-1}^6 \xrightarrow{d} \sigma^6 \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right)^6 dr$$

giving the limit for KSS_μ as

$$KSS_\mu \xrightarrow{d} \frac{\kappa \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right)^3 dr + \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right)^3 dW(r) - c \int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right)^3 X(r) \mathcal{G}(g^2 X(r)^2) dr}{\sqrt{\int_0^1 \left(\kappa \left(r - \frac{1}{2} \right) + X_\mu(r) \right)^6 dr}}.$$

The test statistics based on demeaned and detrended data are invariant to β , hence we set $\beta_T = 0$ without loss of generality in the remainder of this proof. The DF_τ test statistic is

$$DF_\tau = \frac{\sum_{t=2}^T y_{\tau,t-1} \Delta y_{\tau,t}}{\sqrt{\hat{\sigma}^2 \sum_{t=2}^T y_{\tau,t-1}^2}}$$

where $\hat{\sigma}^2$ denotes the error variance estimate from the OLS regression of $\Delta y_{\tau,t}$ on $y_{\tau,t-1}$ and $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$. Here,

$$\begin{aligned} \Delta y_{\tau,t} &= y_t - \hat{\mu} - \hat{\beta}t - \{y_{t-1} - \hat{\mu} - \hat{\beta}(t-1)\} \\ &= \Delta y_t - \hat{\beta} \\ &= -\frac{c}{T} u_{t-1} G(\theta_T, u_{t-1}) + \varepsilon_t - \hat{\beta} \end{aligned}$$

and

$$\begin{aligned} T^{1/2} \hat{\beta} &= \frac{T^{-5/2} \sum_{t=1}^T (t - \bar{t}) y_t}{T^{-3} \sum_{t=1}^T (t - \bar{t})^2} \\ &\xrightarrow{d} 12 \int_0^1 \left(r - \frac{1}{2}\right) X(r) dr. \end{aligned}$$

Also,

$$\begin{aligned} T^{-1/2} y_{\tau, [rT]} &= T^{-1/2} y_{[rT]} - T^{-1/2} \hat{\mu} - T^{1/2} \hat{\beta} r \\ &= T^{-1/2} y_{[rT]} - T^{-1/2} \left(\bar{y} - \hat{\beta} \bar{t} \right) - T^{1/2} \hat{\beta} r \\ &= T^{-1/2} y_{[rT]} - \left(T^{-1/2} \bar{y} - T^{1/2} \hat{\beta} T^{-1} \bar{t} \right) - T^{1/2} \hat{\beta} r \\ &\xrightarrow{d} X(r) - \int_0^1 X(s) ds + 6 \int_0^1 \left(s - \frac{1}{2}\right) X(s) ds - 12r \int_0^1 \left(s - \frac{1}{2}\right) X(s) ds \\ &= X(r) - \int_0^1 X(s) ds - 12 \left(r - \frac{1}{2}\right) \int_0^1 \left(s - \frac{1}{2}\right) X(s) ds \\ &= X(r) + (6r - 4) \int_0^1 X(s) ds + (6 - 12r) \int_0^1 s X(s) ds \\ &\equiv X_\tau(r). \end{aligned}$$

The numerator of DF_τ is then given by

$$\begin{aligned}
T^{-1} \sum_{t=2}^T y_{\tau,t-1} \Delta y_{\tau,t} &= T^{-1} \sum_{t=2}^T y_{\tau,t-1} \left\{ -\frac{c}{T} u_{t-1} G(\theta_T, u_{t-1}) + \varepsilon_t - \widehat{\beta} \right\} \\
&= T^{-1} \sum_{t=2}^T y_{\tau,t-1} \varepsilon_t - c T^{-2} \sum_{t=2}^T y_{\tau,t-1} u_{t-1} G(\theta_T, u_{t-1}) \\
&\quad - T^{1/2} \widehat{\beta} T^{-3/2} \sum_{t=2}^T y_{\tau,t-1} \\
&\xrightarrow{d} \sigma^2 \int_0^1 X_\tau(r) dW(r) - c \sigma^2 \int_0^1 X_\tau(r) X(r) \mathcal{G}(g^2 X(r)^2) dr
\end{aligned}$$

as

$$\sum_{t=2}^T y_{\tau,t-1} = 0$$

and for the denominator we obtain

$$T^{-2} \sum_{t=2}^T y_{\tau,t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 X_\tau(r)^2 dr$$

giving

$$DF_\tau \xrightarrow{d} \frac{\int_0^1 X_\tau(r) dW(r) - c \int_0^1 X_\tau(r) X(r) \mathcal{G}(g^2 X(r)^2) dr}{\sqrt{\int_0^1 X_\tau(r)^2 dr}}.$$

Finally, the KSS_τ test statistic is

$$KSS_\tau = \frac{\sum_{t=2}^T y_{\tau,t-1}^3 \Delta y_{\tau,t}}{\sqrt{\widehat{\sigma}^2 \sum_{t=2}^T y_{\tau,t-1}^6}}$$

where $\widehat{\sigma}^2$ is the error variance estimate from the OLS regression of $\Delta y_{\tau,t}$ on $y_{\tau,t-1}^3$. Now $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$ as before,

$$\begin{aligned}
T^{-2} \sum_{t=2}^T y_{\tau,t-1}^3 \Delta y_{\tau,t} &= T^{-2} \sum_{t=2}^T y_{\tau,t-1}^3 \varepsilon_t - c T^{-3} \sum_{t=2}^T y_{\tau,t-1}^3 u_{t-1} G(\theta_T, u_{t-1}) \\
&\quad - T^{1/2} \widehat{\beta} T^{-5/2} \sum_{t=2}^T y_{\tau,t-1}^3 \\
&\xrightarrow{d} \sigma^2 \int_0^1 X_\tau(r)^3 dW(r) - c \sigma^2 \int_0^1 X_\tau(r)^3 X(r) \mathcal{G}(g^2 X(r)^2) dr \\
&\quad - 12 \int_0^1 \left(r - \frac{1}{2}\right) X(r) dr \int_0^1 X_\tau(r)^3 dr
\end{aligned}$$

and

$$T^{-4} \sum_{t=2}^T y_{\tau, t-1}^6 \xrightarrow{d} \int_0^1 X_{\tau}(r)^6 dr$$

so

$$KSS_{\tau} \xrightarrow{d} \frac{\int_0^1 X_{\tau}(r)^3 dW(r) - c \int_0^1 X_{\tau}(r)^3 X(r) \mathcal{G}(g^2 X(r)^2) dr - 12 \int_0^1 \left(r - \frac{1}{2}\right) X(r) dr \int_0^1 X_{\tau}(r)^3 dr}{\sqrt{\int_0^1 X_{\tau}(r)^6 dr}}.$$

Proof of Lemma 3.2

As before, the DF_μ test statistic is given by

$$DF_\mu = \frac{\sum_{t=2}^T y_{\mu,t-1} \Delta y_t}{\sqrt{\hat{\sigma}^2 \sum_{t=2}^T y_{\mu,t-1}^2}}$$

where $\hat{\sigma}^2$ is the error variance estimate from the OLS regression of Δy_t on $y_{\mu,t-1}$.

We find

$$\begin{aligned} T^{-3/2} \sum_{t=2}^T y_{\mu,t-1} \Delta y_t &= T^{-3/2} \sum_{t=2}^T (\beta(t-1-\bar{t}) + u_{t-1} - \bar{u}) \left(\beta - \frac{c}{T} u_{t-1} G(\theta_T, u_{t-1}) + \varepsilon_t \right) \\ &= \sigma^2 \kappa^2 T^{-3/2} \sum_{t=2}^T (t-\bar{t}) - \sigma \kappa c T^{-5/2} \sum_{t=2}^T (t-\bar{t}) u_{t-1} G(\theta_T, u_{t-1}) \\ &\quad + \sigma \kappa T^{-3/2} \sum_{t=2}^T (t-\bar{t}) \varepsilon_t + \sigma \kappa T^{-3/2} \sum_{t=2}^T (u_{t-1} - \bar{u}) \\ &\quad - c T^{-5/2} \sum_{t=2}^T (u_{t-1} - \bar{u}) u_{t-1} G(\theta_T, u_{t-1}) \\ &\quad + T^{-3/2} \sum_{t=2}^T (u_{t-1} - \bar{u}) \varepsilon_t + o_p(1) \\ &= \sigma \kappa T^{-3/2} \sum_{t=2}^T (t-\bar{t}) \varepsilon_t - \sigma \kappa c T^{-5/2} \sum_{t=2}^T (t-\bar{t}) u_{t-1} G(\theta_T, u_{t-1}) \\ &\quad + o_p(1) \\ &\xrightarrow{d} \sigma^2 \kappa \int_0^1 \left(r - \frac{1}{2}\right) dW(r) - \sigma^2 \kappa c \int_0^1 \left(r - \frac{1}{2}\right) X(r) \mathcal{G}(g^2 X(r)^2) dr \end{aligned}$$

and

$$\begin{aligned} T^{-3} \sum_{t=2}^T y_{\mu,t-1}^2 &= T^{-3} \sum_{t=2}^T (\beta(t-1-\bar{t}) + u_{t-1} - \bar{u})^2 \\ &= \sigma^2 \kappa^2 T^{-3} \sum_{t=2}^T (t-\bar{t})^2 + T^{-3} \sum_{t=2}^T (u_{t-1} - \bar{u})^2 \\ &\quad + 2\sigma \kappa T^{-3} \sum_{t=2}^T (t-\bar{t}) (u_{t-1} - \bar{u}) + o_p(1) \\ &= \sigma^2 \kappa^2 T^{-3} \sum_{t=2}^T (t-\bar{t})^2 + o_p(1) \\ &\xrightarrow{p} \sigma^2 \kappa^2 / 12. \end{aligned}$$

Also,

$$\begin{aligned}
\hat{\sigma}^2 &= T^{-1} \sum_{t=2}^T \left(\Delta y_t - \frac{\sum_{s=2}^T (y_{s-1} - \bar{y}) \Delta y_s}{\sum_{s=2}^T (y_{s-1} - \bar{y})^2} (y_{t-1} - \bar{y}) \right)^2 \\
&= T^{-1} \sum_{t=2}^T \left(\Delta y_t - \frac{T^{-3/2} \sum_{s=2}^T (y_{s-1} - \bar{y}) \Delta y_s}{T^{-3} \sum_{s=2}^T (y_{s-1} - \bar{y})^2} T^{-3/2} (y_{t-1} - \bar{y}) \right)^2 \\
&= T^{-1} \sum_{t=2}^T (\Delta y_t)^2 + o_p(1) \\
&= \sigma^2 \kappa^2 + T^{-1} \sum_{t=2}^T \varepsilon_t^2 + o_p(1) \\
&\xrightarrow{p} \sigma^2 (\kappa^2 + 1).
\end{aligned}$$

Therefore the DF_μ test statistic has the limiting distribution

$$DF_\mu \xrightarrow{d} \frac{\int_0^1 (r - \frac{1}{2}) dW(s) - c \int_0^1 (r - \frac{1}{2}) X(r) \mathcal{G}(g^2 X(r)^2) dr}{\sqrt{(\kappa^2 + 1)/12}}.$$

For KSS_μ we again have

$$KSS_\mu = \frac{\sum_{t=2}^T y_{\mu,t-1}^3 \Delta y_t}{\sqrt{\hat{\sigma}^2 \sum_{t=2}^T y_{\mu,t-1}^6}}$$

where $\hat{\sigma}^2$ is the error variance estimate from the OLS regression of Δy_t on $y_{\mu,t-1}^3$.

We obtain

$$\begin{aligned}
T^{-7/2} \sum_{t=2}^T y_{\mu, t-1}^3 \Delta y_t &= T^{-7/2} \sum_{t=2}^T (\sigma \kappa (t - \bar{t}) + u_{t-1} - \bar{u})^3 \left(\sigma \kappa - \frac{c}{T} u_{t-1} G(\theta_T, u_{t-1}) + \varepsilon_t \right) \\
&\quad + o_p(1) \\
&= T^{-7/2} \sigma^4 \kappa^4 \sum_{t=1}^T (t - \bar{t})^3 + \sigma \kappa T^{-7/2} \sum_{t=1}^T (u_{t-1} - \bar{u})^3 \\
&\quad + 3\sigma^3 \kappa^3 T^{-7/2} \sum_{t=1}^T (t - \bar{t})^2 (u_{t-1} - \bar{u}) \\
&\quad + 3\sigma^2 \kappa^2 T^{-7/2} \sum_{t=1}^T (t - \bar{t}) (u_{t-1} - \bar{u})^2 \\
&\quad + \sigma^3 \kappa^3 T^{-7/2} \sum_{t=1}^T (t - \bar{t})^3 \varepsilon_t + T^{-7/2} \sum_{t=1}^T (u_{t-1} - \bar{u})^3 \varepsilon_t \\
&\quad + 3\sigma^2 \kappa^2 T^{-7/2} \sum_{t=1}^T (t - \bar{t})^2 (u_{t-1} - \bar{u}) \varepsilon_t \\
&\quad + 3\sigma \kappa T^{-7/2} \sum_{t=1}^T (t - \bar{t}) (u_{t-1} - \bar{u})^2 \varepsilon_t \\
&\quad - c \sigma^3 \kappa^3 T^{-9/2} \sum_{t=1}^T (t - \bar{t})^3 u_{t-1} G(\theta_T, u_{t-1}) \\
&\quad - c T^{-9/2} \sum_{t=1}^T (u_{t-1} - \bar{u})^3 u_{t-1} G(\theta_T, u_{t-1}) \\
&\quad - 3c \sigma^2 \kappa^2 T^{-9/2} \sum_{t=1}^T (t - \bar{t})^2 (u_{t-1} - \bar{u}) u_{t-1} G(\theta_T, u_{t-1}) \\
&\quad - 3c \sigma \kappa T^{-9/2} \sum_{t=1}^T (t - \bar{t}) (u_{t-1} - \bar{u})^2 u_{t-1} G(\theta_T, u_{t-1}) + o_p(1) \\
&= 3\sigma^3 \kappa^3 T^{-7/2} \sum_{t=2}^T (t - \bar{t})^2 (u_{t-1} - \bar{u}) + \sigma^3 \kappa^3 T^{-7/2} \sum_{t=2}^T (t - \bar{t})^3 \varepsilon_t \\
&\quad - c \sigma^3 \kappa^3 T^{-9/2} \sum_{t=2}^T (t - \bar{t})^3 u_{t-1} G(\theta_T, u_{t-1}) + o_p(1) \\
&\xrightarrow{d} 3\sigma^4 \kappa^3 \int_0^1 \left(r - \frac{1}{2}\right)^2 X_\mu(r) dr + \sigma^4 \kappa^3 \int_0^1 \left(r - \frac{1}{2}\right)^3 dW(r) \\
&\quad - c \sigma^4 \kappa^3 \int_0^1 \left(r - \frac{1}{2}\right)^3 X(r) \mathcal{G}(g^2 X(r)^2) dr
\end{aligned}$$

and

$$\begin{aligned} T^{-7} \sum_{t=2}^T y_{\mu, t-1}^6 &= \sigma^6 \kappa^6 T^{-7} \sum_{t=2}^T (t - \bar{t})^6 + o_p(1) \\ &\xrightarrow{p} \sigma^6 \kappa^6 / 448 \end{aligned}$$

together with

$$\begin{aligned} \hat{\sigma}^2 &= T^{-1} \sum_{t=2}^T \left(\Delta y_t - \frac{T^{-7/2} \sum_{s=2}^T (y_{s-1} - \bar{y})^3 \Delta y_s}{T^{-7} \sum_{s=2}^T (y_{s-1} - \bar{y})^6} T^{-7/2} (y_{t-1} - \bar{y})^3 \right)^2 \\ &= T^{-1} \sum_{t=2}^T (\Delta y_t)^2 + o_p(1) \\ &\xrightarrow{p} \sigma^2 (\kappa^2 + 1) \end{aligned}$$

giving the KSS_μ limit distribution

$$KSS_\mu \xrightarrow{d} \frac{3 \int_0^1 \left(r - \frac{1}{2}\right)^2 X_\mu(r) dr + \int_0^1 \left(r - \frac{1}{2}\right)^3 dW(r) - c \int_0^1 \left(r - \frac{1}{2}\right)^3 X(r) \mathcal{G}(g^2 X(r)^2) dr}{\sqrt{(\kappa^2 + 1) / 448}}$$

Chapter 4

Testing equal forecast accuracy and forecast encompassing in small samples

In this chapter we show that when computing standard Diebold-Mariano-type tests for equal forecast accuracy and forecast encompassing, the long-run variance can frequently be negative when dealing with multi-step-ahead predictions in small, but empirically relevant, sample sizes. We subsequently consider a number of alternative approaches to dealing with this problem, including direct inference in the problem cases and use of long-run variance estimators that guarantee positivity. The finite sample size and power of the different approaches are evaluated using an extensive Monte Carlo simulation exercise. Overall, for multi-step-ahead forecasts, we find that the recently proposed Coroneo and Iacone (2015) test, which is based on a weighted periodogram long-run variance estimator, offers the best finite sample size and power performance.

4.1 Introduction

Given the critical role that forecasting plays in economic and financial research and policy-making, the evaluation of competing forecasts has become an extensive and prominent field in the econometric and empirical economic literatures. Within this field, the most common forecast evaluation exercise typically undertaken is to compare the accuracy of two or more sets of forecasts on the basis of some measure of loss associated with the forecast errors, such as mean squared forecast error. While comparison of such accuracy measures provides an indication as to the relative performance of the different forecasting methodologies, inevitably one forecast will achieve a lower mean squared error than another and thus we are almost always able to declare a ‘winning’ forecast. In order to determine whether a forecast’s superior predictive performance is statistically significant, we require more formal statistical evaluation in the form of testing a null hypothesis of equal forecast accuracy.

One of the most widely used tests for forecast accuracy is that of Diebold and Mariano (1995) [DM] who consider the null hypothesis that two forecasts have equal accuracy using a test statistic based on the sample mean of a loss differential series (i.e. the difference in the measure of loss associated with the two sets of forecast errors). Harvey et al. (1997) [HLNa] demonstrate that the DM test can be over-sized when only a small number of out of sample forecasts are available, particularly as the forecast horizon increases. They propose a finite sample bias correction and use of different critical values to better control size in these small samples. An alternative evaluation exercise commonly undertaken is to assess whether one forecast encompasses another. If a forecast provides no additional, relevant information than is already contained within another forecast, it is said to be encompassed by this competing forecast. Harvey et al. (1998) [HLNb] propose a test for encompassing that uses a DM-type approach, with the loss differential series redefined to consider a null hypothesis of forecast encompassing.

In this paper we focus on the behaviour of the DM/HLNa tests based on squared error loss, and the HLNb test for encompassing, in small samples. The DM test statistic is fundamentally comprised of the loss differential series sample mean divided by an estimate of the long-run variance. DM make use of the fact that optimal h -steps-ahead forecasts are at most $(h - 1)$ -order dependent to advocate use of a rectangular kernel in the long-run variance estimator which truncates at lag $h - 1$. While this approach results in decent finite sample size and power properties for many sample size and h settings, the long-run variance estimator is not guaranteed to be positive whenever $h > 1$. DM note this possibility,

but suggest that such an outcome would be rare; similarly, Clark (1999) finds a low occurrence of negative long-run variance estimates in his equal accuracy test simulations (always less than 3% of replications). However, these observations were made on the basis of results that considered predictions only up to two steps ahead. The first contribution of this chapter is to highlight that the prevalence of negative long-run variance estimates can be much greater in small samples when longer horizon forecasts are considered. For example, when testing equal mean squared forecast error with $h = 6$, we find that negative variance estimates arise approximately 20% of the time for a sample size of 16, rising to over 40% of the time for a sample size of 8.

In practical applications, often due to the limitations of economic or forecast data, it is not uncommon for forecast evaluation to be conducted using sample size and forecast horizon settings that lie in the region where negative variance estimates occur frequently. For example, in the context of testing for equal forecast accuracy, the recent papers Caporale and Gil-Alana (2014), Dreger and Wolters (2014), Dib et al. (2008) and Qin et al. (2008) all implement the DM/HLNa tests in forecast samples smaller than 25 observations with horizons of $h = 6$ or greater. Further, Mehl (2009) and Chow and Choy (2006) have reported finding negative DM/HLNa long-run variance estimates when using samples of 18 and 24 forecasts at horizons of 6 and 5-6, respectively.

Given that negative variance estimates can arise frequently in situations of practical relevance, it is important to determine the best approach to deliver a reliable testing procedure in terms of small sample size and power properties. DM suggest treating a negative variance estimate as a zero, thereby rejecting the null in such cases. However, given the low occurrence of negative variance estimates in their simulations, the size implications of such an approach are not fully explored. In the simulation work of Clark (1999), the relatively few replications where negative variance estimates were obtained were excluded from the simulations, thereby abstracting from the effects of dealing with certain problematic cases. HLNa and HLNb simulated combinations of sample size and forecast horizon where negative variances can occur frequently, but their simulations failed to correctly deal with negative variance estimates, thus again the impact of negativity in the variance estimate is not clear. Of course, other long-run variance estimators exist which ensure non-negativity; for example, DM, Clark (1999) and others discuss the possible use of the Bartlett kernel, and in a recent paper on testing equal forecast accuracy, Coroneo and Iacone (2015) recommend use of the nonparametric periodogram estimator of Hualde and Iacone (2017), combined with use of bandwidth-dependent critical values. The second contribution

of this chapter is therefore to formally assess the behaviour of different strategies for dealing with the potential problem of negative long-run variance estimation in tests of equal accuracy and encompassing. We conduct an extensive Monte Carlo simulation exercise to establish the small sample size and power properties of different approaches. Broadly, we find that for multi-step-ahead forecasts, the Coroneo and Iacone (2015) approach outperforms other methods, and the attractive finite sample properties reported in their paper for moderate sample sizes and forecast horizons extends to the small sample and longer horizon region under focus in this paper for both equal accuracy and encompassing tests, where the DM/HLNa and HLNb tests can suffer from negative variance estimates. To assess the empirical validity of our results, we consider an application of the tests in this chapter to errors calibrated from forecasts of Euro-area inflation.

The outline of the paper is as follows. In Section 4.2, we discuss some of the key contributions to the forecast evaluation literature. In Section 4.3 we outline the DM, HLNa and HLNb tests for equal mean squared forecast error and forecast encompassing. Section 4.4 highlights by simulation the frequency with which negative long-run variance estimates can arise for different sample sizes and forecast horizons. In Section 4.5, a number of ways of dealing with these cases are considered, including alternative long-run variance estimators that are guaranteed to be non-negative. Section 4.6 reports results from our finite sample size and power simulations. In 4.6.3 we consider an application of these tests to forecasts of Euro-area inflation. Section 4.7 concludes.

4.2 Forecast evaluation

Tests of equal mean squared forecast error originated with Granger and Newbold (1977), with their approach extended to account for serially correlated forecast errors by Meese and Rogoff (1988). In a key contribution to the literature, DM propose an alternative approach that is valid for potentially contemporaneously correlated, serially correlated and non-normal forecast errors. Contemporaneous correlation is a particularly important consideration as the two forecasts being compared are of the same economic time series. The two forecasters are likely to have largely overlapping information sets and may therefore produce forecasts that are contemporaneously correlated to one another. It is also well known that multi-step ahead forecasts are likely to produce serially correlated forecast errors, as the forecast horizon is longer than the interval at which the data is sampled. The presence of very large, occasional errors in forecasting exercises gives rise to the possibility that the distribution of forecast errors may be heavy-tailed com-

pared to the normal distribution. DM test for a zero mean in a ‘loss differential’ series, defined as the difference between two forecasts’ error loss functions. This framework is substantially more general than previous work, allowing testing for equal accuracy using any defined loss function. The sample mean of the loss differential series is divided by an estimator of the long-run variance. The estimator used employs a rectangular kernel that truncates at lag $h - 1$, and is not guaranteed to be positive. DM state that such occurrences would be rare, and suggest treating a negative variance as zero and thus a rejection of the null hypothesis. Alternatively, non-negativity of the estimator can be enforced using the Bartlett kernel of Newey and West (1987). Considering 2-step-ahead forecasts, such that errors follow an $MA(1)$ structure, size simulations of DM demonstrate that the test is robust to serially correlated, contemporaneously correlated and heavy-tailed errors. However severe size distortion is seen in small samples. For example, when considering normally distributed errors, with no serial correlation or contemporaneous correlation, a nominal 0.10-level DM test has size of 0.111 for $T = 256$ and 0.314 for $T = 8$.

HLNa suggest two finite sample modifications to the DM statistic to improve size control in small samples. First, they demonstrate that the long-run variance estimator used in the DM test is biased in small samples, and propose a finite sample bias correction to the test statistic. Second, it is known that in small samples, and when the population standard deviation is unknown, an estimate of the mean of a normally distributed population follows a Student’s t distribution with $T - 1$ degrees of freedom. HLNa therefore make use of Student’s t critical values rather than those from a standard normal distribution. In a Monte Carlo simulation exercise they demonstrate that the DM test becomes increasingly oversized as the forecast horizon increases. For example, when $T = 16$, the DM test has a size of 0.135 when $h = 1$, 0.345 when $h = 5$, and 0.453 when $h = 9$. Comparatively, the MDM test has empirical size much closer to nominal size, with sizes of 0.091 ($h = 1$), 0.199 ($h = 4$) and 0.151 ($h = 9$) for the same sample size.

Application of the DM test or its HLNa variant have now become prevalent in empirical forecasting research, to the extent that it is now routine for the results of such forecast accuracy tests to be reported alongside any forecast comparisons. Beginning primarily with West (1996), subsequent work on testing for equal forecast accuracy has focused on cases where the forecasts have been produced by estimated models, either non-nested or nested, and more sophisticated methods have been proposed to properly account for the impact that model parameter estimation uncertainty can have on the distributions of forecast accuracy tests

in such circumstances. For reviews of this literature, see West (2006) and Clark and McCracken (2013). Additionally, if two forecast models are nested, such that one model is comprised only of terms included in another model, then forecast errors from the two models will be perfectly correlated under the null hypothesis of equal accuracy. In this circumstance, more appealing test procedures exist, see Clark and McCracken (2001) for example. In some situations where forecasts have been obtained from estimated models, the original DM approach is asymptotically valid without the need for any modification. As outlined by West (1996), examples are where the forecast models are non-nested, linear and estimated by ordinary least squares, along with the loss function being mean squared forecast error, or when the number of forecast observations is small relative to the number of observations used for model estimation. Also, many forecast evaluation exercises are conducted using predictions where the underlying models are unknown (as is the case with the use of forecast survey data), or where the nature of the models makes accounting for parameter estimation uncertainty intractable. For these reasons, the DM/HLNa tests are still very widely applied in practice.

In a large Monte Carlo exercise, Clark (1999) evaluates the finite sample properties of the DM and HLNa tests. The size and power performance of the DM test is examined under three alternative heteroskedasticity and autocorrelation consistent (HAC) variance estimators: the rectangular kernel with a bandwidth of $h = 1$, and the Bartlett kernel where the bandwidth is determined using two different data-based methods of lag selection, the Newey and West (1994) and the pre-whitened Newey-West procedures. When the rectangular kernel is employed, replications where negative long-run variance estimates occur are excluded from their results. Whilst a range of other kernels exist, Newey and West (1994) show that the choice of kernel is less important than the choice of bandwidth, and Clark (1999) therefore employ the Bartlett kernel with data-determined lag lengths. Examining one and two step forecast horizons only, and when comparing the three different HAC variance estimators, the rectangular kernel offers the smallest size distortions when $h = 1$. When $h = 2$ and the degree of serial correlation in the errors is greater than zero, no general ranking of the HAC estimators is possible. In small samples, the HLNa test suffers smaller size distortions than the DM test with any of the three HAC variance estimators used.

Testing for equal forecast accuracy is just one approach to evaluating the predictive ability of rival forecasts. A second popular evaluation method is to test for whether one set of forecasts encompasses another. When faced with competing forecasts it may be the case that regardless of one forecast's superior predictive performance over another, the least favoured forecast still contains information

which is relevant and valuable to the forecasting problem at hand. In these circumstances a combination of two or more forecasts may be optimal, rather than choosing to use a single forecast in isolation. In seminal work, Bates and Granger (1969) discuss the possibility that competing forecasts contain independent information. This may be due to one forecast using variables that the competing forecast hasn't considered, or due to different assumptions made by competing forecasts about the relationship between variables. They show that in these instances, a combination forecast can yield lower MSEs than either of the original forecasts considered separately. Forecast combination is only justified when a competing forecast offers additional, relevant information to that contained in the preferred forecast. If this is not the case, the preferred forecast is said to encompass the competing forecast, in the sense that the encompassed forecast does not result in a reduction in forecast accuracy when used in combination with the encompassing forecast. The idea of forecast encompassing began with the Nelson (1972) and Granger and Newbold (1973) notion of conditional efficiency, and was further developed by, *inter alia*, Chong and Hendry (1986) and Fair and Shiller (1989). Motivated by developing an encompassing test that is robust to serial correlation and non-normality in the forecast errors, Harvey et al. (1998) [HLNb] propose a test based on a DM-type approach. Here, the loss differential is redefined to permit testing a forecast encompassing null hypothesis, and the approach has become standard in cases where one abstracts from model parameter estimation uncertainty.

4.3 Standard tests for equal accuracy and encompassing

Consider two competing forecasts of the same economic variable which we wish to evaluate. Denote the actual series by y_t and the competing forecasts by f_{1t} and f_{2t} , $t = 1, \dots, T$, such that T is the number of out-of-sample forecasts and consider a given loss function $L(\cdot)$ that depends on the forecast errors, so that the cost of error associated with the forecast f_{it} is $L(e_{it})$, $i = 1, 2$. Now define the loss differential series

$$d_t = L(e_{1t}) - L(e_{2t}), \quad t = 1, \dots, T.$$

The null hypothesis of equal forecast accuracy, according to the specified loss function $L(\cdot)$, can then be expressed as

$$H_0 : E(d_t) = 0.$$

For example, under squared error loss, $d_t = e_{1t}^2 - e_{2t}^2$ and the null hypothesis entails the equality of population mean squared forecast errors.

Under the assumptions that d_t is covariance stationary and short memory, DM propose a test of H_0 based on the asymptotic distribution of the sample mean loss differential

$$\sqrt{T}(\bar{d} - E(d_t)) \xrightarrow{d} N(0, \omega^2)$$

where $\bar{d} = T^{-1} \sum_{t=1}^T d_t$ and ω^2 denotes the long-run variance of d_t , i.e. $\omega^2 = \sum_{j=-\infty}^{\infty} \gamma_j$ with $\gamma_j = \text{Cov}(d_t, d_{t-j})$. Denoting a consistent long-run variance estimator by $\hat{\omega}^2$, the DM test statistic is then given by

$$DM = \sqrt{T} \left(\frac{\bar{d}}{\hat{\omega}} \right)$$

which has an asymptotic standard normal distribution under the null. DM suggest use of a long-run variance estimator comprised of a weighted sum of sample autocovariances, and, motivated by the fact that optimal h -steps-ahead forecast errors are at most $(h-1)$ -dependent, they advocate using a rectangular kernel truncated at lag $h-1$, i.e.

$$\hat{\omega}^2 = \hat{\gamma}_0 + 2 \sum_{j=1}^{h-1} \hat{\gamma}_j \quad (4.1)$$

where $\hat{\gamma}_j = T^{-1} \sum_{t=j+1}^T (d_t - \bar{d})(d_{t-j} - \bar{d})$.¹ HLNa propose a modification of the DM statistic designed to improve the small sample size behaviour of the test. Their statistic is based on an approximate bias correction to the long-run variance estimator, and can be written as

$$MDM = \sqrt{T+1-2h+T^{-1}h(h-1)} \left(\frac{\bar{d}}{\hat{\omega}} \right). \quad (4.2)$$

These authors also suggest use of t_{T-1} critical values in place of those from the standard normal, again to better control small sample size.

Next consider investigating whether one set of forecasts encompasses another, in that the accuracy of one set of (encompassing) forecasts f_{1t} cannot be improved through linear combination with a second set of (encompassed) forecasts f_{2t} .

¹Note that ARCH-type behaviour in the forecast errors induces additional autocorrelation into d_t , requiring use of higher order lags; see Harvey et al. (1999).

HLNb develop a test for forecast encompassing based on the Bates and Granger (1969)) forecast combination scheme, where the combination weights sum to one.² Denoting the combined forecast by f_{ct} , the combination is

$$f_{ct} = (1 - \delta)f_{1t} + \delta f_{2t}$$

where δ ($0 \leq \delta \leq 1$) determines the weights associated with the constituent forecasts. We can define the forecast error of the combined forecast as

$$\begin{aligned} e_{ct} &= y_t - f_{ct} \\ &= y_t - [(1 - \delta)f_{1t} + \delta f_{2t}] \\ &= y_t - [(1 - \delta)(y_t - e_{1t}) + \delta(y_t - e_{2t})] \\ &= (1 - \delta)e_{1t} + \delta e_{2t} \end{aligned}$$

The optimal weight is one which minimises the expected value of the squared forecast error

$$E[e_{ct}^2] = (1 - \delta)^2 E[e_{1t}^2] + \delta^2 E[e_{2t}^2] + 2\delta(1 - \delta)E[e_{1t}e_{2t}]$$

The first order condition is given by

$$\frac{dE[e_{ct}^2]}{d\delta} = -2(1 - \delta)E[e_{1t}^2] + 2\delta E[e_{2t}^2] + 2(1 - 2\delta)E[e_{1t}e_{2t}] = 0$$

which can be rearranged to give

$$\delta_{opt} = \frac{E(e_{1t}^2) - E(e_{1t}e_{2t})}{E(e_{1t}^2) + E(e_{2t}^2) - 2E(e_{1t}e_{2t})}$$

In this context, forecast f_{1t} encompasses forecast f_{2t} if the optimal mean squared error-minimising combination weight δ_{opt} is equal to zero. The null of forecast encompassing can then be expressed in a DM-type form:

$$H_0 : E(d_t) = 0$$

with d_t in this case defined as

$$d_t = e_{1t}(e_{1t} - e_{2t}). \quad (4.3)$$

HLNb therefore propose applying the DM approach to this testing problem, along

²Extensions of the test to allow for biased forecasts and combination weights that are not constrained to sum to one are discussed in Clements and Harvey (2009).

with the HLN_a bias correction and use of t_{T-1} critical values. The test statistic is then (4.2) but with d_t given by (4.3). The test is conducted against the one-sided alternative $E(d_t) > 0$ (i.e. $\delta > 0$), given the assumption of a non-negative combination weight.

When considering equal accuracy, the more under-sized a test statistic is the less likely we are to reject the null hypothesis that there is no significant difference in the accuracy of two forecasts. In this sense, under-sizing can be seen as a conservative approach. This same interpretation of under-sizing cannot be made in an encompassing context; the more under-sized an encompassing test statistic is, the less likely we are to reject the null that one forecast encompasses the other, and consequently the more likely we are to exclude a forecast rather than use it in forecast combination. In this sense, failing to reject the null hypothesis is the more ‘extreme’ result. It could be argued that in the forecast encompassing case, under-sizing is equally as undesirable as over-sizing.

4.4 Frequency of negative long-run variance estimates

The long-run variance estimator (4.1), based on the rectangular kernel, is not guaranteed to be positive whenever $h > 1$. In practice, of course, a negative outcome is highly problematic since the DM, HLN_a and HLN_b statistics for testing equal accuracy or encompassing cannot be computed. In such circumstances, a practitioner must then decide how to deal with such a result. DM suggest treating such estimates as zero or using an alternative long-run variance estimator that guarantees positivity. Whatever strategy is followed will have implications for the size and power of the resulting testing procedure, so it is therefore valuable to quantify how frequently negative long-run variance estimates are likely to be encountered in practice. While DM, Clark (1999) and Coroneo and Iacone (2015), *inter alios*, note that (4.1) can produce a negative result, little evidence has so far been provided as to the extent of this potential problem. To shed more light on the issue, in this section we report results from Monte Carlo simulation experiments to determine the frequency with which negative long-run variance estimates arise for different sample sizes and forecast horizons, both for equal accuracy and encompassing tests.

In the case of testing for equal forecast accuracy, we assume mean squared error loss, so that $d_t = e_{1t}^2 - e_{2t}^2$, $t = 1, \dots, T$, generating the forecast errors according to the following data generating process [DGP], which allows for h -

steps-ahead forecasts to follow moving average [MA] processes of order $h - 1$:

$$\begin{aligned} e_{1t} &= v_{1t} + \sum_{j=1}^{h-1} \theta v_{1,t-j} \\ e_{2t} &= \sqrt{R} \left(v_{2t} + \sum_{j=1}^{h-1} \theta v_{2,t-j} \right) \end{aligned}$$

where $[v_{1t}, v_{2t}]' = P[u_{1t}, u_{2t}]' \sim N(0, Q)$ where Q is the contemporaneous correlation matrix

$$Q = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

with $\rho \in [0, 1)$. The Choleski factor is:

$$P = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1 - \rho^2} \end{bmatrix}$$

yielding the contemporaneously correlated series

$$v_{1t} = u_{1t} \tag{4.4}$$

$$v_{2t} = \rho u_{1t} + \left(\sqrt{1 - \rho^2} \right) u_{2t} \tag{4.5}$$

where $[u_{1t}, u_{2t}]' \sim N(0, I_2)$, $t = 1, \dots, T$.

The DGP adopts the simplified approach of setting all MA parameters to the same value. The ratio of the variances of the two forecast errors is given by $R > 0$, with $R = 1$ giving the null and $R \neq 1$ the alternative. Focusing on the small samples that are often employed in forecast evaluation exercises, we simulate this DGP for $T = \{8, 16, 32, 64\}$, $h = \{2, 3, 4, 5, 6\}$, $\rho = \{0, 0.5, 0.9\}$, $\theta = \theta^*/(h-1)$ with $\theta^* = \{0, 0.5, 0.9\}$, and calculate the frequency with which negative values of the long-run variance estimator (4.1) arise. Here and throughout the paper, simulations are conducted using 10,000 Monte Carlo replications. Table 4.1 reports the results under the null ($R = 1$), and under the alternative ($R > 1$), with the settings $R = 12$, $R = 7$, $R = 3$ and $R = 2$ for $T = 8$, $T = 16$, $T = 32$ and $T = 64$, respectively (chosen to ensure that the test powers considered in section 4.6 are roughly comparable across sample sizes).

As might be expected, we find negative long-run variance estimates occur with a frequency that increases with the forecast horizon, and decreases with the sample size. While the occurrence of negative estimates is rare when $T = 64$, the problem can be substantial for the smaller sample sizes considered, particularly for longer forecast horizons where the frequency can rise above 40%. In such

circumstances, a practitioner would be unable to compute the standard DM or HLNa test statistics almost half the time. The pattern of frequencies for negative long-run variance estimates has very little dependence on whether the simulations are conducted under the null or alternative hypotheses, or on the contemporaneous correlation parameter ρ , with no clear pattern emerging as ρ increases. An increase in the moving average parameter θ^* results in a slight reduction in the frequency of negative estimates (for example, when $T = 8$, $h = 6$ and $\rho = 0$, the frequency of negative estimates is 42.2% for $\theta = 0$, 41.6% for $\theta = 0.5$, and 40.6% for $\theta = 0.9$) but the overall features of the results are similar to the serially uncorrelated case. These results highlight a potentially serious issue with the implementation of standard tests for equal forecast accuracy in small samples.

Turning now to testing for forecast encompassing, we let $d_t = e_{1t}(e_{1t} - e_{2t})$, $t = 1, \dots, T$, where the forecast errors are generated according to the following DGP, again allowing for MA($h - 1$)-dependence in the errors of h -steps-ahead forecasts:

$$e_{it} = v_{it} + \sum_{j=1}^{h-1} \theta v_{i,t-j}, \quad i = 1, 2$$

where $\theta = \theta^*/(h - 1)$ and

$$\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \sim N \left(0, \begin{bmatrix} 1 & \rho \\ \rho & \kappa^2 \end{bmatrix} \right), \quad t = 1 - (h - 1), \dots, T$$

Note that $E[e_{1t}^2] = 1$ and $E[e_{1t}e_{2t}] = \rho$. The null hypothesis that forecast f_{1t} encompasses f_{2t} is therefore obtained by setting $\rho = 1$, while a setting of $\rho < 1$ gives the alternative. To ensure that the correlation between the forecast errors is less than one in absolute value, we require $Var[e_{2t}] > Cov[e_{1t}, e_{2t}]^2$, and therefore set $\kappa^2 > \rho^2$. Under the alternative, Harvey et al. (1998) show that power depends only on the single parameter $k = \sqrt{\kappa^2 - \rho^2}/(1 - \rho)$. Table 4.2 reports results for the frequency of negative long-run variance estimates using the same settings for T , h and θ^* as in Table 4.1. Results are reported under both the null and alternative, with the settings for k under the alternative being $k = 1.25$, $k = 2.00$, $k = 3.00$ and $k = 4.50$ for $T = 8$, $T = 16$, $T = 32$ and $T = 64$, respectively (again chosen to broadly align the test power levels considered in section 4.6 across sample sizes).

The pattern of negative estimates for the long-run variance is very similar in the case of testing for forecast encompassing to that for testing for equal forecast accuracy. Indeed, on comparing Tables 4.1 and 4.2 for a given combination of T , h and θ^* , it is clear that the numerical frequencies are very close to each other,

suggesting that the prevalence of negative long-run variance estimates is driven more by the interplay of sample size and the number of estimated autocovariances $(h - 1)$ than by the precise form of d_t . We again see a rising incidence of negative estimates as T decreases and as h increases, with frequencies above 10% for $h > 2$ when $T = 8$ and for $h > 3$ when $T = 16$. As with the equal accuracy results, the role of serial correlation is small, and it makes little difference whether the long-run variance is being calculated under the null or alternative. The overall finding is that negative long-run variance estimates can occur with very high probability for equal accuracy and encompassing tests when using multi-step-ahead forecasts with small, yet empirically relevant, sample sizes.

4.5 Adjusted Diebold-Mariano type tests

Given the prevalence of negative long-run variance estimates that arise for multi-step-ahead forecasts in small samples when using the standard long-run variance estimator in the DM-type tests, it is important to establish methods for dealing with this potential problem. In this section we consider a number of possible approaches, all based on the DM-type tests for equal accuracy and encompassing. The following section then evaluates their relative performance in terms of finite sample size and power.

The first approach we consider is the suggested method of DM in the equal accuracy testing context, which is to treat any occurrence of a negative long-run variance as a zero, viewing the negative estimate as indicative of a very small long-run variance. This of course implies a test statistic of $\pm\infty$, depending on the sign of the numerator $\bar{d}$. In a two-sided testing context, as in DM, such a treatment induces an immediate rejection of the null hypothesis, so a negative long-run variance estimate always indicates evidence in favour of the alternative hypothesis under this approach. When testing against a one-sided alternative, as is common in applications of equal accuracy tests and always the case when testing for encompassing, treating a negative long-run variance as zero will either induce automatic rejection or non-rejection, depending on whether the implied test statistic value of $+\infty$ or $-\infty$ lies in the relevant one-tailed critical region. Applying this approach to the DM test, we can express the method as

$$DM_{rej} = \begin{cases} DM & \text{if } \hat{\omega}^2 > 0 \\ \text{sign}(\bar{d}) \times \infty & \text{otherwise} \end{cases}$$

with the test statistic to be compared to standard normal critical values, and

application to the MDM test of HLN_a and HLN_b gives

$$MDM_{rej} = \begin{cases} MDM & \text{if } \hat{\omega}^2 > 0 \\ \text{sign}(\bar{d}) \times \infty & \text{otherwise} \end{cases}$$

with the test statistic to be compared with t_{T-1} critical values.

Given the frequency with which negative long-run variance estimates can occur, the DM_{rej} and MDM_{rej} approaches will induce substantial over-size in two-sided equal accuracy testing procedures for $h > 1$ and small T , as all occurrences of a negative $\hat{\omega}^2$ trigger a rejection of the null. A similar, albeit reduced, feature of over-size would also be expected for one-sided equal accuracy tests and tests for forecast encompassing, with rejections of the null occurring whenever a negative $\hat{\omega}^2$ coincides with the appropriate sign of $\bar{d}$. In addition, the results in Section 4.4 demonstrate that a negative variance estimate is no more likely to be obtained under the alternative hypothesis, than under the null hypothesis. A simple conservative approach which would avoid the substantial over-sizing problem likely to be exhibited by DM_{rej} and MDM_{rej} is to treat the occurrence of a negative long-run variance estimate as a failure to correctly estimate the true long-run variance, and default to *non-rejection* of the null in such instances. One way of writing such a method would be to define the adjusted MDM test statistic as

$$MDM_{non} = \begin{cases} MDM & \text{if } \hat{\omega}^2 > 0 \\ 0 & \text{otherwise} \end{cases}$$

with the test statistic again being compared with t_{T-1} critical values. A potential down-side of this approach is that the greater size control afforded by treating negative estimate cases as non-rejections is also likely to be associated with low power under the respective test alternative. It is well established that the MDM approach of HLN_a and HLN_b leads to greater size control than the DM test in small samples, thus we consider an MDM version of the test statistic only.

An ad-hoc approach to dealing with negative long-run variance estimates is to use an adaptive bandwidth such that we initially estimate (4.1) with a bandwidth of $h - 1$, but then reduce the bandwidth if a negative estimate occurs. We suggest doing this sequentially, such that if a bandwidth of $h - 1$ results in a negative estimate, (4.1) is re-estimated with a bandwidth of $h - 2$, and so on. We continue to reduce the bandwidth until either the estimate is positive, or the bandwidth is equal to zero such that the short-run variance estimate, $\hat{\gamma}_0$, is used. As we no longer always use $h - 1$ autocovariances in the estimate of (4.1), we require a different bias correction than that of HLN_a/HLN_b. The bandwidth of the long-

run variance estimator is now given by h' , where h' is an integer between 0 and $h - 1$. Substituting $h' = h - 1$ into the MDM bias correction, yields the new test statistic

$$MDM_{h'} = \sqrt{\frac{T - 1 - 2h' + T^{-1}h'(h' + 1)}{T}} \left(\frac{\bar{d}}{\sqrt{\hat{\omega}_{h'}^2}} \right)$$

where

$$\hat{\omega}_{h'}^2 = \hat{\gamma}_0 + 2 \sum_{j=1}^{h'} \hat{\gamma}_j$$

and this testing strategy can be written as

$$MDM_{adapt} = \begin{cases} MDM & \text{if } \hat{\omega}^2 > 0 \\ MDM_{h'} & \text{otherwise} \end{cases}$$

with the test compared to t_{T-1} critical values. Given the superior size properties of MDM over DM , we evaluate the size and power of MDM_{adapt} only.

Another simple approach is to deal with a negative long-run variance estimate by replacing it directly with the corresponding short-run variance estimate $\hat{\gamma}_0$, thereby reducing the bandwidth in (4.1) from $h - 1$ to zero. While this approach neglects the impact of autocorrelation terms, it can be argued that the very presence of a negative estimate indicates that estimation of such components is highly unreliable in these situations. When the short-run variance estimator is used, the appropriate bias correction in the MDM statistic is that for $h = 1$, i.e.

$$MDM_0 = \sqrt{T - 1} \left(\frac{\bar{d}}{\sqrt{\hat{\gamma}_0}} \right)$$

and the overall test that adopts this statistic when a negative long-run variance is encountered can be written as

$$MDM_{SR} = \begin{cases} MDM & \text{if } \hat{\omega}^2 > 0 \\ MDM_0 & \text{otherwise} \end{cases}.$$

Critical values from the t_{T-1} distribution are again to be used. Of course, when $h = 2$, MDM_{SR} and MDM_{adapt} are the same testing procedure, as both will use MDM if $\hat{\omega}^2 > 0$ and default to using the short-run variance otherwise.

While the above methods replace negative long-run variances with simple decision rules or reduced bandwidth estimates, the following approaches we consider retain a proper estimate of the long-run variance, but make use of estimators that impose positivity. An obvious possibility in this class is to replace the rectangular

kernel in (4.1) with the Bartlett kernel, i.e.

$$\hat{\omega}_B^2 = \hat{\gamma}_0 + 2 \sum_{j=1}^m \left(1 - \frac{j}{m+1}\right) \hat{\gamma}_j$$

where m denotes the bandwidth. The Bartlett kernel therefore imposes the largest weight on the first sample autocovariance and decreasing weights on all subsequent autocovariances. A range of other kernels exist which also guarantee a positive long-run variance estimate, such as the quadratic-spectral (Andrews, 1991) kernel. However, Newey and West (1994) show that the choice of kernel is of less importance for size results than the choice of bandwidth, m . For consistency with the optimal forecast-motivated choice of truncation $h - 1$ in (4.1), along with the fact that use of the Bartlett kernel is most likely to arise in small samples, we set the Bartlett bandwidth to $m = h - 1$. As the HLN_a bias correction does not apply to $\hat{\omega}_{Bart}^2$ (and an equivalent bias correction is not possible to obtain without effectively reducing $\hat{\omega}_{Bart}^2$ to $\hat{\omega}^2$), we define the test statistic

$$DM_b = \sqrt{T} \left(\frac{\bar{d}}{\hat{\omega}_{Bart}} \right).$$

which is compared to standard normal critical values.

Clark (1999) considered using the Bartlett kernel with Newey-West and pre-whitened Newey-West bandwidth selection. While Clark's simulations abstracted from issues of negative variance estimation, it was found that a Bartlett-based approach could result in greater finite sample over-size than when using the rectangular kernel. This suggests that it is not recommended to use the Bartlett kernel in all circumstances, particularly when the rectangular kernel does not have negative variance estimate problems. We therefore also consider a hybrid approach, whereby the standard MDM test is used provided the long-run variance estimate is positive, but in the case of a negative estimate, the statistic switches to one based on the Bartlett kernel. We define this testing approach as follows

$$MDM_b = \begin{cases} MDM & \text{if } \hat{\omega}^2 > 0 \\ DM_b & \text{otherwise} \end{cases}$$

with t_{T-1} critical values.

The original DM test, and the variants outlined above, all make use of weighted sample autocovariances in the long-run variance estimator. An alternative method for dealing with the problem of negative long-run variance estimates is proposed by Hering and Genton (2011) in the spatial statistics literature. They propose fitting an exponential covariogram to the sample autocovariances thus guarantee-

ing the long-run variance estimate is positive. Similar to the Bartlett kernel, the covariogram imposes the largest weight on the first sample autocovariance and decreasing weights on all subsequent autocovariances. However, whilst the Bartlett weights decline linearly, Hering and Genton (2011) opt for an exponential decline such that proportionally less weight is given to higher order autocovariances here compared to the Bartlett kernel. Denoting a vector of the autocovariances, γ_k , as $C(k)$, they suggest that an exponential covariogram of the form:

$$C(k) = \sigma^2 \exp(-3k / \psi)$$

is fitted, where the parameters σ^2 and ψ are estimated using nonlinear least squares. The restrictions $\sigma^2 > 0$ and $\psi > 0$ are imposed to ensure that the exponential covariogram is giving decreasing rather than increasing weight to the sample autocovariances as k increases. Given that fewer observations are available to compute the autocovariances at higher lags, only half the autocovariances are used to fit the covariogram, such that $k = \{1, \dots, (T-1)/2\}$. This is a common rule of thumb in spatial statistics. The long-run variance estimator is then:

$$\hat{\omega}_{cov}^2 = \hat{C}(0) + 2 \sum_{k=1}^{T-1} \hat{C}(k)$$

where the fitted coefficients from the nonlinear least squares regression are used to compute $\hat{C}(0)$ and $\hat{C}(k)$, where all lags of the sample autocovariances are now used in the estimation of the long-run variance.

This spatial statistics approach does not allow the forecast horizon, h , to play any role in determining the value of the test statistic. We therefore adjust the method of Hering and Genton (2011) so that it is consistent with the time series approach used by the other test statistics considered. Instead of using $(T-1)/2$ sample autocovariances to compute the fitted model, and then using $T-1$ lags to calculate $\hat{C}_k$, we choose to use $h-1$ lags for both the fitted model and the estimation of $\hat{C}_k$, such that:

$$\hat{\omega}_{HG}^2 = \hat{C}(0) + 2 \sum_{k=1}^{h-1} \hat{C}(k)$$

This test statistic is given by

$$DM_{HG} = \sqrt{T} \left(\frac{\bar{d}}{\hat{\omega}_{HG}} \right)$$

which is compared to standard normal critical values. The size properties of a test statistic using $\hat{\omega}_{cov}^2$ rather than $\hat{\omega}_{HG}^2$ were found to be very similar, therefore we report results only for DM_{HG} .

An alternative approach proposed by Coroneo and Iacone (2015) is to use a weighted periodogram estimator, and these authors recommend construction of a DM-type test using the estimator of Hualde and Iacone (2017). Denoting the periodogram of d_t for Fourier frequency $\lambda_j = 2\pi j/T$ by

$$I(\lambda_j) = \left| \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T d_t e^{-i\lambda_j t} \right|^2$$

with i the imaginary unit, they suggest use of the Daniell kernel with bandwidth m to construct the weighted periodogram estimator of the long-run variance

$$\hat{\omega}_{Dan}^2 = 2\pi \frac{1}{m} \sum_{j=1}^m I(\lambda_j)$$

which is then used to construct the DM-type test statistic

$$DM_{CI} = \sqrt{T} \left(\frac{\bar{d}}{\hat{\omega}_{Dan}} \right).$$

If the bandwidth is treated as fixed, $\hat{\omega}_{Dan}^2$ is not a consistent estimator of ω^2 , but is asymptotically unbiased, and under the null hypothesis of $E(d_t) = 0$, DM_{CI} follows an asymptotic t_{2m} distribution. This fixed- m treatment results in a test with appealing finite sample properties, offering better size control relative to the $m \rightarrow \infty$ treatment that results in standard normal limit theory. Coroneo and Iacone (2015) observe that the t_{2m} distribution can act as a better approximation of the true null distribution for a smaller bandwidth, whereas larger bandwidths can be associated with higher power, hence a size-power trade-off emerges. Following these authors, we set the bandwidth according to $m = \lfloor T^{1/3} \rfloor$, but for any given sample size, m is then treated as a fixed number so that the fixed- m asymptotic theory can be applied, with critical values drawn from the t_{2m} distribution.

4.6 Finite sample size and power

In this section we consider the finite sample performance of the different methods outlined in Section 4.5 by evaluating both their empirical size and their size-adjusted power.

4.6.1 Forecast accuracy

We first consider testing for equal forecast accuracy, again focusing on mean squared error loss ($d_t = e_{1t}^2 - e_{2t}^2$), and simulate the empirical sizes of the DM_{rej} , MDM_{rej} , MDM_{non} , MDM_{adapt} , MDM_{SR} , DM_b , MDM_b , DM_{HG} and DM_{CI} testing approaches, with the tests conducted against a two-sided alternative at the nominal 0.10-level. Tables 4.3 - 4.5 report the sizes for the same simulation DGPs that were considered in the negative long-run variance simulations of Section 4.4 when the null hypothesis is imposed ($R = 1$) for contemporaneous correlation settings $\rho = 0$, $\rho = 0.5$ and $\rho = 0.9$ respectively.

Consider first Table 4.3 where $\rho = 0$. When $h = 1$, the long-run variance estimate is equivalent to the short-run variance, and therefore DM_{rej} and DM_b are equivalent test statistics. We also note that the estimate of $C(0)$ used in the DM_{HG} procedure must be very similar to that obtained for the short-run variance γ_0 , and therefore DM_{HG} has equivalent size to DM_{rej} and DM_b . Additionally the original MDM statistic cannot suffer from negative long-run variance estimation problems when $h = 1$, so MDM_{rej} , MDM_{non} , MDM_{SR} , MDM_b all amount to simply conducting MDM . Note also that when $h = 1$, no serial correlation is present in the DGP, hence the θ^* setting plays no role. Here, DM_{rej} , DM_b and DM_{HG} exhibit a degree of over-sizing at all sample sizes, with the severity of over-sizing increasing as T decreases. In contrast, the five MDM -type statistics are well behaved, with sizes very close to the nominal level, with only modest under-size displayed for $T = 8$ and $T = 16$. A very similar pattern of size behaviour is also seen for DM_{CI} . Therefore, the five MDM tests and DM_{CI} are reliable for one-step-ahead forecasts and there is little to choose between them in terms of finite sample size.

For multi-step-ahead forecasts ($h > 1$), the possibility of negative long-run variance estimates arises and so the method of dealing with these problem cases results in different size properties for the overall procedures that we consider. The DM_{rej} and MDM_{rej} approaches translate any negative long-run variance estimate into a rejection of the null, thus the high frequency of negative estimates for larger h and smaller T induces a high degree of over-size. In line with the results of Table 4.1, the size of DM_{rej} reaches almost 0.75, whilst the size of MDM_{rej} reaches almost 0.50, and such large upward size distortions render these procedures invalid. The DM_b test, which uses the Bartlett kernel, also exhibits a severe degree of over-sizing, with size over 0.40 in the most extreme case considered in this chapter. The MDM_b hybrid method achieves better size control through use of the Bartlett kernel only in problem cases and the MDM approach elsewhere, but is again subject to quite substantial over-size for mod-

erate values of h and T , with empirical size in excess of 0.20 in the worst cases. Likewise, the sequential bandwidth reduction test, MDM_{adapt} , also exhibits size as high as approximately 0.20 for the longer forecast horizons considered here. Of the MDM -based approaches, the best performing methods are MDM_{non} , which translates negative variance estimates into non-rejections of the null, and MDM_{SR} , which replaces negative long-run variance estimates with a short-run variance estimate. These procedures have similar sizes for most combinations of h and T . Where differences in size occur (typically for $T = 8$ and to a lesser extent for $T = 16$) MDM_{non} always displays a lower size than MDM_{SR} , as would be expected. Although MDM_{non} and MDM_{SR} are the best of the MDM -based methods, the sizes can still be inflated above the nominal level, with maximum size of around 0.16 occurring.

For a given degree of serial correlation, the size of DM_{HG} , which estimates the long-run variance using the Hering and Genton (2011) covariogram approach, is almost identical across $h = 2 - 6$. Figure 4.1 gives the average values of $\widehat{C}(k)$ and the sample autocovariances $\widehat{\gamma}_k$ given from 10,000 Monte Carlo replications of the forecast errors (as described in Section 4.4) for $T = 8$, $h = 6$ where no contemporaneous or serial correlation is present. The exponential covariogram gives a value for $\widehat{C}(0)$ that is very close to the sample short-run variance $\widehat{\gamma}_0$. However, for all subsequent $\widehat{C}(k)$ the values are smaller in absolute value than the sample autocovariances $\widehat{\gamma}_k$ and decline to zero much faster. This suggests that, for multi-step ahead forecasts, increasing h (which implies increasing the bandwidth of the long-run variance estimator) has a minimal effect on the value of the long-run variance estimate and therefore on the size properties of the DM_{HG} test statistic. Conversely, we see that the degree of serial correlation does impact the properties of the test, with the over-size of DM_{HG} being exacerbated by high degrees of serial correlation. For example, considering $T = 8$ and $h = 3$, DM_{HG} has size of 0.146 when $\theta = 0$ and 0.211 when $\theta = 0.9$. This over-sizing under high degrees of serial correlation makes the test statistic unattractive in small samples. In contrast, the DM_{CI} weighted periodogram approach achieves excellent size control across h , T and θ . Apart from some under-size for $T = 8$, the test generally has size very close to 0.10, with the worst upward size distortion being a size of 0.111, offering a clear improvement over the other methods considered. The attractive finite sample size results reported in Coroneo and Iacone (2015) for moderate sample sizes and forecast horizons therefore extend to the small sample and longer horizon region under focus here, suggesting a valuable role for DM_{CI} in delivering forecast accuracy tests with reliable size in small samples. Note that the role of serial correlation in these simulated sizes appears relatively

small, with very similar patterns observed across θ^* .

We note that the sizes of the MDM_{non} , MDM_{adapt} and MDM_{SR} initially increase as T increases before decreasing again, i.e. when $h = 4$, $\rho = 0$ and $\theta = 0$, MDM_{SR} has sizes of 0.112 ($T = 8$), 0.154 ($T = 16$) and 0.130 ($T = 64$). This is the result of the bias correction applied to MDM test procedures which is non-linear in T . We can demonstrate this by plotting the bias correction factor (CF) against T for different forecast horizons where

$$CF = T(T + 1 - 2h + (h(h - 1)/T))^{-1}$$

Figure 4.2 show the bias correction factor against T for $h = 1 - 6$ respectively. The bias correction is evidently a convex function of T , so a larger correction is applied when $T = 8$ than when $T = 16$. As T increases, the curve flattens, so the effect of the bias correction becomes less pronounced. As the forecast horizon increases, the bias correction curves become steeper and therefore the nonlinear behaviour of CF across T is more apparent for longer forecast horizons. Whilst we might expect this behaviour to also be present for MDM_b and MDM_{rej} , the size distortions encountered by the test procedures outweigh the effect of the MDM bias correction.

In Table 4.4, we consider the case of $\rho = 0.5$ so that the two forecast error series are contemporaneously correlated. The size properties of tests are very similar to that displayed in Table 4.3, with the same general ranking of test procedures arising. In Table 4.5, where we set $\rho = 0.9$, again the sizes of all test procedures considered are consistent with the results found in the no contemporaneous correlation case. All the DM-type tests considered in this chapter appear robust to the degree of contemporaneous correlation in the forecast errors.

As discussed in Section 4.2, it may be unreasonable to assume that forecast errors are normally distributed. Following Harvey et al. (1997) we simulate heavy-tailed errors by generating u_{1t} and u_{2t} in 4.4 as two independent t_6 random variables

$$u_{it} = \frac{z_{it}}{\sqrt{\chi_{it,6}^2/6}}$$

where $[z_{1t}, z_{2t}]' \sim N(0, I_2)$ and z_{it} is independent of $\chi_{it,6}^2$ for $i = \{1, 2\}$. Contemporaneous correlation and serial correlation are then incorporated as before. The size properties of the tests considered in this chapter under heavy-tailed errors are displayed in Table 4.6. We consider the no contemporaneous correlation case only given the similarity of sizes over different degrees of contemporaneous correlation encountered in Tables 4.4-4.5. In general, using heavy-tailed errors

appears to have slightly decreased the sizes of tests, regardless of the degree of serial correlation present in the forecast errors. Consequently, DM_{CI} now exhibits some under-sizing in small samples, with sizes ranging from 0.061 to 0.098 across the h , T and θ settings considered in this chapter. Nevertheless, this testing procedure still offers empirical size closest to nominal size across almost all settings. Both MDM_{non} and MDM_{SR} have empirical sizes reasonably close to nominal size with MDM_{non} having size ranging from 0.046 to 0.149, and MDM_{SR} having size ranging from 0.068 to 0.161. As in the case of normally distributed errors, DM_{rej} , MDM_{rej} , DM_b and MDM_b are very over-sized in small samples, whilst DM_{HG} suffers from a large degree of over-sizing when the degree of serial correlation in the errors is high. The general size ranking of tests obtained when using normally distributed forecast errors seems robust to heavy-tailed forecast errors.

We also consider the possibility that forecast errors are negatively serially correlated by setting $\theta^* = \{0, -0.5, -0.9\}$ (i.e. the negative numbers that correspond to the positive degrees of serial correlation considered previously). Size results under negative degrees of serial correlation are reported in Table 4.7. The sizes of all tests when negative serial correlation is present in the forecast errors are broadly similar to their positive serial correlation counterparts. In general, the presence of negative serial correlation causes a slight reduction in size (relative to the positive serial correlation case). For example, when $T = 8$ and $h = 6$, DM_{CI} has size of 0.102 for $\theta = 0.9$ and 0.081 for $\theta = -0.9$. Nevertheless, the overall pattern of results holds and DM_{CI} has empirical size closest to nominal size across almost all settings of T , h and θ .

In addition to evaluating the empirical sizes of the procedures, it is also important to assess their relative powers. It is only possible to directly compare the power of different test statistics when those tests are correctly sized. Clearly, this is not the case with many of the tests considered in this chapter. We therefore use the test statistics themselves to directly simulate empirical 0.10 critical values. When a test is over-sized, these empirical critical values will be larger in absolute value than the nominal 0.10 critical values taken from the appropriate normal or Student- t distributions; whereas, when a test is under-sized, the opposite is true. Given that the DM/HLNa tests use a two-sided alternative, we compute both upper and lower critical values to account for any asymmetry in the distributions of the forecast errors. However, we note that for all test statistics the absolute value of upper and lower critical values are very similar to each other, suggesting a symmetric distribution is appropriate. These empirical critical values are used to compute size-adjusted power for the test statistics, allowing direct comparison of power across different tests. The DM_{rej} and MDM_{rej} approaches

are not amenable to size-adjustment due to the high proportion of automatic rejections induced by negative long-run variance estimates being treated as zero, which cannot be corrected by adjusting critical values. Given their unfavourable size properties, these tests would not be considered candidate tests regardless of their power performance. Likewise, the substantial over-sizing exhibited by DM_{HG} when a high degree of serial correlation is present in the forecast errors excludes this test as a reliable procedure. Additionally, we note that in regards to obtaining empirical size closest to nominal size, MDM_b dominates DM_b and MDM_{SR} dominates MDM_{adapt} across almost all settings of T , h , θ and ρ . We therefore exclude DM_b and MDM_{adapt} from our power analysis. Table 4.8 reports the size-adjusted powers of the equal accuracy test procedures MDM_{non} , MDM_{SR} , MDM_b and DM_{CI} .³

The DGPs are again those used in Section 4.4, with R varied across T to keep the power levels broadly similar across different sample sizes. When $h = 1$, where MDM_{non} , MDM_{SR} and MDM_b simply reduce to MDM , it is clear that the original MDM test can offer decent power gains over DM_{CI} , particularly for smaller samples where gains over 0.16 are seen. Thus in the one-step-ahead context, where the tests are correctly sized and no negative long-run variance estimation problems arise, use of MDM is to be recommended.

When $h > 1$, however, this ranking tends to be reversed, with DM_{CI} becoming the most powerful test. In particular, when $T = 8$, DM_{CI} outperforms all the MDM -based procedures for all forecast horizons, with worthwhile power gains of up to 0.13 displayed. For larger sample sizes, DM_{CI} power can dip a little below that of the MDM -based approaches when h is small, but for the longer forecast horizons, DM_{CI} again offers decent power gains. Among the three MDM -based methods, power differences emerge for smaller values of T , becoming more exaggerated as h increases, with MDM_{non} displaying the lowest relative power (as expected given its conservative approach to dealing with negative long-run variance estimates), followed by MDM_b and then MDM_{SR} . Given that MDM_{SR} was one of the better size controlled MDM -based approaches, this suggests MDM_{SR} to be the best of these modifications. Considering the impact of serial correlation in the errors, the power levels of all the methods decrease as θ^* increases, but the relative rankings of the tests are preserved, hence the above comments apply for any given setting of θ^* .

Taking the multi-step-ahead size and power results together, DM_{CI} therefore

³We considered the effect of contemporaneous correlation on the power of these test procedures but found the key findings to be unchanged and therefore report power results for the no contemporaneous correlation case only.

emerges as the best performing test overall, with reliable finite sample size performance and relatively high levels of power. While MDM_{SR} can suffer from some moderate size distortion, its relative power levels are reasonably competitive, and it represents the best of the MDM -based procedures. Given the fact that MDM_{SR} requires almost no additional computation compared to MDM (as $\hat{\gamma}_0$ is computed as part of $\hat{\omega}^2$), a possible role for this approach could be envisaged for practitioners who desire a simple approach that remains in the MDM -based framework. Otherwise, it is clear that DM_{CI} should be employed whenever $h > 1$.

4.6.2 Forecast encompassing

We next consider testing for forecast encompassing, beginning by simulating the empirical sizes of the DM_{rej} , MDM_{rej} , MDM_{non} , MDM_{adapt} , MDM_{SR} , DM_b , MDM_b , DM_{HG} and DM_{CI} procedures, conducting one-sided tests at the nominal 0.10-level for the same simulation DGPs as in Section 4.4. The results are reported in Table 4.9. As for the equal accuracy case, for one-step-ahead forecasts, we observe that the MDM -based procedures (which are identical for $h = 1$) and the DM_{CI} test display good finite sample size control. Indeed, the MDM test has almost no size distortions even for small T , while only a very modest amount of under-size is displayed for DM_{CI} . The DM_{rej} , DM_b and DM_{HG} procedures (which are also identical for $h = 1$) exhibit a degree of over-sizing for smaller values of T , with maximum size of 0.142.

When $h > 1$, we find a similar picture of size behaviour to that in Table 4.3. Specifically, DM_{rej} , MDM_{rej} and DM_b can again be substantially over-sized for small T and large h , whereas the size of DM_{HG} worsens as the degree of serial correlation increases. DM_{CI} has the best size control, with MDM_{non} , MDM_{SR} , MDM_{adapt} and MDM_b lying inbetween. Note that the over-size of DM_{rej} and MDM_{rej} is less severe than for the equal accuracy case, since here the encompassing test is conducted against a one-sided alternative, hence only a proportion of the negative long-run variances obtained induce a rejection of the null. When $\theta^* = 0.9$, DM_{CI} can suffer from a small amount of upward size distortion, with size around 0.12 for $T = 8$, and for $T = 16$ when $h > 3$, but overall this test still outperforms the other approaches. MDM_{non} , MDM_{SR} , MDM_{adapt} and MDM_b are subject to maximum sizes of around 0.13, 0.14, 0.15 and 0.16 respectively, a little lower than the corresponding size maxima under equal accuracy. DM_{HG} still exhibits an unfavourable degree of over-sizing for high degrees of serial correlation, although this is lower than that exhibited for equal accuracy, with maximum size of 0.175.

Turning to power for forecast encompassing tests, Table 4.10 gives results for the size-adjusted powers of MDM_{non} , MDM_{SR} , MDM_b and DM_{CI} for the relevant DGPs of Section 4.4, with k varying across T as specified in that section. The relative power rankings of the tests are unchanged compared to tests for equal forecast accuracy, therefore the comments and conclusions outlined above are equally applicable in this context. We again find that MDM has a power advantage over DM_{CI} for $h = 1$, while DM_{CI} generally outperforms the other procedures for $h > 1$. MDM_{SR} again offers the best overall power of the MDM -based tests, with MDM_{non} worst due to its inherent conservatism. Once again, therefore, MDM is to be recommended for one-step-ahead forecasts, but for multi-step-ahead forecasts, where negative long-run variance estimates can arise, apart from a potential role for MDM_{SR} when a simple MDM -based modification is desired, it is the DM_{CI} test that is to be preferred, offering the best finite sample performance in terms of both size and relative power.

4.6.3 Simulations using calibrations from empirical data

To consider the extent to which our Monte Carlo exercise is representative of the issues an applied researcher faces in practice, we now focus on obtaining forecast errors from an empirical example. Dreger and Wolters (2014) forecast Euro-area inflation from an AR model using quarterly data. The forecast performance of this benchmark model is then evaluated against competing models which consider additional explanatory variables such as excess liquidity and excess inflation. Forecasts are obtained at 4, 8 and 12 step-ahead horizons, yielding 29, 25 and 21 out of sample forecast observations respectively. Broadly in line with this exercise, we forecast inflation using the same benchmark AR model in order to closer examine the properties of the forecast errors.

Quarterly HICP inflation data is obtained from the authors for the period 1981Q1 - 2010Q4. We construct 4, 8 and 12 quarter inflation rates as follows

$$\pi_t^k = \frac{4}{k} \log(pc_t / pc_{t-k}), \quad k = 4, 8, 12$$

where pc is the consumer price index (HICP). The forecasting equation is then given by

$$\pi_{t+k}^k = \alpha_1 \pi_t^k + \alpha_2 \pi_{t-1}^k + \alpha_3 \pi_{t-2}^k + \alpha_4 \pi_{t-3}^k + \epsilon_t \quad (4.6)$$

i.e. the benchmarking model of Dreger and Wolters (2014). For all forecast horizons we estimate the model over the initial sample 1984Q1 - 2002Q4 such that $t = [1984Q4 : 2001Q4]$ for $h = 4$, $t = [1984Q4 : 2000Q4]$ for $h = 8$ and

$t = [1984Q4 : 1999Q4]$ for $h = 12$. The parameter estimates are then used to obtain the first forecast for periods 2003Q4, 2004Q4 and 2005Q4 for horizons of 4, 8 and 12 steps-ahead respectively. We then extend the sample period by one observation and re-estimate (4.6) to obtain the next forecast for 2004Q1 at $h = 4$, 2005Q1 at $h = 8$ and 2006Q1 at $h = 12$. We continue in this way until the final forecast at time period 2010Q4 (for all forecast horizons) is obtained, therefore producing a total of 29 forecasts at a one-year horizon, 25 at a two-year horizon, and 21 at a three-year horizon. Three forecast error series, corresponding to the three different forecast horizons considered, are then given by

$$fe_t^k = \pi_t^k - \hat{\pi}_t^k$$

Theory suggests that forecast errors should follow an $MA(h-1)$ specification. We therefore use exact maximum likelihood to fit an MA model to each forecast error series. For all horizons, a maximum MA order of 8 is considered. For fe_t^{12} , this is below the theoretically correct lag order of $h-1$. The exact maximum likelihood algorithm involves computing the inverse Hessian matrix. The Hessian can be ill-conditioned when the sample size is small relative to the dimension of the matrix. The exact ML algorithm therefore fails to fit the theoretically appropriate $MA(11)$ model due to the small number of observations available relative to the lag order. In all cases AIC is used to select the appropriate model. We therefore find that the best fit is an $MA(3)$ for fe_t^4 , and an $MA(4)$ for both fe_t^8 and fe_t^{12} indicating that serial correlation is present in all three forecasts. Parameter estimates for these models are given in Table 4.11.

Considering the null hypothesis of equal forecast accuracy, we generate three forecast error series, using the following DGP:

$$\begin{aligned} e_{1t}^k &= \mu + v_{1t} + \theta_1 v_{1,t-1} + \dots + \theta_j v_{1,t-j} \\ e_{2t}^k &= \sqrt{R}(\mu + v_{2t} + \theta_1 v_{2,t-1} + \dots + \theta_j v_{2,t-j}) \end{aligned}$$

where $[v_{1t}, v_{2t}]' \sim N(0, I_2)$, $t = 1-j, \dots, T$ and j is the maximum lag order, i.e. $j = 3$ for e_{it}^4 where $i = 1, 2$ and $j = 4$ for both e_{it}^8 and e_{it}^{12} . The MA parameters θ_l , $l = 1, \dots, j$ and the mean μ correspond to the fitted parameters given in Table 4.11.

To examine the frequency at which negative values of the long-run variance estimator (4.1) occur we simulate e_{it}^4 for $T = 29$, e_{it}^8 for $T = 25$, and e_{it}^{12} for $T = 21$ under both the null (where $R = 1$) and the alternative, with the settings $R = 4$, $R = 6$ and $R = 8$ for e_{it}^4 , e_{it}^8 and e_{it}^{12} respectively. As before these settings of R were chosen such that simulated powers are broadly comparable

across sample sizes. The simulation exercise undertaken in Section 4.4 did not consider forecast horizons greater than $h = 6$, nor did we look at the exact settings of T used in this empirical example. To enable comparison between the two exercises, we therefore also report the frequency of negative variance estimates that are obtained by simulating the DGP in Section 4.4 with the same h and T settings as used here. These frequencies are reported for $\theta^* = 0$ only given that the proportion of negative variances was shown to be very similar across settings of θ^* . We refer to the errors obtained from the DGP simulated in Section 4.4 as *simulated errors* and those obtained from the empirical exercise above as *calibrated errors*. Table 4.12 reports the frequency of negative long-run variance estimates arising in both exercises. For $h = 4$, we see that the proportion of negative variances that arise for the calibrated errors are lower than that which we encounter for our simulated errors. However for both $h = 8$ and $h = 12$, the frequency of negative variances appears to be similar across both exercises. This demonstrates that encountering a high proportion of negative long-run variance estimates is an empirically relevant issue, and suggests that the simulated forecast errors that we analyse throughout this chapter are representative of the forecast errors one might obtain in practice.

We also examine the size and power properties of the MDM_{non} , MDM_{SR} , MDM_b and DM_{CI} tests for equal forecast accuracy when the calibrated forecast errors are used. Empirical size of nominal 0.10 level tests are reported in Table 4.13 for the same three combinations of h and T used above. As before, all tests excluding DM_{CI} exhibit a degree of over-sizing, whilst DM_{CI} provides size reasonably close to 0.10 at all horizons. At $h = 12$, DM_{CI} now has slightly larger size than MDM_{non} (0.120 and 0.099 respectively), but broadly these results reflect the superior size properties of DM_{CI} reported in Section 4.6. Finally size-adjusted power of 0.10 level tests under these calibrated errors is given in Table 4.14. DM_{CI} is shown to have higher power than the other tests considered at all forecast horizons. These results suggest that the simulation exercise in Section 4.4 is able to capture the issue of encountering negative long-run variance estimates that occurs in empirical forecasting exercises. This example confirms the promising size and power performance of DM_{CI} across all forecast horizons and sample sizes in practice.

4.7 Conclusion

In this chapter, we highlight that application of the standard DM-based tests for equal forecast accuracy and forecast encompassing can often result in a neg-

ative long-run variance estimate when dealing with multi-step-ahead predictions and small, but empirically relevant, sample sizes. Examining a number of possible approaches to dealing with this problem, we find that the recently proposed test of Coroneo and Iacone (2015), which uses a weighted periodogram long-run variance estimator combined with fixed-bandwidth asymptotics, offers the best overall finite sample size and power performance. Any size distortions associated with this test are very modest, while power levels are appealing relative to other approaches, permitting reliable inference even in the small sample/long horizon cases we consider. Aside from this preferred approach, a case could possibly be made for a strategy that uses the *MDM* tests of Harvey et al. (1997, 1998) when a positive long-run variance estimate is obtained, then switching to use of a short-run variance estimator whenever a negative long-run variance estimate occurs. While this approach does not perform as well as the Coroneo and Iacone (2015) procedure, it does have the advantage of simplicity, since no additional estimation is required to move from the long-run to the short-run variance estimator. Ultimately, the optimal choice of test statistic relies on the preferences of the practitioner and their desire for size control versus simplicity. Finally, when forecast evaluation is being done with one-step-ahead predictions, no negative long-run variance estimates can arise with the standard tests, and the *MDM* tests provide good size control and superior power to the Coroneo and Iacone (2015) test.

Table 4.1: Frequency of negative long-run variance estimates in tests for equal forecast accuracy

h	$\theta^* = 0.0$				$\theta^* = 0.5$				$\theta^* = 0.9$			
	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$
Panel A. $R = 1$ (null)												
$\rho = 0.0$												
2	0.074	0.015	0.001	0.000	0.050	0.006	0.000	0.000	0.036	0.003	0.000	0.000
3	0.175	0.064	0.013	0.000	0.154	0.048	0.007	0.000	0.131	0.032	0.003	0.000
4	0.275	0.120	0.031	0.002	0.254	0.101	0.025	0.001	0.231	0.077	0.016	0.001
5	0.352	0.175	0.055	0.008	0.340	0.162	0.047	0.005	0.325	0.141	0.034	0.004
6	0.422	0.217	0.085	0.017	0.416	0.201	0.076	0.013	0.406	0.186	0.060	0.008
$\rho = 0.5$												
2	0.072	0.014	0.001	0.000	0.049	0.008	0.000	0.000	0.037	0.003	0.000	0.000
3	0.178	0.060	0.010	0.001	0.154	0.048	0.006	0.000	0.127	0.031	0.004	0.000
4	0.272	0.123	0.032	0.004	0.255	0.103	0.024	0.002	0.228	0.079	0.014	0.001
5	0.357	0.170	0.055	0.010	0.347	0.155	0.045	0.007	0.331	0.136	0.032	0.003
6	0.422	0.218	0.084	0.017	0.419	0.206	0.077	0.015	0.403	0.183	0.062	0.010
$\rho = 0.9$												
2	0.074	0.015	0.001	0.000	0.053	0.006	0.000	0.000	0.039	0.003	0.000	0.000
3	0.176	0.067	0.011	0.000	0.153	0.050	0.006	0.000	0.124	0.030	0.003	0.000
4	0.273	0.125	0.030	0.002	0.249	0.109	0.022	0.002	0.232	0.084	0.012	0.000
5	0.359	0.169	0.055	0.010	0.347	0.153	0.048	0.005	0.330	0.128	0.035	0.003
6	0.418	0.220	0.088	0.019	0.419	0.212	0.075	0.015	0.408	0.182	0.059	0.010
Panel B. $R > 1$ (alternative)												
$\rho = 0.0$												
2	0.056	0.007	0.001	0.000	0.044	0.003	0.000	0.000	0.038	0.002	0.000	0.000
3	0.169	0.052	0.009	0.000	0.154	0.041	0.006	0.000	0.134	0.028	0.003	0.000
4	0.268	0.107	0.027	0.002	0.250	0.094	0.024	0.001	0.228	0.075	0.015	0.000
5	0.347	0.164	0.052	0.007	0.340	0.152	0.044	0.005	0.331	0.135	0.033	0.004
6	0.430	0.211	0.085	0.016	0.423	0.206	0.076	0.012	0.414	0.189	0.060	0.007
$\rho = 0.5$												
2	0.052	0.009	0.000	0.000	0.040	0.003	0.000	0.000	0.034	0.003	0.000	0.000
3	0.170	0.051	0.007	0.000	0.155	0.041	0.004	0.000	0.134	0.028	0.003	0.000
4	0.266	0.106	0.025	0.002	0.253	0.098	0.021	0.001	0.236	0.077	0.011	0.001
5	0.351	0.159	0.050	0.009	0.341	0.149	0.042	0.006	0.332	0.131	0.033	0.003
6	0.431	0.213	0.080	0.017	0.424	0.204	0.068	0.014	0.413	0.182	0.056	0.009
$\rho = 0.9$												
2	0.054	0.006	0.000	0.000	0.039	0.002	0.000	0.000	0.032	0.001	0.000	0.000
3	0.170	0.051	0.007	0.000	0.148	0.041	0.004	0.000	0.130	0.029	0.001	0.000
4	0.261	0.107	0.022	0.002	0.249	0.091	0.016	0.001	0.230	0.078	0.010	0.000
5	0.353	0.159	0.048	0.006	0.345	0.147	0.039	0.004	0.331	0.134	0.029	0.002
6	0.429	0.206	0.079	0.013	0.421	0.198	0.069	0.009	0.407	0.184	0.054	0.007

Table 4.2: Frequency of negative long-run variance estimates in tests for forecast encompassing

h	$\theta^* = 0.0$				$\theta^* = 0.5$				$\theta^* = 0.9$			
	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$
Panel A. $R = 1$ (null)												
2	0.070	0.015	0.001	0.000	0.049	0.006	0.001	0.000	0.037	0.003	0.000	0.000
3	0.182	0.067	0.012	0.001	0.154	0.048	0.005	0.000	0.130	0.030	0.002	0.000
4	0.275	0.126	0.029	0.003	0.248	0.105	0.022	0.002	0.226	0.083	0.014	0.001
5	0.360	0.169	0.054	0.009	0.346	0.157	0.047	0.005	0.330	0.132	0.033	0.002
6	0.424	0.220	0.089	0.017	0.419	0.206	0.077	0.013	0.406	0.180	0.061	0.008
Panel B. $R > 1$ (alternative)												
2	0.063	0.012	0.001	0.000	0.043	0.005	0.000	0.000	0.033	0.003	0.000	0.000
3	0.169	0.054	0.009	0.000	0.150	0.045	0.005	0.000	0.129	0.029	0.002	0.000
4	0.263	0.116	0.027	0.003	0.246	0.105	0.020	0.001	0.221	0.081	0.012	0.001
5	0.349	0.166	0.051	0.008	0.345	0.148	0.045	0.004	0.331	0.132	0.032	0.002
6	0.433	0.215	0.085	0.018	0.429	0.204	0.078	0.014	0.414	0.184	0.060	0.009

Table 4.3: Empirical size of nominal 0.10-level tests for equal forecast accuracy, $\rho = 0$

	$\theta^* = 0.0$				$\theta^* = 0.5$				$\theta^* = 0.9$			
	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$
$h = 1$												
DM_{rej}	0.165	0.127	0.119	0.110								
MDM_{rej}	0.085	0.091	0.102	0.103								
MDM_{non}	0.085	0.091	0.102	0.103								
MDM_{adapt}	0.085	0.091	0.102	0.103								
MDM_{SR}	0.085	0.091	0.102	0.103								
DM_b	0.165	0.127	0.119	0.110								
MDM_b	0.085	0.091	0.102	0.103								
DM_{HG}	0.165	0.127	0.119	0.110								
DM_{CI}	0.085	0.087	0.099	0.098								
$h = 2$												
DM_{rej}	0.323	0.206	0.153	0.124	0.305	0.189	0.149	0.121	0.296	0.186	0.148	0.122
MDM_{rej}	0.203	0.145	0.124	0.109	0.175	0.127	0.120	0.106	0.165	0.123	0.116	0.107
MDM_{non}	0.129	0.130	0.123	0.109	0.125	0.121	0.120	0.106	0.129	0.120	0.116	0.107
MDM_{adapt}	0.137	0.131	0.123	0.109	0.132	0.121	0.120	0.106	0.135	0.120	0.116	0.107
MDM_{SR}	0.137	0.131	0.123	0.109	0.132	0.121	0.120	0.106	0.135	0.120	0.116	0.107
DM_b	0.207	0.150	0.127	0.113	0.240	0.173	0.155	0.137	0.254	0.188	0.168	0.148
MDM_b	0.153	0.133	0.123	0.109	0.145	0.122	0.120	0.106	0.144	0.121	0.116	0.107
DM_{HG}	0.145	0.111	0.104	0.094	0.179	0.137	0.129	0.116	0.197	0.153	0.138	0.119
DM_{CI}	0.081	0.086	0.095	0.102	0.089	0.085	0.099	0.097	0.092	0.086	0.097	0.099
$h = 3$												
DM_{rej}	0.446	0.287	0.194	0.143	0.442	0.280	0.184	0.138	0.439	0.269	0.176	0.132
MDM_{rej}	0.294	0.207	0.153	0.122	0.275	0.198	0.142	0.119	0.256	0.183	0.134	0.113
MDM_{non}	0.119	0.143	0.140	0.122	0.121	0.150	0.135	0.119	0.125	0.150	0.131	0.112
MDM_{adapt}	0.160	0.159	0.143	0.122	0.159	0.163	0.137	0.119	0.163	0.161	0.132	0.112
MDM_{SR}	0.138	0.150	0.142	0.122	0.139	0.156	0.136	0.119	0.145	0.156	0.132	0.112
DM_b	0.257	0.176	0.141	0.119	0.279	0.201	0.156	0.137	0.311	0.222	0.174	0.153
MDM_b	0.183	0.163	0.144	0.122	0.180	0.166	0.137	0.119	0.180	0.163	0.132	0.112
DM_{HG}	0.146	0.114	0.101	0.095	0.174	0.139	0.119	0.117	0.211	0.169	0.143	0.126
DM_{CI}	0.082	0.092	0.094	0.098	0.095	0.097	0.098	0.098	0.109	0.099	0.100	0.097
$h = 4$												
DM_{rej}	0.550	0.356	0.226	0.156	0.543	0.349	0.223	0.150	0.548	0.340	0.210	0.145
MDM_{rej}	0.366	0.263	0.179	0.131	0.351	0.247	0.166	0.125	0.340	0.229	0.158	0.117
MDM_{non}	0.090	0.143	0.148	0.128	0.098	0.146	0.141	0.124	0.108	0.144	0.142	0.116
MDM_{adapt}	0.149	0.180	0.158	0.129	0.158	0.178	0.150	0.124	0.168	0.173	0.147	0.117
MDM_{SR}	0.112	0.157	0.151	0.128	0.121	0.157	0.144	0.124	0.136	0.156	0.144	0.116
DM_b	0.299	0.203	0.151	0.119	0.308	0.219	0.163	0.132	0.340	0.233	0.181	0.150
MDM_b	0.183	0.183	0.157	0.129	0.186	0.179	0.149	0.124	0.196	0.173	0.147	0.117
DM_{HG}	0.143	0.115	0.100	0.092	0.166	0.133	0.116	0.110	0.205	0.160	0.144	0.129
DM_{CI}	0.074	0.091	0.095	0.092	0.089	0.094	0.095	0.093	0.108	0.100	0.097	0.093
$h = 5$												
DM_{rej}	0.628	0.421	0.273	0.180	0.632	0.417	0.263	0.173	0.643	0.409	0.254	0.170
MDM_{rej}	0.415	0.304	0.215	0.150	0.409	0.296	0.201	0.142	0.404	0.282	0.187	0.136
MDM_{non}	0.063	0.130	0.160	0.142	0.069	0.134	0.155	0.137	0.079	0.141	0.153	0.132
MDM_{adapt}	0.141	0.182	0.178	0.145	0.149	0.187	0.173	0.138	0.161	0.192	0.165	0.133
MDM_{SR}	0.091	0.147	0.165	0.143	0.102	0.154	0.159	0.137	0.118	0.162	0.157	0.132
DM_b	0.338	0.219	0.161	0.128	0.353	0.236	0.171	0.138	0.384	0.257	0.192	0.160
MDM_b	0.186	0.182	0.173	0.144	0.199	0.187	0.168	0.138	0.207	0.192	0.161	0.133
DM_{HG}	0.145	0.117	0.102	0.090	0.160	0.130	0.115	0.109	0.197	0.160	0.142	0.137
DM_{CI}	0.079	0.093	0.097	0.097	0.089	0.099	0.100	0.100	0.107	0.106	0.103	0.099
$h = 6$												
DM_{rej}	0.732	0.473	0.310	0.204	0.733	0.470	0.301	0.195	0.742	0.470	0.294	0.184
MDM_{rej}	0.469	0.340	0.239	0.165	0.466	0.332	0.231	0.159	0.463	0.320	0.216	0.148
MDM_{non}	0.047	0.123	0.154	0.148	0.050	0.131	0.154	0.145	0.056	0.134	0.156	0.140
MDM_{adapt}	0.122	0.193	0.186	0.155	0.126	0.197	0.185	0.151	0.138	0.199	0.181	0.143
MDM_{SR}	0.082	0.145	0.161	0.150	0.087	0.154	0.163	0.147	0.101	0.159	0.165	0.141
DM_b	0.371	0.253	0.174	0.136	0.387	0.259	0.184	0.141	0.417	0.278	0.199	0.158
MDM_b	0.193	0.196	0.176	0.153	0.206	0.198	0.178	0.149	0.221	0.202	0.174	0.142
DM_{HG}	0.145	0.117	0.101	0.092	0.157	0.130	0.118	0.108	0.185	0.158	0.141	0.134
DM_{CI}	0.077	0.092	0.100	0.097	0.084	0.097	0.107	0.100	0.102	0.106	0.111	0.100

Table 4.4: Empirical size of nominal 0.10-level tests for equal forecast accuracy, $\rho = 0.5$

	$\theta^* = 0.0$				$\theta^* = 0.5$				$\theta^* = 0.9$			
	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$
$h = 1$												
DM_{rej}	0.165	0.128	0.118	0.106								
MDM_{rej}	0.084	0.090	0.100	0.099								
MDM_{non}	0.084	0.090	0.100	0.099								
MDM_{adapt}	0.084	0.090	0.100	0.099								
MDM_{SR}	0.084	0.090	0.100	0.099								
DM_b	0.165	0.128	0.118	0.106								
MDM_b	0.084	0.090	0.100	0.099								
DM_{HG}	0.165	0.128	0.118	0.106								
DM_{CI}	0.086	0.088	0.095	0.101								
$h = 2$												
DM_{rej}	0.317	0.212	0.152	0.127	0.302	0.193	0.150	0.125	0.294	0.184	0.144	0.124
MDM_{rej}	0.198	0.153	0.121	0.113	0.172	0.131	0.117	0.112	0.163	0.122	0.116	0.109
MDM_{non}	0.126	0.139	0.120	0.113	0.124	0.123	0.116	0.112	0.127	0.119	0.116	0.109
MDM_{adapt}	0.134	0.140	0.120	0.113	0.132	0.123	0.116	0.112	0.133	0.119	0.116	0.109
MDM_{SR}	0.134	0.140	0.120	0.113	0.132	0.123	0.116	0.112	0.133	0.119	0.116	0.109
DM_b	0.206	0.158	0.126	0.114	0.230	0.173	0.155	0.140	0.252	0.185	0.168	0.150
MDM_b	0.150	0.143	0.120	0.113	0.141	0.124	0.116	0.112	0.142	0.119	0.116	0.109
DM_{HG}	0.146	0.118	0.102	0.097	0.179	0.141	0.128	0.119	0.197	0.151	0.135	0.123
DM_{CI}	0.081	0.092	0.095	0.101	0.089	0.084	0.098	0.101	0.092	0.084	0.097	0.101
$h = 3$												
DM_{rej}	0.452	0.287	0.187	0.143	0.443	0.281	0.179	0.133	0.436	0.269	0.173	0.133
MDM_{rej}	0.296	0.208	0.149	0.125	0.275	0.195	0.141	0.114	0.250	0.175	0.129	0.112
MDM_{non}	0.118	0.148	0.140	0.125	0.121	0.147	0.134	0.114	0.122	0.144	0.126	0.112
MDM_{adapt}	0.160	0.162	0.141	0.125	0.158	0.161	0.136	0.114	0.157	0.154	0.127	0.112
MDM_{SR}	0.135	0.153	0.140	0.125	0.139	0.154	0.135	0.114	0.142	0.148	0.126	0.112
DM_b	0.261	0.172	0.135	0.119	0.281	0.196	0.156	0.132	0.311	0.218	0.175	0.151
MDM_b	0.181	0.166	0.142	0.125	0.176	0.163	0.136	0.114	0.174	0.155	0.127	0.112
DM_{HG}	0.149	0.110	0.102	0.099	0.175	0.135	0.122	0.111	0.215	0.164	0.140	0.127
DM_{CI}	0.080	0.088	0.094	0.098	0.092	0.090	0.096	0.094	0.106	0.093	0.095	0.095
$h = 4$												
DM_{rej}	0.552	0.358	0.230	0.158	0.544	0.357	0.219	0.152	0.552	0.348	0.207	0.146
MDM_{rej}	0.363	0.264	0.182	0.134	0.349	0.254	0.163	0.125	0.337	0.237	0.151	0.118
MDM_{non}	0.091	0.141	0.150	0.130	0.094	0.150	0.139	0.123	0.109	0.158	0.138	0.117
MDM_{adapt}	0.150	0.178	0.160	0.131	0.155	0.184	0.146	0.124	0.167	0.184	0.142	0.118
MDM_{SR}	0.112	0.154	0.154	0.130	0.122	0.163	0.141	0.124	0.139	0.170	0.140	0.118
DM_b	0.302	0.200	0.144	0.121	0.321	0.222	0.159	0.136	0.351	0.246	0.179	0.150
MDM_b	0.181	0.179	0.159	0.130	0.189	0.183	0.145	0.124	0.202	0.186	0.141	0.118
DM_{HG}	0.144	0.118	0.095	0.094	0.167	0.137	0.112	0.115	0.206	0.170	0.138	0.129
DM_{CI}	0.080	0.091	0.088	0.094	0.095	0.098	0.093	0.093	0.114	0.102	0.094	0.094
$h = 5$												
DM_{rej}	0.644	0.415	0.273	0.179	0.645	0.410	0.262	0.175	0.655	0.402	0.250	0.168
MDM_{rej}	0.421	0.303	0.209	0.152	0.414	0.294	0.199	0.141	0.409	0.271	0.185	0.133
MDM_{non}	0.064	0.133	0.154	0.141	0.067	0.139	0.154	0.134	0.078	0.135	0.153	0.130
MDM_{adapt}	0.138	0.187	0.174	0.145	0.145	0.190	0.170	0.137	0.165	0.184	0.166	0.131
MDM_{SR}	0.094	0.151	0.159	0.142	0.100	0.157	0.158	0.135	0.119	0.155	0.157	0.131
DM_b	0.338	0.219	0.164	0.130	0.361	0.233	0.174	0.137	0.393	0.252	0.190	0.156
MDM_b	0.185	0.187	0.169	0.143	0.195	0.190	0.168	0.136	0.211	0.184	0.163	0.131
DM_{HG}	0.143	0.112	0.100	0.094	0.162	0.128	0.117	0.106	0.200	0.157	0.144	0.134
DM_{CI}	0.079	0.086	0.096	0.099	0.090	0.092	0.099	0.094	0.109	0.102	0.105	0.100
$h = 6$												
DM_{rej}	0.731	0.480	0.307	0.201	0.740	0.477	0.298	0.198	0.751	0.468	0.295	0.194
MDM_{rej}	0.472	0.346	0.236	0.166	0.471	0.337	0.228	0.160	0.462	0.323	0.214	0.152
MDM_{non}	0.050	0.128	0.152	0.149	0.053	0.131	0.151	0.145	0.058	0.140	0.152	0.142
MDM_{adapt}	0.126	0.192	0.183	0.155	0.128	0.199	0.181	0.150	0.141	0.206	0.177	0.146
MDM_{SR}	0.086	0.147	0.159	0.150	0.091	0.152	0.160	0.146	0.102	0.164	0.160	0.143
DM_b	0.371	0.251	0.171	0.133	0.393	0.266	0.183	0.142	0.421	0.282	0.197	0.162
MDM_b	0.203	0.194	0.173	0.152	0.209	0.200	0.173	0.147	0.221	0.207	0.171	0.145
DM_{HG}	0.145	0.121	0.097	0.094	0.158	0.130	0.113	0.107	0.183	0.158	0.138	0.138
DM_{CI}	0.076	0.098	0.095	0.098	0.084	0.099	0.102	0.098	0.099	0.109	0.104	0.106

Table 4.5: Empirical size of nominal 0.10-level tests for equal forecast accuracy, $\rho = 0.9$

	$\theta^* = 0.0$				$\theta^* = 0.5$				$\theta^* = 0.9$			
	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$
$h = 1$												
DM_{rej}	0.165	0.126	0.118	0.108								
MDM_{rej}	0.082	0.090	0.099	0.099								
MDM_{non}	0.082	0.090	0.099	0.099								
MDM_{adapt}	0.082	0.090	0.099	0.099								
MDM_{SR}	0.082	0.090	0.099	0.099								
DM_b	0.165	0.126	0.118	0.108								
MDM_b	0.082	0.090	0.099	0.099								
DM_{HG}	0.165	0.126	0.118	0.108								
DM_{CI}	0.081	0.088	0.091	0.099								
$h = 2$												
DM_{rej}	0.325	0.210	0.155	0.128	0.315	0.192	0.151	0.125	0.308	0.185	0.148	0.122
MDM_{rej}	0.203	0.150	0.124	0.116	0.180	0.130	0.120	0.111	0.168	0.122	0.118	0.108
MDM_{non}	0.129	0.136	0.122	0.116	0.127	0.124	0.119	0.111	0.129	0.119	0.118	0.108
MDM_{adapt}	0.137	0.137	0.123	0.116	0.135	0.124	0.119	0.111	0.135	0.119	0.118	0.108
MDM_{SR}	0.137	0.137	0.123	0.116	0.135	0.124	0.119	0.111	0.135	0.119	0.118	0.108
DM_b	0.210	0.152	0.130	0.118	0.243	0.177	0.158	0.142	0.259	0.189	0.171	0.151
MDM_b	0.154	0.140	0.123	0.116	0.147	0.125	0.119	0.111	0.143	0.119	0.118	0.108
DM_{HG}	0.147	0.115	0.104	0.099	0.185	0.141	0.132	0.119	0.203	0.155	0.139	0.121
DM_{CI}	0.082	0.092	0.100	0.100	0.089	0.085	0.099	0.100	0.094	0.084	0.098	0.100
$h = 3$												
DM_{rej}	0.444	0.288	0.189	0.142	0.443	0.276	0.182	0.140	0.434	0.266	0.174	0.133
MDM_{rej}	0.293	0.207	0.149	0.124	0.275	0.195	0.139	0.123	0.248	0.173	0.132	0.115
MDM_{non}	0.117	0.140	0.138	0.123	0.123	0.145	0.133	0.122	0.125	0.142	0.130	0.115
MDM_{adapt}	0.153	0.156	0.140	0.123	0.158	0.159	0.134	0.122	0.160	0.149	0.130	0.115
MDM_{SR}	0.133	0.149	0.139	0.123	0.138	0.152	0.133	0.122	0.144	0.146	0.130	0.115
DM_b	0.257	0.170	0.138	0.122	0.280	0.197	0.156	0.139	0.311	0.217	0.179	0.155
MDM_b	0.176	0.160	0.140	0.123	0.178	0.161	0.134	0.122	0.176	0.152	0.130	0.115
DM_{HG}	0.152	0.114	0.102	0.098	0.173	0.137	0.126	0.119	0.214	0.164	0.144	0.128
DM_{CI}	0.077	0.087	0.093	0.101	0.094	0.089	0.095	0.100	0.110	0.090	0.096	0.102
$h = 4$												
DM_{rej}	0.554	0.363	0.223	0.156	0.553	0.365	0.216	0.154	0.556	0.349	0.209	0.151
MDM_{rej}	0.368	0.270	0.172	0.132	0.352	0.260	0.164	0.127	0.340	0.237	0.153	0.124
MDM_{non}	0.095	0.145	0.142	0.130	0.102	0.152	0.143	0.125	0.108	0.154	0.141	0.123
MDM_{adapt}	0.153	0.184	0.152	0.131	0.164	0.186	0.150	0.126	0.169	0.182	0.146	0.123
MDM_{SR}	0.117	0.158	0.145	0.130	0.127	0.164	0.145	0.125	0.138	0.167	0.142	0.123
DM_b	0.307	0.202	0.149	0.122	0.321	0.223	0.157	0.136	0.346	0.249	0.181	0.159
MDM_b	0.188	0.186	0.150	0.130	0.194	0.187	0.149	0.125	0.200	0.184	0.145	0.123
DM_{HG}	0.151	0.118	0.099	0.096	0.168	0.138	0.111	0.117	0.206	0.173	0.140	0.136
DM_{CI}	0.081	0.093	0.094	0.094	0.089	0.096	0.092	0.097	0.113	0.100	0.097	0.101
$h = 5$												
DM_{rej}	0.648	0.415	0.272	0.184	0.644	0.414	0.263	0.172	0.651	0.401	0.253	0.169
MDM_{rej}	0.428	0.303	0.209	0.150	0.415	0.294	0.198	0.141	0.405	0.275	0.189	0.138
MDM_{non}	0.068	0.134	0.154	0.140	0.068	0.141	0.150	0.135	0.075	0.145	0.154	0.135
MDM_{adapt}	0.145	0.187	0.174	0.144	0.145	0.188	0.166	0.137	0.154	0.192	0.169	0.136
MDM_{SR}	0.098	0.150	0.159	0.141	0.101	0.158	0.155	0.136	0.115	0.164	0.159	0.136
DM_b	0.344	0.222	0.164	0.129	0.354	0.235	0.176	0.137	0.387	0.257	0.195	0.160
MDM_b	0.191	0.187	0.170	0.143	0.194	0.190	0.165	0.136	0.208	0.192	0.166	0.136
DM_{HG}	0.145	0.114	0.104	0.096	0.161	0.131	0.116	0.111	0.192	0.162	0.142	0.139
DM_{CI}	0.081	0.087	0.096	0.094	0.088	0.093	0.100	0.093	0.104	0.102	0.105	0.102
$h = 6$												
DM_{rej}	0.735	0.475	0.306	0.198	0.738	0.476	0.301	0.196	0.745	0.471	0.292	0.184
MDM_{rej}	0.471	0.339	0.236	0.162	0.472	0.332	0.228	0.156	0.464	0.318	0.217	0.144
MDM_{non}	0.053	0.120	0.148	0.142	0.053	0.120	0.153	0.141	0.056	0.135	0.158	0.135
MDM_{adapt}	0.129	0.185	0.181	0.150	0.129	0.188	0.179	0.146	0.137	0.198	0.180	0.139
MDM_{SR}	0.088	0.142	0.156	0.144	0.091	0.145	0.160	0.142	0.104	0.161	0.166	0.136
DM_b	0.381	0.245	0.172	0.130	0.398	0.264	0.181	0.136	0.421	0.285	0.201	0.151
MDM_b	0.203	0.188	0.172	0.146	0.209	0.191	0.172	0.143	0.221	0.204	0.174	0.137
DM_{HG}	0.147	0.120	0.097	0.092	0.159	0.136	0.110	0.104	0.188	0.164	0.141	0.129
DM_{CI}	0.078	0.093	0.091	0.094	0.088	0.102	0.098	0.096	0.107	0.114	0.107	0.098

Table 4.6: Empirical size of nominal 0.10-level tests for equal forecast accuracy with heavy-tailed errors, $\rho = 0$

	$\theta^* = 0.0$				$\theta^* = 0.5$				$\theta^* = 0.9$			
	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$
$h = 1$												
DM_{rej}	0.143	0.123	0.100	0.107								
MDM_{rej}	0.068	0.086	0.083	0.099								
MDM_{non}	0.068	0.086	0.083	0.099								
MDM_{adapt}	0.068	0.086	0.083	0.099								
MDM_{SR}	0.068	0.086	0.083	0.099								
DM_b	0.143	0.123	0.100	0.107								
MDM_b	0.068	0.086	0.083	0.099								
DM_{HG}	0.144	0.124	0.101	0.109								
DM_{CI}	0.064	0.072	0.079	0.089								
$h = 2$												
DM_{rej}	0.303	0.201	0.138	0.116	0.289	0.189	0.133	0.112	0.282	0.176	0.130	0.111
MDM_{rej}	0.176	0.135	0.107	0.102	0.154	0.120	0.102	0.096	0.145	0.113	0.101	0.096
MDM_{non}	0.116	0.123	0.106	0.102	0.116	0.113	0.102	0.096	0.119	0.112	0.101	0.096
MDM_{adapt}	0.121	0.124	0.106	0.102	0.120	0.114	0.102	0.096	0.123	0.112	0.101	0.096
MDM_{SR}	0.121	0.124	0.106	0.102	0.120	0.114	0.102	0.096	0.123	0.112	0.101	0.096
DM_b	0.193	0.146	0.115	0.105	0.232	0.177	0.144	0.133	0.258	0.194	0.161	0.152
MDM_b	0.134	0.125	0.106	0.102	0.129	0.116	0.102	0.096	0.130	0.112	0.101	0.096
DM_{HG}	0.134	0.106	0.091	0.091	0.178	0.142	0.121	0.109	0.203	0.154	0.126	0.111
DM_{CI}	0.061	0.077	0.082	0.088	0.075	0.074	0.084	0.087	0.087	0.079	0.084	0.088
$h = 3$												
DM_{rej}	0.442	0.271	0.178	0.140	0.443	0.267	0.173	0.132	0.435	0.256	0.168	0.130
MDM_{rej}	0.271	0.184	0.137	0.117	0.257	0.172	0.129	0.111	0.240	0.164	0.123	0.111
MDM_{non}	0.110	0.136	0.127	0.116	0.113	0.137	0.124	0.111	0.123	0.139	0.121	0.110
MDM_{adapt}	0.149	0.146	0.128	0.116	0.151	0.145	0.125	0.111	0.158	0.145	0.122	0.110
MDM_{SR}	0.124	0.139	0.127	0.116	0.130	0.141	0.124	0.111	0.142	0.142	0.121	0.110
DM_b	0.243	0.163	0.129	0.115	0.276	0.190	0.153	0.135	0.311	0.220	0.180	0.159
MDM_b	0.166	0.148	0.129	0.116	0.169	0.147	0.125	0.111	0.174	0.147	0.122	0.110
DM_{HG}	0.132	0.102	0.092	0.091	0.168	0.131	0.122	0.112	0.217	0.167	0.146	0.130
DM_{CI}	0.065	0.070	0.081	0.091	0.078	0.079	0.085	0.090	0.107	0.086	0.090	0.090
$h = 4$												
DM_{rej}	0.554	0.337	0.215	0.157	0.558	0.334	0.207	0.152	0.558	0.326	0.201	0.150
MDM_{rej}	0.348	0.233	0.159	0.130	0.339	0.226	0.152	0.124	0.323	0.214	0.142	0.121
MDM_{non}	0.093	0.137	0.137	0.127	0.098	0.141	0.137	0.122	0.110	0.147	0.133	0.121
MDM_{adapt}	0.155	0.164	0.144	0.127	0.155	0.168	0.141	0.123	0.168	0.170	0.136	0.121
MDM_{SR}	0.110	0.146	0.139	0.127	0.118	0.152	0.138	0.122	0.136	0.158	0.135	0.121
DM_b	0.298	0.185	0.138	0.124	0.314	0.206	0.151	0.134	0.353	0.233	0.174	0.157
MDM_b	0.178	0.165	0.142	0.127	0.182	0.169	0.140	0.123	0.193	0.172	0.135	0.121
DM_{HG}	0.137	0.105	0.092	0.095	0.161	0.127	0.111	0.114	0.208	0.163	0.138	0.137
DM_{CI}	0.066	0.073	0.079	0.090	0.080	0.081	0.086	0.093	0.107	0.089	0.092	0.094
$h = 5$												
DM_{rej}	0.647	0.410	0.251	0.175	0.652	0.404	0.245	0.168	0.658	0.399	0.240	0.160
MDM_{rej}	0.413	0.280	0.186	0.143	0.412	0.270	0.179	0.136	0.401	0.257	0.169	0.128
MDM_{non}	0.064	0.132	0.141	0.134	0.069	0.136	0.142	0.130	0.077	0.143	0.140	0.126
MDM_{adapt}	0.147	0.183	0.156	0.138	0.151	0.186	0.156	0.132	0.165	0.188	0.153	0.127
MDM_{SR}	0.090	0.147	0.145	0.135	0.097	0.153	0.145	0.130	0.113	0.161	0.145	0.126
DM_b	0.346	0.219	0.148	0.129	0.366	0.233	0.159	0.136	0.397	0.260	0.184	0.155
MDM_b	0.194	0.182	0.152	0.137	0.200	0.186	0.151	0.131	0.214	0.189	0.149	0.126
DM_{HG}	0.141	0.106	0.091	0.094	0.155	0.125	0.108	0.111	0.197	0.160	0.141	0.139
DM_{CI}	0.065	0.079	0.076	0.090	0.076	0.085	0.083	0.092	0.096	0.096	0.092	0.092
$h = 6$												
DM_{rej}	0.747	0.465	0.293	0.202	0.751	0.465	0.290	0.195	0.759	0.472	0.280	0.191
MDM_{rej}	0.482	0.312	0.217	0.163	0.477	0.306	0.213	0.157	0.474	0.303	0.200	0.145
MDM_{non}	0.046	0.122	0.145	0.148	0.046	0.124	0.146	0.145	0.055	0.131	0.149	0.138
MDM_{adapt}	0.137	0.183	0.172	0.154	0.133	0.186	0.170	0.150	0.147	0.193	0.168	0.141
MDM_{SR}	0.076	0.138	0.153	0.150	0.080	0.141	0.153	0.146	0.095	0.151	0.155	0.139
DM_b	0.387	0.232	0.157	0.133	0.401	0.247	0.168	0.142	0.430	0.271	0.190	0.159
MDM_b	0.212	0.183	0.164	0.152	0.217	0.184	0.163	0.147	0.234	0.194	0.163	0.140
DM_{HG}	0.134	0.102	0.092	0.089	0.142	0.114	0.103	0.103	0.177	0.144	0.137	0.136
DM_{CI}	0.062	0.070	0.077	0.092	0.070	0.076	0.081	0.094	0.089	0.095	0.092	0.098

Table 4.7: Empirical size of nominal 0.10-level tests for equal forecast accuracy with negative MA coefficients, $\rho = 0$

	$\theta^* = 0.0$				$\theta^* = -0.5$				$\theta^* = -0.9$			
	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$
$h = 1$												
DM_{rej}	0.165	0.127	0.119	0.110								
MDM_{rej}	0.085	0.091	0.102	0.103								
MDM_{non}	0.085	0.091	0.102	0.103								
MDM_{adapt}	0.085	0.091	0.102	0.103								
MDM_{SR}	0.085	0.091	0.102	0.103								
DM_b	0.165	0.127	0.119	0.110								
MDM_b	0.085	0.091	0.102	0.103								
DM_{HG}	0.165	0.127	0.119	0.110								
DM_{CI}	0.085	0.087	0.099	0.098								
$h = 2$												
DM_{rej}	0.323	0.206	0.153	0.124	0.307	0.199	0.146	0.124	0.299	0.194	0.146	0.123
MDM_{rej}	0.203	0.145	0.124	0.109	0.178	0.133	0.117	0.108	0.161	0.130	0.113	0.108
MDM_{non}	0.129	0.130	0.123	0.109	0.131	0.127	0.117	0.108	0.127	0.126	0.113	0.108
MDM_{adapt}	0.137	0.131	0.123	0.109	0.138	0.128	0.117	0.108	0.134	0.126	0.113	0.108
MDM_{SR}	0.137	0.131	0.123	0.109	0.138	0.128	0.117	0.108	0.134	0.126	0.113	0.108
DM_b	0.207	0.150	0.127	0.113	0.236	0.182	0.153	0.141	0.257	0.198	0.167	0.152
MDM_b	0.153	0.133	0.123	0.109	0.148	0.129	0.117	0.108	0.142	0.127	0.113	0.108
DM_{HG}	0.145	0.111	0.104	0.094	0.180	0.146	0.126	0.118	0.200	0.162	0.137	0.121
DM_{CI}	0.081	0.086	0.095	0.102	0.088	0.089	0.099	0.100	0.095	0.091	0.100	0.102
$h = 3$												
DM_{rej}	0.446	0.287	0.194	0.143	0.445	0.278	0.186	0.140	0.447	0.271	0.180	0.138
MDM_{rej}	0.294	0.207	0.153	0.122	0.281	0.195	0.145	0.120	0.272	0.186	0.139	0.118
MDM_{non}	0.119	0.143	0.140	0.122	0.119	0.146	0.138	0.120	0.123	0.145	0.134	0.118
MDM_{adapt}	0.160	0.159	0.143	0.122	0.163	0.159	0.139	0.120	0.166	0.156	0.135	0.118
MDM_{SR}	0.138	0.150	0.142	0.122	0.138	0.151	0.138	0.120	0.143	0.150	0.135	0.118
DM_b	0.257	0.176	0.141	0.119	0.280	0.192	0.151	0.133	0.293	0.206	0.167	0.146
MDM_b	0.183	0.163	0.144	0.122	0.181	0.162	0.139	0.120	0.182	0.159	0.135	0.118
DM_{HG}	0.146	0.114	0.101	0.095	0.169	0.130	0.122	0.113	0.183	0.149	0.134	0.129
DM_{CI}	0.082	0.092	0.094	0.098	0.092	0.094	0.099	0.099	0.105	0.097	0.101	0.102
$h = 4$												
DM_{rej}	0.550	0.356	0.226	0.156	0.556	0.355	0.228	0.154	0.557	0.350	0.223	0.154
MDM_{rej}	0.366	0.263	0.179	0.131	0.365	0.256	0.179	0.130	0.353	0.244	0.173	0.130
MDM_{non}	0.090	0.143	0.148	0.128	0.096	0.145	0.153	0.128	0.097	0.149	0.152	0.129
MDM_{adapt}	0.149	0.180	0.158	0.129	0.156	0.180	0.162	0.129	0.157	0.179	0.159	0.129
MDM_{SR}	0.112	0.157	0.151	0.128	0.118	0.157	0.156	0.128	0.122	0.163	0.154	0.129
DM_b	0.299	0.203	0.151	0.119	0.316	0.210	0.161	0.131	0.326	0.226	0.174	0.144
MDM_b	0.183	0.183	0.157	0.129	0.192	0.182	0.160	0.128	0.189	0.183	0.158	0.129
DM_{HG}	0.143	0.115	0.100	0.092	0.154	0.124	0.113	0.106	0.163	0.141	0.124	0.120
DM_{CI}	0.074	0.091	0.095	0.092	0.082	0.096	0.098	0.098	0.090	0.102	0.104	0.102
$h = 5$												
DM_{rej}	0.628	0.421	0.273	0.180	0.635	0.412	0.272	0.178	0.640	0.407	0.266	0.177
MDM_{rej}	0.415	0.304	0.215	0.150	0.411	0.298	0.208	0.147	0.413	0.287	0.196	0.143
MDM_{non}	0.063	0.130	0.160	0.142	0.065	0.132	0.158	0.140	0.073	0.136	0.153	0.139
MDM_{adapt}	0.141	0.182	0.178	0.145	0.142	0.186	0.177	0.143	0.151	0.187	0.170	0.142
MDM_{SR}	0.091	0.147	0.165	0.143	0.097	0.152	0.162	0.141	0.102	0.155	0.158	0.140
DM_b	0.338	0.219	0.161	0.128	0.351	0.232	0.175	0.140	0.360	0.239	0.183	0.148
MDM_b	0.186	0.182	0.173	0.144	0.194	0.186	0.173	0.142	0.199	0.190	0.166	0.141
DM_{HG}	0.145	0.117	0.102	0.090	0.151	0.128	0.112	0.105	0.161	0.136	0.121	0.116
DM_{CI}	0.079	0.093	0.097	0.097	0.085	0.095	0.102	0.101	0.087	0.102	0.109	0.105
$h = 6$												
DM_{rej}	0.732	0.473	0.310	0.204	0.733	0.471	0.303	0.200	0.736	0.471	0.301	0.198
MDM_{rej}	0.469	0.340	0.239	0.165	0.472	0.335	0.234	0.164	0.467	0.331	0.231	0.158
MDM_{non}	0.047	0.123	0.154	0.148	0.050	0.126	0.156	0.150	0.050	0.135	0.159	0.147
MDM_{adapt}	0.122	0.193	0.186	0.155	0.127	0.194	0.185	0.156	0.129	0.204	0.186	0.152
MDM_{SR}	0.082	0.145	0.161	0.150	0.088	0.149	0.164	0.151	0.089	0.158	0.167	0.148
DM_b	0.371	0.253	0.174	0.136	0.380	0.259	0.188	0.142	0.389	0.268	0.197	0.150
MDM_b	0.193	0.196	0.176	0.153	0.204	0.197	0.178	0.153	0.205	0.205	0.179	0.150
DM_{HG}	0.145	0.117	0.101	0.092	0.151	0.125	0.113	0.103	0.153	0.131	0.121	0.111
DM_{CI}	0.077	0.092	0.100	0.097	0.081	0.096	0.105	0.099	0.081	0.104	0.111	0.102

Table 4.8: Size-adjusted power of nominal 0.10-level tests for equal forecast accuracy, $\rho = 0$

	$\theta^* = 0.0$				$\theta^* = -0.5$				$\theta^* = -0.9$			
	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$
$h = 1$												
MDM_{non}	0.709	0.820	0.677	0.699								
MDM_{SR}	0.709	0.820	0.677	0.699								
MDM_b	0.709	0.820	0.677	0.699								
DM_{CI}	0.572	0.655	0.591	0.642								
$h = 2$												
MDM_{non}	0.454	0.688	0.638	0.679	0.436	0.627	0.565	0.595	0.392	0.594	0.521	0.559
MDM_{SR}	0.480	0.689	0.638	0.679	0.453	0.627	0.565	0.595	0.402	0.596	0.521	0.559
MDM_b	0.443	0.686	0.638	0.679	0.410	0.624	0.565	0.595	0.386	0.596	0.521	0.559
DM_{CI}	0.607	0.648	0.606	0.612	0.526	0.589	0.514	0.541	0.504	0.548	0.471	0.499
$h = 3$												
MDM_{non}	0.385	0.574	0.601	0.656	0.354	0.517	0.535	0.599	0.339	0.470	0.466	0.509
MDM_{SR}	0.434	0.596	0.605	0.656	0.399	0.531	0.534	0.599	0.374	0.481	0.466	0.509
MDM_b	0.393	0.565	0.601	0.656	0.377	0.509	0.530	0.598	0.348	0.468	0.465	0.509
DM_{CI}	0.564	0.635	0.619	0.643	0.526	0.572	0.547	0.576	0.474	0.517	0.460	0.487
$h = 4$												
MDM_{non}	0.388	0.505	0.531	0.638	0.369	0.476	0.512	0.598	0.339	0.424	0.454	0.523
MDM_{SR}	0.506	0.560	0.540	0.640	0.462	0.522	0.520	0.598	0.401	0.454	0.459	0.523
MDM_b	0.428	0.524	0.535	0.638	0.403	0.492	0.509	0.598	0.352	0.427	0.456	0.523
DM_{CI}	0.604	0.627	0.592	0.647	0.559	0.591	0.568	0.598	0.493	0.520	0.485	0.512
$h = 5$												
MDM_{non}	0.388	0.478	0.483	0.599	0.352	0.451	0.466	0.581	0.319	0.412	0.432	0.514
MDM_{SR}	0.513	0.556	0.507	0.603	0.478	0.512	0.483	0.584	0.425	0.458	0.443	0.516
MDM_b	0.418	0.516	0.497	0.601	0.401	0.474	0.476	0.583	0.364	0.423	0.444	0.516
DM_{CI}	0.593	0.627	0.598	0.626	0.561	0.576	0.571	0.588	0.503	0.514	0.510	0.527
$h = 6$												
MDM_{non}	0.318	0.441	0.460	0.577	0.304	0.420	0.441	0.557	0.297	0.386	0.397	0.505
MDM_{SR}	0.505	0.545	0.488	0.581	0.477	0.518	0.470	0.561	0.419	0.471	0.417	0.506
MDM_b	0.433	0.506	0.482	0.583	0.414	0.465	0.463	0.563	0.374	0.427	0.406	0.504
DM_{CI}	0.597	0.632	0.582	0.638	0.572	0.592	0.541	0.596	0.495	0.554	0.491	0.533

Table 4.9: Empirical size of nominal 0.10-level tests for forecast encompassing

	$\theta^* = 0.0$				$\theta^* = 0.5$				$\theta^* = 0.9$			
	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$
$h = 1$												
DM_{rej}	0.142	0.119	0.111	0.108								
MDM_{rej}	0.100	0.100	0.102	0.103								
MDM_{non}	0.100	0.100	0.102	0.103								
MDM_{adapt}	0.100	0.100	0.102	0.103								
MDM_{SR}	0.100	0.100	0.102	0.103								
DM_b	0.142	0.119	0.111	0.108								
MDM_b	0.100	0.100	0.102	0.103								
DM_{HG}	0.142	0.119	0.111	0.108								
DM_{CI}	0.091	0.095	0.102	0.103								
$h = 2$												
DM_{rej}	0.257	0.168	0.127	0.117	0.237	0.161	0.129	0.117	0.230	0.158	0.129	0.117
MDM_{rej}	0.189	0.138	0.111	0.110	0.167	0.127	0.112	0.109	0.156	0.122	0.112	0.108
MDM_{non}	0.118	0.123	0.110	0.110	0.118	0.121	0.112	0.109	0.119	0.118	0.112	0.108
MDM_{adapt}	0.125	0.125	0.110	0.110	0.124	0.122	0.112	0.109	0.124	0.119	0.112	0.108
MDM_{SR}	0.125	0.125	0.110	0.110	0.124	0.122	0.112	0.109	0.124	0.119	0.112	0.108
DM_b	0.174	0.136	0.116	0.112	0.186	0.157	0.135	0.129	0.196	0.162	0.143	0.134
MDM_b	0.134	0.126	0.110	0.110	0.129	0.123	0.112	0.109	0.129	0.119	0.112	0.108
DM_{HG}	0.139	0.115	0.100	0.100	0.156	0.135	0.118	0.113	0.168	0.141	0.124	0.116
DM_{CI}	0.095	0.097	0.098	0.102	0.106	0.100	0.104	0.106	0.110	0.101	0.105	0.107
$h = 3$												
DM_{rej}	0.359	0.229	0.161	0.125	0.346	0.219	0.154	0.127	0.333	0.211	0.150	0.127
MDM_{rej}	0.276	0.186	0.136	0.112	0.252	0.175	0.127	0.115	0.235	0.159	0.123	0.115
MDM_{non}	0.094	0.119	0.125	0.111	0.098	0.127	0.122	0.115	0.106	0.129	0.120	0.115
MDM_{adapt}	0.124	0.132	0.127	0.111	0.127	0.134	0.123	0.115	0.131	0.136	0.121	0.115
MDM_{SR}	0.113	0.127	0.126	0.111	0.117	0.131	0.122	0.115	0.124	0.133	0.121	0.115
DM_b	0.189	0.147	0.130	0.114	0.199	0.160	0.140	0.125	0.217	0.178	0.155	0.137
MDM_b	0.135	0.133	0.127	0.111	0.135	0.137	0.123	0.115	0.139	0.136	0.121	0.115
DM_{HG}	0.133	0.112	0.107	0.100	0.150	0.129	0.120	0.112	0.175	0.147	0.134	0.123
DM_{CI}	0.095	0.095	0.104	0.105	0.106	0.102	0.106	0.103	0.121	0.105	0.106	0.106
$h = 4$												
DM_{rej}	0.455	0.298	0.176	0.134	0.443	0.287	0.172	0.133	0.439	0.281	0.170	0.135
MDM_{rej}	0.351	0.243	0.149	0.118	0.332	0.229	0.141	0.117	0.321	0.215	0.137	0.119
MDM_{non}	0.076	0.116	0.121	0.115	0.084	0.124	0.120	0.115	0.095	0.132	0.123	0.118
MDM_{adapt}	0.125	0.144	0.126	0.116	0.128	0.149	0.124	0.116	0.136	0.150	0.127	0.118
MDM_{SR}	0.105	0.132	0.123	0.116	0.112	0.139	0.122	0.116	0.123	0.144	0.125	0.118
DM_b	0.207	0.169	0.128	0.116	0.221	0.178	0.137	0.126	0.236	0.194	0.152	0.143
MDM_b	0.141	0.146	0.126	0.116	0.145	0.150	0.124	0.116	0.150	0.152	0.126	0.118
DM_{HG}	0.132	0.122	0.101	0.098	0.146	0.133	0.112	0.114	0.167	0.153	0.129	0.129
DM_{CI}	0.093	0.106	0.100	0.098	0.104	0.109	0.102	0.105	0.123	0.116	0.107	0.108
$h = 5$												
DM_{rej}	0.530	0.337	0.206	0.142	0.526	0.332	0.203	0.136	0.522	0.322	0.194	0.136
MDM_{rej}	0.417	0.270	0.172	0.124	0.406	0.264	0.167	0.116	0.391	0.247	0.157	0.117
MDM_{non}	0.057	0.102	0.117	0.115	0.059	0.108	0.121	0.111	0.060	0.115	0.124	0.116
MDM_{adapt}	0.115	0.140	0.130	0.117	0.119	0.145	0.132	0.112	0.120	0.147	0.132	0.116
MDM_{SR}	0.092	0.122	0.123	0.115	0.096	0.128	0.127	0.111	0.103	0.134	0.129	0.116
DM_b	0.222	0.173	0.136	0.114	0.230	0.181	0.141	0.120	0.243	0.196	0.154	0.133
MDM_b	0.142	0.141	0.130	0.116	0.145	0.147	0.132	0.112	0.146	0.148	0.133	0.116
DM_{HG}	0.132	0.112	0.101	0.095	0.139	0.126	0.109	0.105	0.156	0.148	0.127	0.122
DM_{CI}	0.092	0.098	0.100	0.098	0.100	0.105	0.102	0.098	0.113	0.117	0.108	0.103
$h = 6$												
DM_{rej}	0.608	0.393	0.246	0.153	0.606	0.385	0.241	0.156	0.604	0.373	0.232	0.148
MDM_{rej}	0.468	0.315	0.205	0.135	0.466	0.308	0.200	0.134	0.457	0.287	0.187	0.126
MDM_{non}	0.044	0.095	0.116	0.117	0.047	0.101	0.123	0.120	0.051	0.107	0.126	0.117
MDM_{adapt}	0.105	0.143	0.138	0.121	0.113	0.149	0.143	0.123	0.118	0.149	0.142	0.119
MDM_{SR}	0.086	0.118	0.125	0.119	0.094	0.127	0.131	0.121	0.103	0.135	0.135	0.118
DM_b	0.248	0.183	0.141	0.115	0.257	0.195	0.151	0.125	0.272	0.208	0.164	0.134
MDM_b	0.148	0.143	0.134	0.120	0.154	0.151	0.140	0.123	0.164	0.153	0.140	0.119
DM_{HG}	0.133	0.118	0.098	0.091	0.143	0.128	0.112	0.105	0.163	0.150	0.134	0.120
DM_{CI}	0.095	0.099	0.099	0.095	0.099	0.104	0.106	0.103	0.116	0.117	0.116	0.103

Table 4.10: Size-adjusted power of nominal 0.10-level tests for forecast encompassing

	$\theta^* = 0.0$				$\theta^* = 0.5$				$\theta^* = 0.9$			
	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$	$T = 8$	$T = 16$	$T = 32$	$T = 64$
$h = 1$												
MDM_{non}	0.648	0.690	0.693	0.678								
MDM_{SR}	0.648	0.690	0.693	0.678								
MDM_b	0.648	0.690	0.693	0.678								
DM_{CI}	0.602	0.613	0.640	0.644								
$h = 2$												
MDM_{non}	0.529	0.630	0.680	0.657	0.497	0.570	0.595	0.578	0.460	0.537	0.555	0.542
MDM_{SR}	0.553	0.633	0.680	0.657	0.511	0.571	0.595	0.578	0.463	0.539	0.555	0.542
MDM_b	0.528	0.633	0.680	0.657	0.490	0.569	0.595	0.578	0.452	0.538	0.555	0.542
DM_{CI}	0.603	0.622	0.659	0.634	0.547	0.552	0.568	0.548	0.516	0.518	0.530	0.519
$h = 3$												
MDM_{non}	0.496	0.578	0.635	0.653	0.471	0.523	0.598	0.592	0.422	0.482	0.531	0.513
MDM_{SR}	0.542	0.597	0.637	0.653	0.508	0.539	0.599	0.592	0.449	0.491	0.531	0.513
MDM_b	0.505	0.583	0.635	0.653	0.467	0.530	0.599	0.592	0.419	0.481	0.532	0.513
DM_{CI}	0.598	0.621	0.639	0.630	0.559	0.560	0.594	0.579	0.510	0.509	0.524	0.507
$h = 4$												
MDM_{non}	0.467	0.537	0.617	0.647	0.446	0.502	0.592	0.604	0.411	0.460	0.531	0.529
MDM_{SR}	0.541	0.560	0.626	0.649	0.516	0.527	0.597	0.603	0.458	0.480	0.535	0.530
MDM_b	0.479	0.532	0.625	0.649	0.466	0.507	0.595	0.604	0.418	0.465	0.534	0.530
DM_{CI}	0.609	0.592	0.646	0.658	0.562	0.563	0.616	0.594	0.513	0.512	0.548	0.520
$h = 5$												
MDM_{non}	0.436	0.523	0.599	0.639	0.415	0.504	0.577	0.610	0.401	0.464	0.521	0.538
MDM_{SR}	0.539	0.567	0.619	0.643	0.513	0.543	0.588	0.610	0.489	0.499	0.529	0.539
MDM_b	0.472	0.536	0.609	0.642	0.448	0.516	0.578	0.610	0.428	0.474	0.525	0.539
DM_{CI}	0.599	0.610	0.653	0.650	0.579	0.582	0.625	0.613	0.532	0.528	0.564	0.543
$h = 6$												
MDM_{non}	0.389	0.491	0.580	0.625	0.392	0.475	0.558	0.606	0.363	0.446	0.514	0.550
MDM_{SR}	0.530	0.576	0.616	0.634	0.498	0.536	0.591	0.611	0.464	0.482	0.527	0.555
MDM_b	0.468	0.532	0.600	0.634	0.447	0.504	0.579	0.609	0.407	0.456	0.523	0.554
DM_{CI}	0.590	0.612	0.664	0.662	0.579	0.593	0.626	0.636	0.526	0.533	0.564	0.567

Table 4.11: Fitted MA coefficients of fe_t^4 , fe_t^8 and fe_t^{12}

	fe_t^4	fe_t^8	fe_t^{12}
θ_1	0.926	1.335	1.170
θ_2	1.075	1.414	1.232
θ_3	0.832	1.390	1.235
θ_4		0.692	0.826
μ	0.001	0.018	0.012

Table 4.12: Frequency of negative long-run variance estimates in tests for equal forecast accuracy for $h = 4, 8, 12$

	$h = 4$	$h = 8$	$h = 12$
<i>Panel A: Calibrated errors</i>			
$R = 1$	0.006	0.135	0.330
$R > 1$	0.004	0.139	0.334
<i>Panel B: Simulated errors</i>			
$R = 1$	0.031	0.175	0.343
$R > 1$	0.032	0.188	0.349

Table 4.13: Empirical size of 0.10 level tests under calibrated MA errors

	$h = 4$	$h = 8$	$h = 12$
MDM_{rej}	0.152	0.291	0.429
MDM_{non}	0.146	0.156	0.099
MDM_{SR}	0.148	0.203	0.212
MDM_b	0.149	0.210	0.238
DM_{CI}	0.105	0.107	0.120

Table 4.14: Size-adjusted power of 0.10 level tests under calibrated MA errors

	$h = 4$	$h = 8$	$h = 12$
MDM_{non}	0.506	0.374	0.328
MDM_{SR}	0.501	0.423	0.472
MDM_b	0.498	0.403	0.421
DM_{CI}	0.556	0.627	0.665

Figure 4.1: $\hat{C}(k)$ and γ_k values for $T = 8, h = 6$

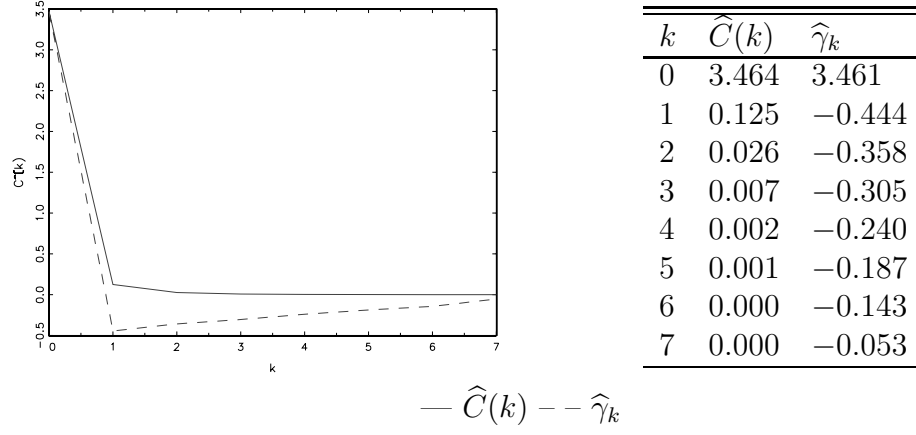
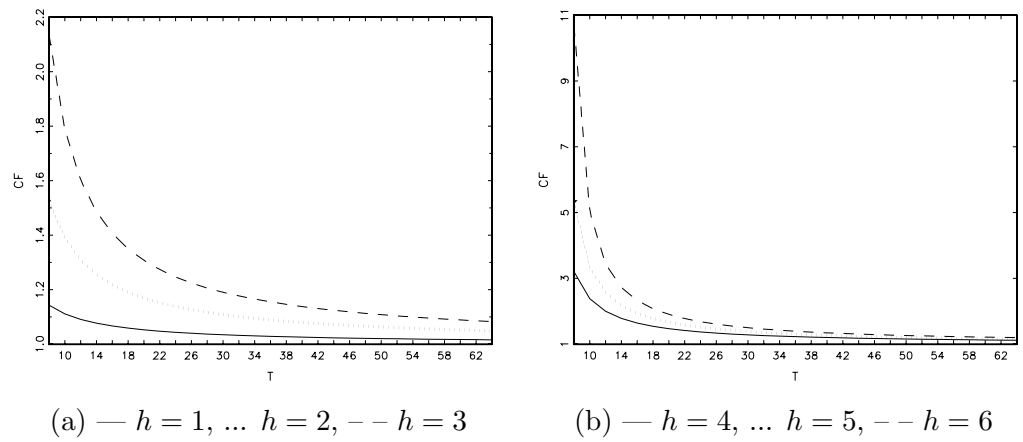


Figure 4.2: Bias Correction Factor (CF)



Chapter 5

Conclusion

This thesis has discussed robust testing methodologies in three areas of time series econometrics: detection of explosive processes, unit root testing against nonlinear alternatives, and forecast evaluation in small samples.

Chapter 2 proposed a union of rejections strategy for detecting temporary explosiveness based on inference from OLS and GLS variants of the PWY test of Phillips et al. (2011) [PWY]. We demonstrated that the original OLS version of the PWY test achieved higher relative power when a small proportion of the data was explosive, but that the GLS-based version had superior power for explosive regimes of longer duration. These power differences were small when the initial condition was asymptotically negligible, but were magnified by the presence of a non-negligible initial condition. We discussed two possible ways of modelling the initial value of a time series. Either by setting a non-zero mean, or by setting an asymptotically non-negligible initial condition. We demonstrated that, under an alternative hypothesis of explosiveness, the first of these would have no impact on the properties of any mean-invariant test, but that the second would affect the magnitude of the explosive process, and therefore the power of any test procedure. Whilst PWY considered only a mean-invariant test for explosive processes, the presence of linear trends in many economic time series motivated us to consider trend-invariant versions of both OLS and GLS based PWY tests. We found a similar pattern of behaviour across mean-invariant and trend-invariant PWY tests. The mean-invariant and trend-invariant union of rejections procedures were shown to capitalise on the respective power advantages of the individual tests. An empirical application to daily NASDAQ prices demonstrated the differing inference that could be made from OLS and GLS based approaches, and the advantages of employing a union of rejections testing strategy.

In this chapter we allowed for only one explosive regime within the sample, but of course it is possible that the time series being analysed features more

than one period of temporary explosiveness. Phillips et al. (2015) propose a generalised version of the PWY testing procedure which is able to detect these multiple bubbles, again based on sub-sample OLS DF tests. A possible extension of the work in this chapter would be to consider the power performance of a GLS version of this testing procedure for different lengths of the explosive period, different numbers of explosive regimes, and different values of the initial condition. Additionally, this chapter addressed only the detection of explosive periods and did not consider how a GLS based procedure might perform when dating explosive regimes. Future research could consider this.

Chapter 3 proposed a union of rejections strategy to test for a unit root against an alternative hypothesis of nonlinear, ESTAR stationarity under two sources of uncertainty: the degree of nonlinearity in the model; and the possible presence of a linear deterministic trend. The Kapetanios et al. (2003) [KSS] test was shown to have superior power over the standard Dickey-Fuller [DF] test when the degree of nonlinearity in the model was large, but this power ranking was reversed for small degrees of nonlinearity. Limiting distributions of KSS and DF tests were derived under a local ESTAR alternative when a local trend is present in the DGP. We demonstrated that whilst mean-invariant tests had a power advantage over trend-invariant tests when no trend was present in the DGP, the power of mean-invariant tests deteriorates as the magnitude of the trend increases. Our union of rejections procedure combined inference from all four individual tests considered in this chapter, and we showed that it captured most of the available power from each individual test across different degrees of nonlinearity and trend magnitudes. Incorporating a trend detection procedure into this union testing strategy resulted in higher power when a large trend is present in the DGP. An empirical application of the proposed union of rejections procedures to energy consumption data in 180 countries demonstrated the practical use of these tests.

As discussed in this chapter, this work focused exclusively on the issue of testing for a unit root when the alternative hypothesis could be either linear or nonlinear stationarity. Estimating ESTAR models, or distinguishing whether a stationary time series follows either a linear or specific nonlinear process, is beyond the scope of this thesis. However, we note that this is another important issue for practitioners, and an interesting area for future research. In this chapter, we only allow a trend to enter the model linearly in line with KSS and DK. A possible extension to this work could consider allowing trends to enter the model in a nonlinear fashion. In much the same way, we also assume that any serial correlation in the data enters linearly, implying that unit root tests can be augmented in the standard way to deal with this. Allowing for nonlinear serial

correlation might require new robust testing strategies. Finally, we could consider extending our proposed union of rejections strategy to tests for cointegration or testing within a panel data setting. We leave these ideas for future work.

Chapter 4 highlighted the frequency with which negative long-run variance estimates could occur when computing the Diebold and Mariano (1995) and Harvey et al. (1997) tests for equal accuracy and the Harvey et al. (1998) test for forecast encompassing in small samples. Given the limitations of economic data, and the need for forecasters to split their total sample into an in sample period for model estimation and an out of sample period for forecast evaluation, it is not uncommon for tests of equal accuracy and encompassing to be conducted using sample size and forecast horizon settings that lie in the region where negative variance estimates occur frequently. A number of different strategies for dealing with the problem of negative long-run variance estimation in tests of equal accuracy and encompassing were therefore considered in an extensive Monte Carlo exercise. The recently proposed Coroneo and Iacone (2015) approach was found to outperform the other tests considered in this chapter, with the attractive finite sample properties reported in their paper for moderate sample sizes and forecast horizons extending to the small sample and long horizon region under focus here. The empirical relevance of our results was confirmed through an application of the tests in this chapter to errors calibrated from forecasts of Euro-area inflation.

Forecast evaluation is an important topic not only for point forecasts, but also in probability forecasting where the use of DM tests for equal accuracy is widespread. A possible extension to the work in this chapter could be to examine whether negative long-run variance estimates affect probability forecast evaluation in the same way as they do here.

This thesis contains three substantive chapters which each focus on an area of time series econometrics in which we find that the performance of test procedures is dependent on a series of factors which are either unobservable or unavoidable in practice. In Chapter 2 we showed that recently proposed tests for detecting temporary explosive processes are dependent on the length of the explosive period and the value of the initial condition. Chapter 3 highlighted how the performance of unit root tests were affected by the degree of nonlinearity under the alternative hypothesis and the presence of a linear trend. In Chapter 4 we showed that tests for equal forecast accuracy and forecast encompassing can have undesirable finite sample properties as the number of out of sample observations decreases and the forecast horizon increases. In all three chapters we recommend a robust testing methodology which achieves desirable size and power properties regardless of these factors. Applications of these procedures to empirical data highlighted

the potential advantages these test procedures can have over existing tests in practice.

Bibliography

- Abuaf, N. and Jorion, P. (1990), ‘Purchasing power parity in the long run’, *Journal of Finance* **45**, 157–174.
- Akdogan, K. (2014), ‘Threshold adjustment in the current account: sustainability for danger zone economies?’, *Applied Economics Letters* **21**,14, 1006–1009.
- Andrews, D. K. (1991), ‘Heteroskedasticity and autocorrelation consistent covariance matrix estimation’, *Econometrica* **59**,3, 817–858.
- Baharumshah, A. Z., Liew, V. K. S. and Haw, C. T. (2009), ‘The real interest rate differential: international evidence based on non-linear unit root tests’, *Bulletin of Economic Research* **61**, 83–94.
- Bates, J. M. and Granger, C. W. J. (1969), ‘The combination of forecasts’, *Operational Research Quarterly* **20**,4, 451–468.
- Beechey, M. and Osterholm, P. (2008), ‘Revisiting the uncertain unit root in GDP and CPI: Testing for non-linear trend reversion’, *Economics Letters* **100**, 221–223.
- Blanchard, O. J. and Watson, M. (1982), Bubbles, rational expectations and financial markets, in P. Wachtel, ed., ‘Crises in the Economic and Financial Structure’, Lexington Books, Lexington, MA., pp. 295–315.
- Caporale, G. M. and Gil-Alana, L. (2014), ‘Long-run and cyclical dynamics in the US stock market’, *Journal of Forecasting* **33**, 147–161.
- Chen, S. W. (2014), ‘Smooth transition, non-linearity and current account sustainability: Evidence from the European countries’, *Economic Modelling* **38**, 541–554.
- Chong, Y. Y. and Hendry, D. F. (1986), ‘Econometric evaluation of linear macroeconomic models’, *Review of Economic Studies* **53**, 671–690.

- Chow, H. K. and Choy, K. M. (2006), ‘Forecasting the global electronics cycle with leading indicators: A Bayesian VAR approach’, *International Journal of Forecasting* **22**, 301–315.
- Christidou, M., Panagiotidis, T. and Sharma, A. (2013), ‘On the stationarity of per capita carbon dioxide emissions over a century’, *Economic Modelling* **33**, 918–925.
- Clark, T. E. (1999), ‘Finite-sample properties of tests for equal forecast accuracy’, *Journal of Forecasting* **18**, 7, 489–504.
- Clark, T. E. and McCracken, M. W. (2001), ‘Tests of equal forecast accuracy and encompassing for nested models’, *Journal of Econometrics* **105**, 1, 85–110.
- Clark, T. E. and McCracken, M. W. (2013), Advances in forecast evaluation, in G. Elliott and A. Timmerman, eds, ‘Handbook of Economic Forecasting, Volume 2’, Elsevier, pp. 1107–1201.
- Clements, M. P. and Harvey, D. I. (2009), Forecast combination and encompassing, in T. C. Mills and K. Patterson, eds, ‘Palgrave Handbook of econometrics, Volume 2, Part B’, Palgrave Macmillan, Basingstoke, pp. 169–198.
- Cook, S. (2008), ‘More uncertainty: on the trending nature of real GDP in the US and UK’, *Applied Economics Letters* **15**, 9, 667–670.
- Coroneo, L. and Iacone, F. (2015), ‘Comparing predictive accuracy in small samples’, *Discussion Papers in Economics No.15/15, Department of Economics, University of York*.
- Davies, R. B. (1977), ‘Hypothesis testing when a nuisance parameter is present only under the alternative’, *Biometrika* **64**, 247–254.
- Davies, R. B. (1987), ‘Hypothesis testing when a nuisance parameter is present only under the alternatives’, *Biometrika* **74**, 33–43.
- Demetrescu, M. and Kruse, R. (2012), ‘The power of unit root tests against nonlinear local alternatives’, *Journal of Time Series Analysis* **34**, 40–61.
- Dib, A., Gammoudi, M. and Moran, K. (2008), ‘Forecasting Canadian time series with the New Keynesian model’, *Canadian Journal of Economics* **41**, 138–165.
- Diba, B. and Grossman, H. (1988), ‘Explosive rational bubbles in stock prices’, *American Economic Review* **78**, 520–530.

- Dickey, D. A. and Fuller, W. A. (1979), ‘Distribution of the estimators for autoregressive time series with a unit root’, *Journal of the American Statistical Association* **74**, 427–431.
- Diebold, F. X. and Mariano, R. S. (1995), ‘Comparing predictive accuracy’, *Journal of Business and Economics Statistics* **13,3**, 253–263.
- Dreger, C. and Wolters, J. (2014), ‘Money demand and the role of monetary indicators in forecasting euro area inflation’, *International Journal of Forecasting* **30**, 303–312.
- Edison, H. and Klovland, J. (1987), ‘A quantitative reassessment of the PPP hypothesis’, *Journal of Applied Econometrics* **2**, 309–333.
- Elliott, G., Rothenberg, T. and Stock, J. (1996), ‘Efficient tests for an autoregressive unit root’, *Econometrica* **64,4**, 813–836.
- Evans, G. W. (1991), ‘Pitfalls in testing for explosive bubbles in asset prices’, *American Economic Review* **81**, 922–930.
- Fair, R. C. and Shiller, R. J. (1989), ‘The informational content of ex ante forecasts’, *Review of Economics and Statistics* **71**, 325–331.
- Flood, R. P. and Garber, P. M. (1980), ‘Market fundamentals versus price-level bubbles: The first tests’, *Journal of Political Economy* **88,4**, 745–770.
- Gikhman, I. I. and Skorokhod, A. V. (1969), *Introduction to the Theory of Stochastic Processes*, Saunders, Philadelphia.
- Granger, C. W. J. and Newbold, P. (1973), ‘Some comments on the evaluation of economic forecasts’, *Applied Economics* **5,1**, 35–47.
- Granger, C. W. J. and Newbold, P. (1977), *Forecasting Economic Time Series*, Academic Press, Orlando, FL.
- Guney, P. O. and Hasanov, M. (2014), ‘Real interest rate parity hypothesis in post-Soviet countries: Evidence from unit root tests’, *Economic Modelling* **36**, 120–129.
- Hamilton, J. D. (1986), ‘On testing for self-fulfilling speculative price bubbles’, *International Economic Review* **27,3**, 545–552.
- Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

- Hamilton, J. D. and Whiteman, C. H. (1985), ‘The observable implications of self-fulfilling speculative price bubbles’, *Journal of Monetary Economics* **16**, 353–373.
- Hanck, C. (2012), ‘On the asymptotic distribution of a unit root test against ESTAR alternatives’, *Statistics and Probability Letters* **82**, 360–364.
- Harvey, D. I., Leybourne, S. J. and Newbold, P. (1997), ‘Testing the equality of prediction mean squared errors’, *International Journal of Forecasting* **13**, 2, 281–291.
- Harvey, D. I., Leybourne, S. J. and Newbold, P. (1998), ‘Tests for forecast encompassing’, *Journal of Business and Economic Statistics* **16**, 2, 254–259.
- Harvey, D. I., Leybourne, S. J. and Newbold, P. (1999), ‘Forecast evaluation tests in the presence of ARCH’, *Journal of Forecasting* **18**, 435–445.
- Harvey, D. I., Leybourne, S. J. and Sollis, R. (2015), ‘Recursive right-tailed unit root tests for an explosive asset price bubble’, *Journal of Financial Econometrics* **13**, 166–187.
- Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2009), ‘Unit root testing in practice: Dealing with uncertainty over the trend and initial condition’, *Econometric Theory* **25**, 587–636.
- Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2012), ‘Testing for unit roots in the presence of uncertainty over both the trend and initial condition’, *Journal of Econometrics* **169**, 188–195.
- Harvey, D. and Leybourne, S. (2014), ‘Asymptotic behaviour of tests for a unit root against an explosive alternative’, *Economics Letters* **122**, 64–68.
- Hasanov, M. and Telatar, E. (2011), ‘A re-examination of stationarity of energy consumption: Evidence from new unit root tests’, *Energy Policy* **39**, 7726–7738.
- Hering, A. S. and Genton, M. G. (2011), ‘Comparing spatial predictions’, *Technometrics* **53**, 4, 414–425.
- Hualde, J. and Iacone, F. (2017), ‘Autocorrelation robust inference using the Daniell kernel with fixed bandwidth’, *Economics Letters* **forthcoming**.
- Hubbert, M. K. (1956), ‘Nuclear energy and the fossil fuels’, *American Petroleum Institute, Shell Development Company* pp. San Antonio, Texas.

- Kapetanios, G., Shin, Y. and Snell, A. (2003), ‘Testing for a unit root in the nonlinear STAR framework’, *Journal of Econometrics* **112**, 359–379.
- Kisswani, K. M. and Nusair, S. A. (2013), ‘Non-linearities in the dynamics of oil prices’, *Energy Economics* **36**, 341–353.
- Kurtz, T. G. and Protter, P. (1991), ‘Weak limit theorems for stochastic integrals and stochastic differential equations’, *The Annals of Probability* **19**, 1035–1070.
- Leroy, S. and Porter, R. (1981), ‘The present value relation: Tests based on implied variance bounds’, *Econometrica* **64**, 555–574.
- Lothian, J. R. and Taylor, M. P. (1996), ‘Real exchange rate behavior: The recent float from the perspective of the past two centuries’, *Journal of Political Economy* **104**, 3, 488–509.
- Marsh, P. (2007), ‘The available information for invariant tests of a unit root’, *Econometric Theory* **23**, 686–710.
- Maslyuk, S. and Smyth, R. (2009), ‘Non-linear unit root properties of crude oil production’, *Energy Economics* **31**, 109–118.
- Meese, R. and Rogoff, K. (1988), ‘Was it real? The exchange rate-interest differential relation over the modern floating-rate period’, *Journal of Finance* **43**, 4, 933–948.
- Mehl, A. (2009), ‘The yield curve as a predictor and emerging economies’, *Open Economics Review* **20**, 683–716.
- Michael, P., Nobay, R. A. and Peel, D. A. (1997), ‘Transactions costs and non-linear adjustment in real exchange rates: an empirical investigation’, *Journal of Political Economy* **105**, 862–879.
- Muller, U. K. and Elliott, G. (2003), ‘Tests for unit roots and the initial condition’, *Econometrica* **71**, 1269–1286.
- Nelson, C. R. (1972), ‘The prediction performance of the FRBMIT-PENN model of the US economy’, *American Economic Review* **62**, 902–917.
- Newey, W. K. and West, K. D. (1987), ‘A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix’, *Econometrica* **55**, 703–708.
- Newey, W. K. and West, K. D. (1994), ‘Automatic lag selection in covariance matrix estimation’, *Review of Economic Studies* **61**, 4, 631–653.

- Ng, S. and Perron, P. (1995), ‘Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag’, *Journal of the American Statistical Association* **90**, 268–281.
- Ng, S. and Perron, P. (2001), ‘Lag length selection and the construction of unit root tests with good size and power’, *Econometrica* **69**, 1519–1554.
- Norman, S. (2009), ‘Testing for a unit root against ESTAR nonlinearity with a delay parameter greater than one’, *Economics Bulletin* **29(3)**, 2152–2173.
- Phillips, P. C. B. (1987), ‘Towards a unified asymptotic theory for autoregression’, *Biometrika* **74**, 535–547.
- Phillips, P., Shi, S. and Yu, J. (2015), ‘Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500’, *International Economic Review* **56**, 1043–1078.
- Phillips, P., Wu, Y. and Yu, J. (2011), ‘Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values?’, *International Economic Review* **52,1**, 201–226.
- Qin, D., Cagas, M., Ducanes, G., Magtibay-Ramos, N. and Quising, P. (2008), ‘Automatic leading indicators versus macroeconomic structural models: A comparison of inflation and GDP growth forecasting’, *International Journal of Forecasting* **24**, 399–413.
- Sercu, P., Uppal, R. and Van Hulle, C. (1995), ‘The exchange rate in the presence of transaction costs: implications for tests of purchasing power parity’, *Journal of Finance* **50**, 1309–1319.
- Shiller, R. J. (1981), ‘Do stock prices move too much to be justified by subsequent changes in dividends?’, *American Economic Review* **71,3**, 421–436.
- Smyth, R. (2013), ‘Are fluctuations in energy variables permanent or transitory? a survey of the literature on the integration properties of energy consumption and production’, *Applied Energy* **104**, 371–378.
- Sollis, R. (2004), ‘Asymmetric adjustment and smooth transitions: A combination of some unit root tests’, *Journal of Time Series Analysis* **25**, 409–417.
- Su, J. J., Cheung, A. W. K. and Roca, E. (2013), ‘Lag selection of the augmented Kapetanios-Shin-Snell nonlinear unit root test’, *Journal of Mathematics and Statistics* **9,2**, 102–111.

- Taylor, M. P., Peel, D. A. and Sarno, L. (2001), ‘Nonlinear mean-reversion in real exchange rates: towards a solution to the purchasing power parity puzzles’, *International Economic Review* **42**, 1015–1042.
- Terasvirta, T. (1994), ‘Specification, estimation, and evaluation of smooth transition autoregressive models’, *Journal of the American Statistical Association* **89**, 208–218.
- Tsong, C. C. and Lee, C. F. (2013), ‘Further evidence on real interest rate equalization: Panel information, non-linearities and structural changes’, *Bulletin of Economic Research* **65**, 85–105.
- van Dijk, D., Terasvirta, T. and Franses, P. H. (2002), ‘Smooth transition autoregressive models a survey of recent developments’, *Econometric Reviews* **21**, 1–47.
- Vogelsang, T. (1998), ‘Trend function hypothesis testing in the presence of serial correlation’, *Econometrica* **66**, 123–148.
- West, K. D. (1988), ‘Dividend innovations and stock price volatility’, *Econometrica* **56**, 1, 37–61.
- West, K. D. (1996), ‘Asymptotic inference about predictive ability’, *Econometrica* **64**, 1067–1084.
- West, K. D. (2006), Forecast evaluation, *in* G. Elliott and A. Timmerman, eds, ‘Handbook of Economic Forecasting, Volume 1’, Elsevier, pp. 100–132.