Testing for non-linearity and asymmetry in time series

Testing for Non-linearity and Asymmetry in Time Series

Marián Vávra

Department of Economics, Mathematics, and Statistics

Birkbeck College

University of London

United Kingdom

A thesis submitted for the degree of Doctor of Philosophy

October 2013

Dedication

To my family

Declaration

I declare that the work presented in this Ph.D. thesis is my own original work. Where information has been derived from other sources, I confirm that this has been clearly and fully identified and acknowledged. No part of the thesis contains material previously submitted to this University or to any other institution for a degree.

....

Marián Vávra
London, UK
October 2013

Abstract

The Ph.D. thesis, called *Testing for Non-linearity and Asymmetry in Time Series*, focuses on various issues related to testing for non-linearity and marginal asymmetry of economic time series. This is an important issue since testing for non-linearity and/or asymmetry represents an early, yet crucial, step in the whole process of time series modelling. A mistake in this preliminary step may lead to model misspecification, and, subsequently, to a sequence of related issues throughout all the modelling steps (i.e. identification, estimation, and forecasting). As a result, this type of mistakes is very likely to result in wrong business or economic policy decisions. The thesis is divided into six chapters.

The first chapter explains the motivation for the thesis.

The second chapter, called *Robustness of the Power of Non-linearity Tests*, examines the statistical properties of the selected univariate non-linearity tests under different conditions. In particular, special attention is paid to the robustness of the power properties of the tests against moment condition failure of innovations, asymmetry of innovations, and the parameter configuration of data generating processes. Since analytical results are available only for a very limited number of the test statistics, an extensive Monte Carlo approach is implemented instead. The Monte Carlo results reveal that the power of the selected non-linearity tests is statistically significantly inflated under asymmetry of innovations and moment condition failure.

In the third chapter, called *Testing for Non-linearity Using a Modified Q Test*, a new version of the portmanteau Q test, based on auto- and cross-correlations, is

developed. The main task of this chapter is to propose a new type of the Q test in order to bypass some of the shortcomings of the McLeod and Li Q test discovered in Chapter 2. Our results, based on extensive Monte Carlo experiments, suggest the proposed Q test significantly improves the power against some non-linear time series models (e.g. threshold autoregressive and moving average models) and is capable to detect some interesting non-linear processes (e.g. non-linear moving average models), for which the standard Mcleod and Li Q test completely fails.

In the fourth chapter, called *Testing for Marginal Asymmetry in Time Series*, a modified test for symmetry of the marginal law of weakly dependent processes is proposed. The test statistic is based on sample quantiles. It is shown that the test has an intuitive interpretation, it is easy and fast to calculate, it it follows a standard limiting distribution, and much more importantly, it is robust against weak dependence of observations. Especially the last feature makes the test very attractive for the use in applied economics since it minimizes inferential errors due to the incorrect configuration of the test. The finite sample properties of the test are examined via Monte Carlo experiments. The results suggest that the quantile-based test of symmetry performs very well.

In the fifth chapter, called *Testing for Non-linearity in Multivariate Time Series*, two new principal component-based multivariate non-linearity tests are considered. The main goal of this chapter is to modify two well known multivariate test statistics which suffer from the curse of dimensionality. It is shown that a dimensionality problem can be easily bypassed by means of a principal component analysis. Our results, based on extensive Monte Carlo experiments, suggest that a principal component analysis reduces the dimensionality problem very efficiently without any systematic power distortion. The results also reveal that the BIC stopping rule performs best in determining the number of components for the selected multivariate non-linearity tests.

The last chapter summarizes the results of this thesis and discusses directions for further research.

Acknowledgements

I would like to express my deep appreciation and thankfulness to my supervisor, Professor Zacharias Psaradakis, for his invaluable guidance throughout the period of writing this Ph.D. thesis. In addition, I wish to thank Professor John Driffill, Professor Ron Smith, Professor Peter Tinsley, Dr. Yunus Aksoy, and Dr. Walter Beckert, all from the University of London, for their constructive remarks and interesting suggestions which improved the thesis. Finally, I am very grateful to Professor Adrian Pagan from the University of Sydney and Professor Timo Teräsvirta from the Aarhus University for fruitful comments and recommendations.

I am grateful for the hospitality I received at the ECB, where a part of my research was carried out.

And finally, from all my heart I wish to thank my family for their encouragement, courageous support, help, and understanding during all these years. My special thanks go to Associate Professor Luboš Marek and Associate Professor Jiří Trešl, my university teachers from the Prague School of Economics, for encouraging me to study a doctorate in the UK.

Marian Vavra

London, UK

October 2013

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Since the Great depression in 1930's, the question whether economic variables do exhibit the non-linear and/or asymmetric behaviour over business cycles or crises has attracted much interest in economics and econometrics. Burns and Mitchell (1946), pioneers of a modern business cycle analysis, have suggested, based on investigating the behaviour of the selected US economic indicators such as industrial production, investments, employment, etc., that business cycles do display significant asymmetry in the sense that contraction phases are short and deep as compared to the expansion phases. Neftci (1984), Sichel (1993), McQueen and Thorley (1993), Speight and McMillan (1998), Psaradakis and Sola (2003), Clements and Krolzig (2003), among others, have used modern econometric techniques to test for various types of business cycle asymmetry (i.e. steepness, deepness, or sharpness).[1] All these studies have confirmed the previous results that some economic indicators do behave asymmetrically, although noticeable differences are observed across the papers.[2] Recently, a form of business cycle asymmetry has been recognized in the second moment of economic indicators as well. French and Sichel (1993) have pointed out that eco-

[1]Definitions of the above mentioned types of business cycle asymmetries can be found in Clements and Krolzig (2003, p. 199).

[2]For example, Neftci (1984) has found evidence of steepness in the US unemployment rate, whereas Sichel (1993) has reported evidence of deepness in US industrial production.

nomic indicators from cyclically sensitive sectors (e.g. durable consumption goods and investment goods) do exhibit significant asymmetry in a conditional variance in the sense that negative shocks significantly increase conditional volatility whereas positive shocks do not. Further investigation links this asymmetry to the phase of the business cycle (i.e. volatility appears to be largest around the business cycle troughs). Apart from the presence of business cycle asymmetry, overwhelming evidence of some other non-linear features (e.g. bilinearity, regime-switching, conditional volatility) has been found in economic time series such as stock returns (Hinich and Patterson (1985)), exchange rates (Peel and Speight (1994), Hsieh (1989), Sarno et al. (2004)), interest rates (Anderson (1997), Ang and Bekaert (2002)), and industrial production or gross domestic (national) product (Teräsvirta and Anderson (1992), Peel and Speight (1998a,b), van Dijk and Franses (1999), Escribano and Jordá (2001)).

Such investigations are an early, yet crucial, step in establishing the stylized facts about economic time series. The understanding of these stylized facts is important for at least three reasons:

(i) Econometrics – The presence of a statistically significant non-linear and/or asymmetric pattern in time series restricts the class of statistical models that can adequately describe the main stochastic features of time series. As a result, very popular linear ARIMA models (see Box et al. (2008) or Brockwell and Davis (1991), among others) are very unlikely to provide a valid data description and an accurate forecast.

(ii) Applied economics – Model misspecification can, in turn, give rise to misleading economic policy decisions. For example, the traditional wisdom of monetary policy is that a central bank should set the policy interest rate according to (expected) economic fundamentals such as the inflation rate and the output gap, see Woodford (2011) for a comprehensive treatment about monetary policy. However, Davig (2007) has examined the implications of the changing slope of the Phillips curve for optimal discretionary monetary policy. He has found that the slope parameter of the Phillips curve following a Markov switching process can create significant instability in the inflation rate. In such a case, the utility-based welfare criterion instructs monetary policy to disregard the switching effect of the Phillips curve and keeps the policy interest rate constant across regimes. This stands in sharp contrast to the traditional wisdom, which would advise monetary policy to change the policy interest rate according to the inflation rate. However, as pointed out by Kozicki and Tinsley (2002), injudicious and/or aggressive changes in the policy interest rate involve an economically important increase in term premia, and, consequently, the long-term interest rate. As a result, high long-term interest rates are very likely to cut economic activity in the economy. So, it can be concluded that omitting business cycle asymmetry and/or non-linearity may cause serious problems for a central bank when conducting monetary policy in practice.

(iii) Theoretical economics – The presence of business cycle asymmetry in economic activity indicators has had implications for developing new theoretical models in economics. For example, Acemoglu and Scott (1997) have built a new economic model where business cycle fluctuations are based on intertemporal increasing returns in the economy. They show that this model specification is helpful in explaining asymmetric business fluctuations even in the case of identically and independently distributed economic shocks hitting the economy because individuals respond differently to shocks depending on their past investment activity.

Theoretical contributions can be found in finance as well. It is nowadays well understood that standard portfolio diversification in a mean-variance setup, assuming a perfect capital market with risk averse investors, concludes that adding randomly selected and equally weighted assets to a portfolio leads to a risk reduction without any effect on returns, see Markowitz (1968) for details. However, as shown in Conine and Tamarkin (1981) and Prakash et al. (2003), the presence of marginal asymmetry in asset returns can fundamentally change portfolio diversification, it means the number of selected assets and their optimal weights.

1.2 Motivation

It is important to emphasize that the stylized facts might be useful only if these facts have been obtained by applying appropriate statistical methods and all conditions, required by these methods, have been satisfied. Unfortunately, this is not always the case. Therefore, I focus on various issues related to testing for non-linearity and marginal asymmetry of economic time series in the submitted Ph.D. thesis. In particular, the thesis brings four contributions to the area of testing for non-linearity and asymmetry of economic time series:

(i) The statistical properties of the selected non-linearity tests are inspected under moment condition failure and asymmetry of model innovations;

(ii) A portmanteau Q test based on generalized correlations is developed;

(iii) A new test of marginal asymmetry based on quantiles is proposed;

(iv) Two new multivariate non-linearity tests based on principal components are discussed.

Motivation for individual theoretical contributions (chapters) is as following:

(i) Chapter 2 – Statistical methods do require some conditions to be satisfied in order to provide correct inference. The non-linearity tests, routinely applied in the literature, are no exception. Among all possible conditions, the existence of a particular number of moments is a joint condition for all the non-linearity tests. The following example illustrates our point. Since a conditional variance has been found to be a common feature of many economic variables, the portmanteau Q test, proposed by McLeod and Li (1983), has become one of the most widely used tests in the literature. The test is based on inspecting the correlation structure of squared residuals. As a result, the test requires the existence of the first eight moments to have a valid limiting distribution. Yet, this is in sharp contrast with empirical findings about economic time series, for which a tail index, determining the highest finite moment of a random variable, usually lies between 2 and 4, see Jansen and Vries (1991), Runde (1997), Koedijk et al. (1990), and Bali (2003), among others. So, a couple of natural questions arise: "*How sensitive are the power properties of the non-linearity tests to moment condition failure?* or "*Do the tests with the minimal moment condition behave better in general?*" Since analytical results are infeasible to get in general, I approach these questions by means of extensive Monte Carlo experiments. In particular, I concentrate on assessing the power properties of the selected eight non-linearity tests under moment condition failure and asymmetry of innovations.

(ii) Chapter 3 – Practitioners always face a dilemma which non-linearity test to apply to a given stochastic process at hand. On the one hand, it might be tempting to apply the whole battery of tests to capture all possible types of non-linear features. On the other hand, it might be then extremely difficult to ensure that all necessary conditions of those tests are satisfied. So, a natural question arises: "*Is there any simple test which can capture all well-known non-linear features such as bilinearity, regime switching, and conditional volatility?*" Keeping in mind the above mentioned trade-off, I propose a modified portmanteau Q test based on auto- and cross-correlations. It is shown that the test is a simple extension of the McLeod and Li Q test and follows a standard limiting distribution. The finite sample properties of the new Q test are assessed via extensive Monte Carlo experiments. An empirical example is provided as well.

(iii) Chapter 4 – Testing for asymmetry of the marginal law of economic time series is by no means easy in practice. The problem is that weak dependence of individual observations of (transformed) economic variables invalidates critical values of standard symmetry tests originally derived for independently and identically distributed random variables. So, a natural question is: "*Is there any simple test which is robust against weak dependence?*" I solve this problem by developing a simple test of marginal symmetry based on sample quantiles. It is shown that the test has an intuitive interpretation, it is easy and fast to calculate, and follows a standard limiting distribution. Finite sample properties are inspected via Monte Carlo experiments. An empirical example is provided as well.

(iv) Chapter 5 – The mainstream literature has focused mainly on the use of the univariate non-linear tests when establishing the stylized facts of business cycle non-linearity and/or asymmetry. However, one can argue that a set of economic variables might be dependent each other in a non-linear way. Many examples can be found in economics and finance, see Sims and Zha (2006), Liu et al. (2009), Schmitt-Grohe and Uribe (2004), Rudebusch and Swanson (2008), Engle and Kroner (1995), or Bollerslev (1990), among others. So, a natural question is: "*How do the univariate non-linearity tests work in the context of multivariate time series?*" In order to approach this question correctly, one should compare appropriate univariate and multivariate non-linearity tests. However, it is by no means easy to test for non-linearity in multivariate time series models. The existing multivariate non-linearity tests (e.g. the TSAY test) suffer from the curse of dimensionality. It means that, due to the construction of these tests, they require a large number of observations, which is not feasible to get in applied economics. Therefore, I first show that the dimensionality problem can be easily bypassed by means of a principal component analysis. Then, the results of two univariate and multivariate test statistics are compared via Monte Carlo experiments. Two empirical examples are provided as well.

Chapter 2

Robustness of the Power of Non-linearity Tests

"Using the term non-linear to describe a time series is like saying that zoology is the study of non-elephant animals."

S. Ulam, Nobel Prize winner in physics

2.1 Introduction

There exist many different non-linear time series models and related non-linearity tests in the literature, see Tong (1990, Chapter 3 and 5) for details. However, there are only few comprehensive studies comparing statistical properties of non-linearity tests, see Luukkonen et al. (1988), Lee et al. (1993), de Lima (1997), or Psaradakis and Spagnolo (2002). It is worth noting, however, that even these studies suffer from some of the following three limitations.

First, non-linearity tests are applied to time series models based on a particular fixed parameter configuration. For instance, Lee et al. (1993, p. 277) consider the following simple threshold autoregressive model given by

$$X_t = 0.9X_{t-1}I(|X_{t-1}| \leq 1) - 0.3X_{t-1}I(|X_{t-1}| > 1) + a_t,$$

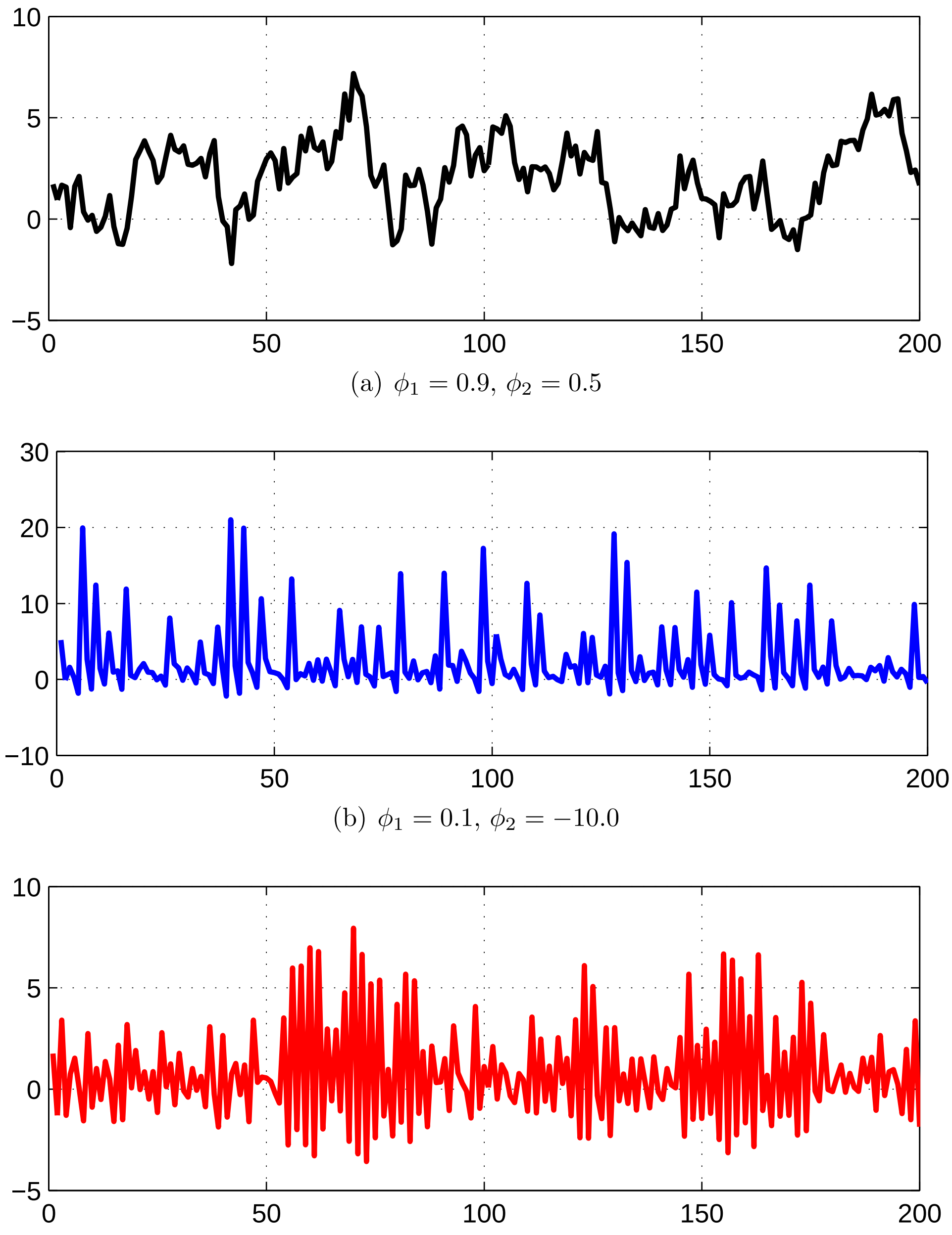
where $I(\cdot)$ is a standard indicator function taking 1 if $|X_{t-1}| \leq 1$ and 0 otherwise, and $\{a_t : t \in \mathbb{Z}\}$ is a sequence of NID(0,1) innovations. A problem is that a change in some parameters of a non-linear model can generate a stochastic process with rather distinct features. To make this point clear, three realizations of a simple threshold autoregressive model with different parameters are presented in Figure 2.1. It is quite clear that realizations are completely different. So, there is no guarantee, at least theoretically, that all non-linearity tests work in the same way for all parameter configurations of a given non-linear process. For this reason, the first part of this chapter examines the robustness of standard non-linearity tests against the parameter configuration of linear and non-linear models.

Second, another problem is that the parameter specification in many research papers does not even satisfy the basic moment conditions required by non-linearity tests. For instance, Luukkonen et al. (1988, p. 170) use the following simple AR-ARCH model given by

$$\begin{aligned} X_t &= 0.6X_{t-1} + \epsilon_t, \\ \epsilon_t &= a_t\sqrt{h_t}, \\ h_t &= 0.2 + 0.8\epsilon_{t-1}^2, \end{aligned}$$

where $\{a_t : t \in \mathbb{Z}\}$ is a sequence of NID(0,1) innovations. A problem is not in the model itself, of course, but in a battery of non-linearity tests applied. Authors consider, among other tests, the Tsay test, which requires the existence of the first four moments, and the McLeod and Li Q test, which requires the existence of even the first eight moments. It is not difficult to show that the above ARCH model does not satisfy either of these two moment conditions. In this case, standard limiting distributions of the above mentioned test statistics are no longer valid and testing non-linearity can lead to misleading results. It would be a serious mistake to think of this particular example as an exception in the literature. Indeed, the opposite is true. In many other papers, although the parameter specification formally satisfies moment conditions, parameters lie very close to or even on the boundary of the parameter space, and thus, do not characterize the stochastic properties of a given

Figure 2.1 Different realizations of a TAR model: Gaussian innovations



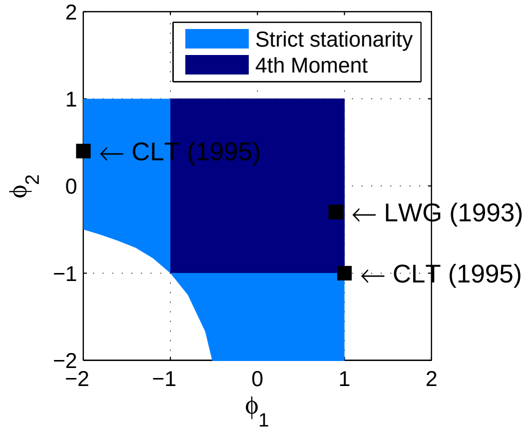
(a) $\phi_1 = 0.9$, $\phi_2 = 0.5$

(b) $\phi_1 = 0.1$, $\phi_2 = -10.0$

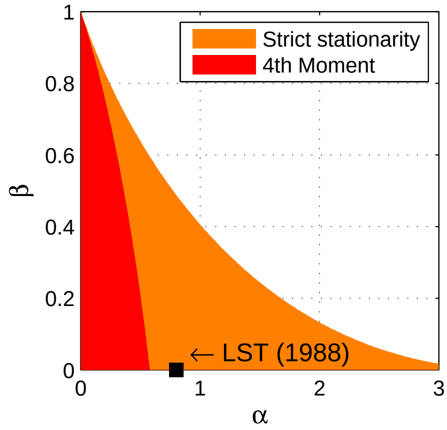
(c) $\phi_1 = 0.4$, $\phi_2 = -2.0$

Note: The series are generated from a simple TAR(2;1,1) model: $X_t = \phi_1 X_{t-1} I(X_{t-1} > 0) + \phi_2 X_{t-1} I(X_{t-1} \leq 0) + a_t$, where $\{a_t : t \in \mathbb{Z}\}$ is a sequence of NID(0,1) innovations. Particular model parameters come from Petruccelli and Woolford (1984).

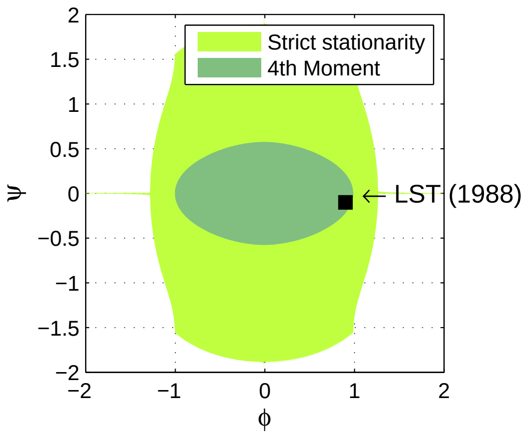
Figure 2.2 Moment failure of non-linear models



(a) TAR model



(b) GARCH model



(c) BL model

Note: CLT (1995) stands for Chen et al. (1995), LST (1988) denotes Luukkonen et al. (1988), and LWG (1993) is Lee et al. (1993). Strict stationarity regions are calculated based on an assumption that a ∼NID(0,1), if necessary, whereas 4th-Moment regions represent the intersection of the 4-th moment stationarity and/or invertibility conditions. Series are generated from the following list of models: (a) a TAR(2;1,1) model: $X_t = \phi_1 X_{t-1} I(X_{t-1} > 0) + \phi_2 X_{t-1} I(X_{t-1} \leq 0) + a_t$, (b) a GARCH(1,1) model: $X_t = a_t\sqrt{h_t} = \epsilon_t, h_t = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}$, (c) a BL(1,0,1,1) model: $X_t = \phi X_{t-1} + \theta a_{t-1} + \psi X_{t-1} a_{t-1} + a_t$. We assume that $\{a_t : t \in \mathbb{Z}\}$ is a sequence of NID(0,1) innovations in all models.

process adequately. See Figure 2.2 for a few examples borrowed from the literature. Figures depict strict stationarity and 4th-moment stationarity regions altogether with particular parameter configurations for three well known non-linear time series models.

The main problem with moment condition failure of non-linearity tests is that we cannot always derive an appropriate limiting distribution for a given test statistic. And even if we could, many other statistical issues arise immediately. To make this point clear, let us consider the following stochastic process with an infinite variance

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j},$$

where $\{Z_t : t \in \mathbb{Z}\}$ is a sequence of IID innovations whose distribution $\mathbb{F}$ has Pareto-like tails with the tail index $\kappa \in (1, 2)$, where κ is defined as $\kappa = \sup_{k>0} \mathbb{E}(|Z_t|^k) < \infty$. Although Adler et al. (1998) show that a standard Box–Jenkins approach can be applied in general, a great deal of care needs to be exercised in individual modelling steps (e.g. identification, estimation, and diagnostic checking). The reason for that can be easily demonstrated using a simple portmanteau Q test originally developed by Box and Pierce (1970). Davis and Resnick (1986) show that the estimated sample autocorrelations are not $O_p(T^{-1/2})$ and their limiting distribution is not Gaussian. In particular, they show that

$$\hat{\rho}_k - \rho_k = O_p\left(\left[\frac{T}{\log T}\right]^{-1/\kappa}\right) = o_p(T^{-1/\beta}),$$

for any real $\beta > \kappa$, where ρ_k and $\hat{\rho}_k$ denote theoretical and sample autocorrelations. It means that the sample autocorrelations have slightly faster rate of convergence as compared to those estimated from a process with a finite variance. Provided we incorrectly assume standard $\sqrt{T}$ convergence (i.e. $2 = \beta > \kappa$), then $\sqrt{T}(\hat{\rho}_k - \rho_k) \xrightarrow{d} 0$, which means that the limiting distribution of the sample autocorrelations is degenerated. Moreover, the authors also show that even if we consider a correct normalizing constant, the limiting distribution is given by

$$\left(\frac{T}{\log T}\right)^{1/\kappa} (\hat{\rho}_k - \rho_k) \xrightarrow{d} S_k/S_0,$$

for some integer $k > 0$, and S_k, S_0 are two independent stable variables, see Corollary 1 in Davis and Resnick (1986, p. 553) for a complete proof. Based on the results above, Runde (1997) derived the limiting distribution of the Box-Pierce Q test, which does not converge to a χ^2 distribution anymore, but to a rather complicate law given by

$$Q(m) = \left(\frac{T}{\log T}\right)^{2/\kappa} \sum_{k=1}^{m} \hat{\rho}_k^2 \xrightarrow{d} W, \tag{2.1}$$

for some integer $m > 0$ and $W = \sum_{k=1}^{m}(S_k/S_0)^2$, see Runde (1997, p. 207) for a proof. Lin and McLeod (2008) confirm that in the case of an infinite variance process, the χ^2 distribution is not a good approximation for the Q test in standard sample sizes available in practice. Another difficulty of this approach is that we have to find the estimate of the tail exponent κ in (2.1). However, as shown by McCulloch (1997) and Kearns and Pagan (1997), an accurate estimate of the tail index is rather difficult to obtain in finite samples. It is also worth pointing out that the importance of the higher-order sample autocorrelations in (2.1) increases in the case of infinite variance processes, see Runde (1997, p. 208) for a discussion. This theoretical finding is confirmed in Lin and McLeod (2008) based on Monte Carlo experiments. Note also that it is not quite clear, at least to our best knowledge, whether or not an automatic lag selection procedure, proposed by Escanciano and Lobato (2009) and used to determined the lag order of the Q tests, works also for infinite variance processes as well. As shown by Davis and Mikosch (2000), moment condition failure is even more peculiar for non-linear time series models. Authors show that the rate of convergence of the sample autocorrelations of some non-linear models (e.g. BL and ARCH models) is actually slower than $\sqrt{T}$, and indeed the slower the heavier the tails. This is in the complete opposite to linear ARMA models with an infinite variance. Unfortunately, the results are not valid for all non-linear time series models in general. The authors also demonstrate that the rate of convergence of the sample autocorrelations of some other non-linear models (e.g. stochastic volatility models) is similar to that derived for a linear ARMA process with an infinite variance. Rather surprisingly, the issue of the robustness of the power properties of non-linearity tests against moment condition failure has not attracted much attention in the literature despite the empirical findings that the tail index $\kappa = \sup_{k>0} \mathbb{E}(|X_t|^k) < \infty$ of economic time series lies usually between 1 and 4, see Runde (1997), Koedijk et al. (1990), and Bali (2003), among others.[1]

[1]Note that de Lima (1997) focuses on a size distortion of non-linearity tests under moment

As will be shown later on in this chapter, this moment range is not sufficient for many routinely applied non-linearity tests. For this reason, the second part of this chapter addresses this issue in detail.

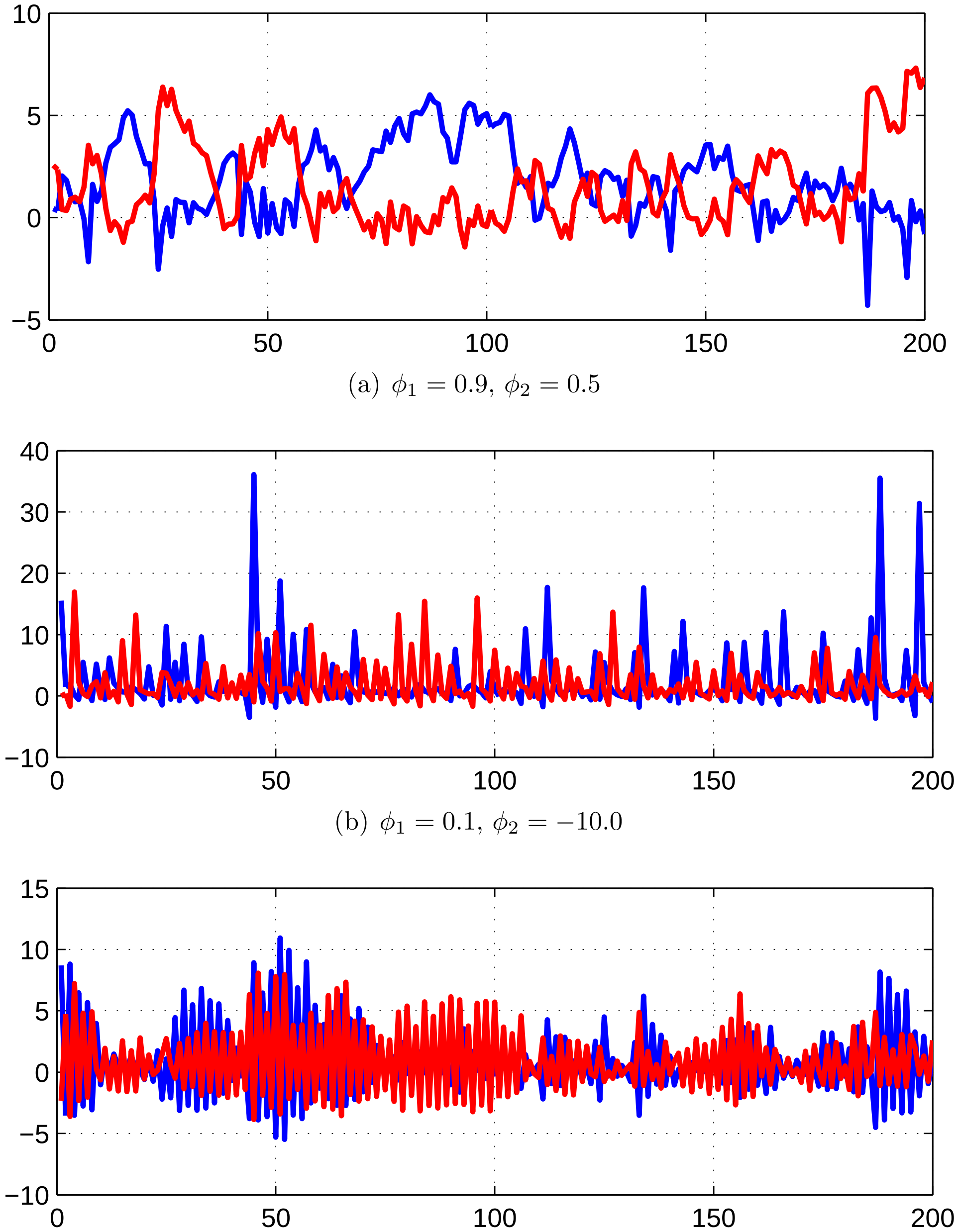
Third, another issue is that statistical properties of non-linearity tests in almost all papers are examined using Gaussian innovations only. However, there is no reason to assume that innovations of time series models are necessarily Gaussian in general. In addition, some non-linearity tests (e.g. the WHITE test) are directly derived based on an assumption of Gaussian innovations. Therefore, it is important to check the robustness of non-linearity tests against non-Gaussian innovations. Intuitively, the problem of non-Gaussian innovations is related especially to regime-switching models with endogenous switching (e.g. a TAR model), where we can expect different allocation of observations into regimes, see Figure 2.3 for an example. There is no paper focusing on this issue in the literature, at least to the best of our knowledge. Therefore, the last part of this chapter focuses on the robustness of non-linearity tests against asymmetry of innovations.

The main task of this chapter is to fill the gap in the literature and assess the robustness of selected non-linearity tests against: (i) a parameter variation of a data generating process; (ii) moment condition failure of innovations; (iii) asymmetry of innovations. This knowledge is important for making correct inference about the tests. In addition, our results may also shed the light on whether there is some objective preference for some non-linearity tests as compared to others.

The rest of the chapter is organized as follows. In Section 2.2, eight the most frequently used non-linearity tests are described. A brief description of nine non-linear time series models and Monte Carlo setup is given in Section 2.3. All Monte Carlo results can be found in Section 2.4.

condition failure only.

Figure 2.3 Different realizations of a TAR model: asymmetric innovations



(a) $\phi_1 = 0.9$, $\phi_2 = 0.5$

(b) $\phi_1 = 0.1$, $\phi_2 = -10.0$

(c) $\phi_1 = 0.4$, $\phi_2 = -2.0$

Note: The series are generated from a simple TAR(2;1,1) model: $X_t = \phi_1 X_{t-1} I(X_{t-1} > 0) + \phi_2 X_{t-1} I(X_{t-1} \leq 0) + a_t$, where $\{a_t : t \in \mathbb{Z}\}$ is a sequence of IID innovations drawn from a generalized lambda distribution: the blue line corresponds to A2(+) configuration (skewness = 1.5, kurtosis = 7.5), and the red line to -A2(+) specification (skewness = -1.5, kurtosis = 7.5), see Table 2.4 for details. Particular model parameters come from Petruccelli and Woolford (1984).

2.2 Non-linearity Tests

2.2.1 Null Hypothesis

Before we proceed to a testing procedure, we state an important assumption about a stochastic process under consideration. The assumption is of the crucial importance for setting the null hypothesis of linearity.

Assumption 1 *Let us assume $\{X_t : t \in \mathbb{Z}\}$ is a zero-mean real-valued finite-order AR(p) model given by*

$$X_t = \xi_0 + \xi_1 X_{t-1} + \cdots + \xi_p X_{t-p} + a_t = \boldsymbol{\xi}' \mathbf{X}_t + a_t, \tag{2.2}$$

where $\{a_t : t \in \mathbb{Z}\}$ is a sequence of IID(0,σ^2) model innovations such that $\mathbb{E}(|a_t|^6) < \infty$. Let $\boldsymbol{\beta} = (\xi_0, \xi_1, \ldots, \xi_p, \sigma)'$ be a $(p+2 \times 1)$ parameter vector, which is assumed to be in the interior of the parameter space

$$\mathbf{B} = \{\boldsymbol{\beta} \in \mathbb{R}^{p+2} \times \mathbb{R}_{++} : \xi(z) = 1 - \sum_{i=1}^{p} \xi_i z^i \neq 0 \quad \textit{for all} \quad |z| \leq 1.$$

$\square$

Provided that all conditions of Assumption 1 are satisfied, then a given stochastic process $\{X_t : t \in \mathbb{Z}\}$ is stationary, an appropriate model is identified and the true parameter vector $\boldsymbol{\beta}$ does not lie on the boundary of the parameter space $\mathbf{B}$. These conditions are sufficient for obtaining consistent estimates of unknown parameters and estimated residuals. It is worth noting that the null hypothesis can be easily extended also to a linear ARMA model, or an ARMA model with other explanatory variables. However, identification and filtration of ARMA models is a bit more computationally expensive for Monte Carlo experiments. For this reason, we consider only a simple AR(p) process. The lag order p is determined by an automatic lag order selection procedure discussed in Ng and Perron (2005). Note that $\{\hat{a}_t\}$ denotes a sequence of estimated residuals from (2.2) and $\hat{\sigma}^2$ is the sample variance of residuals, unless otherwise stated.

2.2.2 Non-linearity Tests

The size and power properties of eight of the most commonly used non-linearity tests are examined in this chapter. In particular, we consider the following set of tests (the moment condition required by each tests is declared in square brackets): the Brock–Dechert–Scheinkman (BDS) test [2], the Mcloed–Li Q (MLQ) test [4], the Monti Q (MQ) test [4], the Tsay (TSAY) test and Keen [KEEN] test [4], the smooth transition autoregressive (STAR) test [6], the dynamic information matrix (WHITE) test [4], and the neural network (NN) test [6]. The moment conditions are taken from de Lima (1997, p. 254).

Brock–Dechert–Scheinkman Test

Brock et al. (1996) developed a test statistic for assessing whether or not a time series is identically and independently distributed. The test statistic is based on a correlation sum defined as follows

$$C(n,\epsilon) = \frac{2}{N(N-1)} \sum\sum_{1\le i<j\le N} I_\epsilon(\hat{\mathbf{a}}_i^n, \hat{\mathbf{a}}_j^n),$$

where $N = T - n + 1$, and the indicator function with n–history is given by

$$I_\epsilon(\hat{\mathbf{a}}_i^n, \hat{\mathbf{a}}_j^n) = I(\|\hat{\mathbf{a}}_i^n - \hat{\mathbf{a}}_j^n\| < \epsilon), \quad \text{for} \quad 1 \le i < j \le N,$$

where $\|\cdot\|$ stands for the sup-norm. The BDS test statistic is then defined as

$$BDS(n,\epsilon) = \sqrt{N}\left(\frac{C(n,\epsilon) - C(1,\epsilon)^n}{\sqrt{\sigma^2(n,\epsilon)}}\right) \xrightarrow{d} NID(0,1), \tag{2.3}$$

where the standard deviation $\sigma(n,\epsilon)$ is estimated as follows

$$\sigma^2(n,\epsilon) = 4\left[K^n + 2\left(\sum_{j=1}^{n-1} K^{n-j}C^{2j}\right) + (n-1)^2C^{2n} - n^2KC^{2n-2}\right],$$

where the quantity C and K are consistently estimated by

$$C = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} I_\epsilon(\hat{a}_i, \hat{a}_j),$$

$$K = \frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} I_\epsilon(\hat{a}_i, \hat{a}_j) I_\epsilon(\hat{a}_j, \hat{a}_k).$$

As pointed out by Hsieh (1989), there are two good reasons for preferring moderate values of n: (a) The BDS test seems to be relatively insensitive to the parameter ϵ for moderate values of the n–history; (b) The standard normal distribution is relatively a good asymptotic approximation for the moderate n–history. Brock et al. (1991) obtain the maximum power of the test for $\epsilon = \hat{\sigma}$, the standard deviation of residuals from the model in (2.2).

McLeod–Li Test

McLeod and Li (1983) proposed a portmanteau test based on inspecting autocorrelations of squared residuals. The test statistic is given by

$$MLQ(m) = T(T+2)\sum_{j=1}^{m}\frac{\hat{\rho}_j^2}{T-j} \xrightarrow{d} \chi^2(m), \tag{2.4}$$

where T is the sample size, m is the lag order of the test, and $\hat{\rho}_j$ is the jth sample autocorrelation coefficient. The test statistic in this form requires the existence of the first eight moments, which might be too difficult to satisfy in practice. Therefore, some authors recommend to use autocorrelations based on absolute residuals given by

$$\hat{\rho}_j = \frac{\sum_{t=j+1}^{T}(|\hat{a}_t| - \hat{\sigma})(|\hat{a}_{t-j}| - \tilde{\sigma})}{\sum_{t=1}^{T}(|\hat{a}_t| - \tilde{\sigma})^2}, \quad j \in \{1, \ldots, m\},$$

where $\hat{a}_t$ is the estimated residual, $\tilde{\sigma}$ denotes the average absolute estimated residual. The advantage of using absolute residuals is that the test statistic requires the existence of only the first four moments. It is worth noting that Q tests are sensitive to the lag order specification m. For this reason, the Q test with the lag order m automatically selected by a procedure developed in Escanciano and Lobato (2009) is implemented.

Monti Test

Monti (1994) proposed a portmanteau test based on inspecting partial autocorrelations of the estimated residuals. It can be shown that the Monti Q test can be easily used for inspecting the partial autocorrelation structure of squared and/or absolute residuals as well. The test statistic is then given by

$$MQ(m) = T(T+2)\sum_{j=1}^{m}\frac{\hat{\pi}_j^2}{T-j} \xrightarrow{d} \chi^2(m), \tag{2.5}$$

where T is the sample size, m is the lag order of the test, and $\hat{\pi}_j$ is the jth sample partial autocorrelation coefficient estimated from the Yule-Walker equations using the above estimated autocorrelations $\hat{\rho}_j$ for $j \in \{1,\ldots,m\}$. As in the case of the MLQ test, the MQ test is sensitive on the lag order specification m as well. For this reason, the test test with the lag order m automatically selected by a procedure developed in Escanciano and Lobato (2009) is implemented.

Tsay and Keen Tests

In order to improve the power of the non-linearity tests developed by Keenan (1985) and Ramsey (1969), Tsay (1986) proposed to use a different set of explanatory variables for the test. The test is based on running an auxiliary equation in the form

$$\hat{a}_t = \boldsymbol{\beta}'\mathbf{Z}_t + u_t,$$

where $\mathbf{Z}_t = \text{vech}(\mathbf{X}_t\mathbf{X}_t')$ is a vector of predetermined variables, their squares and cross products, and vech$(\cdot)$ denotes a half-stacking operator. The LM version of the test statistic is defined as

$$TSAY(p) = TR^2 \xrightarrow{d} \chi^2(p(p+1)/2), \tag{2.6}$$

where T denotes the sample size, and R^2 is the coefficient of determination from an auxiliary model. In a special case when $p = 1$, the TSAY is coincides with the KEEN test proposed by Keenan (1985).

STAR Test

A STAR test is a test used for testing linearity against smooth transition autoregressive models. The model can be written in the form as follows

$$X_t = \boldsymbol{\phi}_1'\mathbf{X}_t + \boldsymbol{\phi}_2'\mathbf{X}_t G(\mathbf{X}_t'\boldsymbol{\theta}, \gamma) + a_t,$$

where $G(\cdot)$ is the so called transition function, $\mathbf{X}_t = (1, X_{t-1}, \ldots, X_{t-p})$ is a $(p+1\times 1)$ vector of predetermined variables, $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, and $\boldsymbol{\theta}$ are $(p+1\times 1)$ vectors of unknown parameters, γ is a smoothing constant. In order to get around the identification problem, see Hansen (1996) for details, Luukkonen et al. (1988) proposed a testing procedure based on an approximating the transition function $G(\cdot)$ by the third-order Taylor expansion. The estimated auxiliary equation is given by

$$\hat{a}_t = \boldsymbol{\alpha}'\mathbf{X}_t + \mathbf{x}_t'\mathbf{A}\mathbf{x}_t + \boldsymbol{\beta}'\mathbf{z}_t + u_t,$$

where $\mathbf{X}_t = (1, X_{t-1}, \ldots, X_{t-p})$ is a $(p+1\times 1)$ vector of predetermined variables, $\mathbf{x}_t = (X_{t-1}, \ldots, X_{t-p})$ is a $(p\times 1)$ vector of predetermined variables, $\mathbf{z}_t = (X_{t-1}^3, \ldots, X_{t-p}^3)$ is a $(p\times 1)$ vector of powers of predetermined variables, $\boldsymbol{\alpha}$ is a $(p+1\times 1)$ and $\boldsymbol{\beta}$ is $(p\times 1)$ vector of real parameters, $\mathbf{A}$ is a $(p\times p)$ upper/lower triangular matrix. The main advantage of the Taylor approximation of the transition function $G(\cdot)$ is that we can apply directly the conventional LM-based test with asymptotic critical values. The LM version of the test statistic is defined as

$$STAR(p) = TR^2 \xrightarrow{d} \chi^2(p(p+1)/2 + p), \tag{2.7}$$

where T denotes the sample size, and R^2 is the coefficient of determination from the auxiliary model.

White Dynamic Information Matrix Test

White (1987) proposed a specification test for time series models. The test is based on the well known fact that for a correctly specified model, a score vector is serially uncorrelated. Assuming Gaussian innovations in (2.2), the score vector $\mathbf{s}_t$ for an AR(p) model can be written as follows

$$\mathbf{s}_t = \frac{\partial l_t(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = \frac{1}{\sigma}(u_t\mathbf{X}_t', u_t^2 - 1)',$$

where $l_t(\cdot)$ denotes the log-likelihood contribution, $\boldsymbol{\omega}$ is $(p+2\times 1)$ complete vector of unknown parameters in the model: $\boldsymbol{\omega} = (\boldsymbol{\xi}', \sigma)'$ in our case, and $u_t = a_t/\sigma$ is a standardized error term. Provided that a model is correctly specified, then it holds that $\mathbb{E}(\mathbf{s}_t) = 0$ and $\mathbb{E}(\mathbf{s}_t\mathbf{s}_{t-1}') = 0$. The test statistic is based on inspecting the relationship between $\hat{u}_t$ and $\mathbf{z}_t = \mathbf{S}\text{vec}(\hat{\mathbf{s}}_t\hat{\mathbf{s}}'_{t-1})/\hat{u}_t$, where $\mathbf{S}$ is the selection matrix, and vec$(\cdot)$ is a stacking operator converting a matrix into a vector, and $\hat{u}_t = \hat{a}_t/\hat{\sigma}$ is the standardized estimated residual term. The test is based on running the following auxiliary equation

$$\hat{u}_t = \boldsymbol{\beta}'\mathbf{X}_t + \boldsymbol{\gamma}'\mathbf{z}_t + e_t,$$

where $\mathbf{X}_t = (1, X_{t-1}, \dots, X_{t-p})'$ is a $(p+1 \times 1)$ vector of predetermined variables, and $\mathbf{z}_t$ is a $(q \times 1)$ vector of selected cross products and powers of the estimated score vector elements $\hat{\mathbf{s}}_t$. The LM test statistic is given by

$$WHITE(q) = TR^2 \xrightarrow{d} \chi^2(q), \tag{2.8}$$

where T denotes the sample size, and R^2 is the coefficient of determination from the auxiliary model. Note that some authors use ad-hoc adjustment of the selection matrix $\mathbf{S}$, see Lee et al. (1993, p. 279). We do not follow this approach here and consider all the elements from the score vector $\hat{\mathbf{s}}_t$.

Neural Network Test

White (1989) proposed a neural network test for testing neglected non-linearity in time series. The test is motivated by the fact that under the null hypothesis of linearity, residuals from the model should be uncorrelated with any $\mathcal{F}_{t-1}$–measurable function: $\mathbb{E}(a_t\boldsymbol{\psi}(\mathcal{F}_{t-1}))$, where $\mathcal{F}_t$ is a Borel-sigma field generated by observation of X up to and including time t. Lee et al. (1993) approximate a vector of squashing functions $\boldsymbol{\psi}(\mathcal{F}_{t-1})$ by a neural network method based on logistic cumulative distribution functions $\boldsymbol{\psi}_t = (\psi_1(\boldsymbol{\gamma}'_1\mathbf{X}_t), \dots, \psi_k(\boldsymbol{\gamma}'_k\mathbf{X}_t))'$, where the individual squashing functions ψ_j are defined as follows

$$\psi_j = \frac{1}{1+\exp(\boldsymbol{\gamma}'_j\mathbf{X}_t)}, \quad \text{for} \quad j = 1, \dots, k.$$

In order to eliminate the identification problem, the authors recommend to use randomly generated real-valued parameter vectors $\boldsymbol{\gamma}_j$, for $j = 1, \dots, k$, from a uniform distribution with support $[-2, 2]$. For computational reasons, the authors also use only the first $k^* < k$ principal components (and exclude the first one) in order to avoid a problem of collinearity in the model. The number of principal component is set to $k^* = p$. The test is based on running an auxiliary regression

$$\hat{a}_t = \boldsymbol{\beta}'\mathbf{X}_t + \boldsymbol{\delta}'\boldsymbol{\psi}_t + u_t,$$

where the vector $\mathbf{X}_t = (1, X_{t-1}, \ldots, X_{t-p})'$ is a $(p + 1 \times 1)$ vector of predetermined variables, and $\boldsymbol{\psi}_t$ is a $(p \times 1)$ vector. The LM test statistic is given by

$$NN(p) = TR^2 \xrightarrow{d} \chi^2(p), \tag{2.9}$$

where T denotes the sample size, and R^2 is the coefficient of determination from the auxiliary model.

Note that we do not consider the NN test modified for testing heteroscedasticity since the test statistic relies on critical values obtained from a bootstrap method, see Blake and Kapetanios (2003). This approach would be very computationally expensive in our case though.

2.3 Time Series Models and Monte Carlo Setup

2.3.1 Time Series Models

The statistical properties of the selected non-linearity tests are examined using: (i) A simple linear autoregressive (AR) model; (ii) The following non-linear time series models: a threshold autoregressive (TAR) model, an exponential autoregressive (EXPAR) model, a mixture autoregressive (MAR) model, a Markov switching autoregressive (MSAR) model, a generalized autoregressive conditional heteroscedasticity (GARCH) model, a bilinear (BL) model, a random coefficient autoregressive (RCAR) model, a non-linear moving average (NLMA) model, and finally, a threshold moving average (TMA) model. Although the list of non-linear time series models is definitely not exhaustive, we are convinced that it includes some of the most commonly used non-linear time series models. The models are summarized in Table 2.1.

2.3.2 Parameters and Innovations

The robustness of the power of the non-linearity tests is examined using different configurations of the key model parameters. In particular, we consider the following number of parameter configurations for individual time series models: $K = 8$ for an AR model, $K = 24$ for TAR, EXPAR, MAR, MSAR, and TMA models, $K = 12$ for GARCH and NLMA models, $K = 18$ for BL and RCAR models, see Table 2.2 for particular parameter configurations. Gaussian innovations are considered for all data generating processes when inspecting the robustness of the power properties of the tests against parameter configurations. Note that parameters of all data generating processes are designed in such a way to satisfy strict stationarity, 6th-moment stationarity and/or invertibility conditions, if necessary, provided that model innovations are from a Gaussian distribution. The only exceptions are S3, S4, S5, S6, A3 specifications of model innovations, for which the 6th-moment stationarity is not satisfied. However, it is worth noting that imposing the moment restriction restricts the parameter space of some non-linear models significantly (e.g. ARCH models are a nice example). To make this point clear, Figure 2.4 depicts the parameter regions ensuring stationarity with NID(0,1) model innovations, denoted as "Strict Stationarity", and the parameter regions satisfying 6th-moment stationarity and/or invertibility conditions, denoted as "Monte Carlo".

Afterwards, we examine the robustness of the selected non-linearity tests against moment condition failure and asymmetry of model innovations. The robustness against moment condition failure is examined using a Student t distribution with different degrees of freedom controlling for the existence of moments. In particular, six different specifications from $t(3)$ to $t(8)$ are considered, see Table 2.3 for details. The robustness against asymmetry of innovations is examined using a generalized lambda distribution (GLD), see Randles et al. (1980). This family provides a wide range of distributions that are easily generated since they are defined in terms of the inverses of the cumulative distribution functions: $F^{-1}(\nu) = \lambda_1 + [\nu^{\lambda_3} - (1-\nu)^{\lambda_4}]/\lambda_2$, for $0 \leq \nu \leq 1$. In particular, we consider six specifications of asymmetric distributions, which differ in the magnitude of asymmetry, see Table 2.4. All generated innovations are normalized to have zero mean and unit variance.

2.3.3 Monte Carlo Setup

Originally, $T+100$ observations in each experiment are generated, but the first 100 of them are discarded in order to eliminate the effect of initial observations. The number of replications of all experiments is set to $R = 1000$. The sample size is $T \in \{200, 500, 1000\}$. In all experiments, the generated series is filtered by an AR(p) model where the lag order p is selected by the Bayesian information criterion (BIC) developed by Schwarz (1978). Following the arguments in Ng and Perron (2005), a modified version of the criterion is used. They show, based on extensive Monte Carlo experiments, that the best method to give the correct lag order is that with the fixed efficient sample size. Therefore, our criterion is defined as follows

$$BIC_l = \log(\hat{\sigma}_l^2) + \frac{l\log(N)}{N},$$
$$\hat{\sigma}_l^2 = \frac{1}{N}\sum_{t=L+1}^{T} \hat{a}_{lt}^2,$$

where $l \in \{1, \ldots, L\}$, and $N = T - L$ is the efficient sample size, where T is the actual sample size and L is the maximum lag order constrained by $L = [8(T/100)^{0.25}]$. Finally, the lag order p is determined by $\hat{p} = \min_{l\in\{1,\ldots,L\}}(BIC_l)$.

We also report two portmanteau tests with the lag order m determined by an optimal selection procedure developed by Escanciano and Lobato (2009).[2] The lag order of the MLQ test is selected by maximizing the following objective function

$$Q_l^* = Q_l - q_l,$$
$$q_l = \begin{cases} p\log(N) & \text{if} \quad \max_{j\in\{1,\ldots,L\}} |\hat{\rho}_j| \leq \sqrt{c\log(N)/N}, \\ \\ 2p & \text{if} \quad \max_{j\in\{1,\ldots,L\}} |\hat{\rho}_j| > \sqrt{c\log(N)/N}, \end{cases}$$

[2] Recall that a given procedure is proposed for a realization of some stochastic process and not a filtered one. Unreported simulations show, however, that the procedure can be adopted for filtered processes as well.

where Q_l is a value of the Q tests, q_l is a penalization function, $c = 2.4$ is a correction constant used in Escanciano and Lobato (2009) based on Monte Carlo experiments. Finally, the lag order m for the Q tests is determined by: $\hat{m} = \max_{l \in \{1,\ldots,L\}}(Q_l^*)$. Note that the lag order of the MQ test is determined using the same procedure but the estimated autocorrelations $\hat{\rho}_j$ are replaced by the estimated partial autocorrelations $\hat{\pi}_j$.

Table 2.1 List of non-linear models

M1: AR model:

$$Y_t = c + \phi Y_{t-1} + \sigma a_t$$

M2: TAR model:

$$Y_t = (c_1 + \phi_1 Y_{t-1} + \sigma_1 a_t) I(Y_{t-1} \leq 0) + (c_2 + \phi_2 Y_{t-1} + \sigma_2 a_t) I(Y_{t-1} > 0)$$

M3: EXPAR model:

$$Y_t = c + (\phi_1 + (\phi_2 - \phi_1) \exp(-Y_{t-1}^2)) Y_{t-1} + \sigma a_t$$

M4: MAR model:

$$Y_t = (c_1 + \phi_1 Y_{t-1} + \sigma_1 a_t) I(S_t = 1) + (c_2 + \phi_2 Y_{t-1} + \sigma_2 a_t) I(S_t = 2)$$

M5: MSAR model:

$$Y_t = (c_1 + \phi_1 Y_{t-1} + \sigma_1 a_t) I(S_t = 1) + (c_2 + \phi_2 Y_{t-1} + \sigma_2 a_t) I(S_t = 2)$$

M6: GARCH model:

$$\begin{aligned} &Y_t = c + \phi Y_{t-1} + \epsilon_t, \quad \epsilon_t = a_t \sqrt{h_t}, \\ &h_t = \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1} \end{aligned}$$

M7: RCAR model:

$$Y_t = c + (\phi + \psi u_t) Y_{t-1} + a_t$$

M8: TMA model:

$$Y_t = (c_1 + \phi_1 a_{t-1}) I(Y_{t-1} \leq 0) + (c_2 + \phi_2 a_{t-1}) I(Y_{t-1} > 0) + \sigma a_t$$

M9: BL model:

$$Y_t = c + \phi Y_{t-1} + \psi Y_{t-1} a_{t-1} + \sigma a_t$$

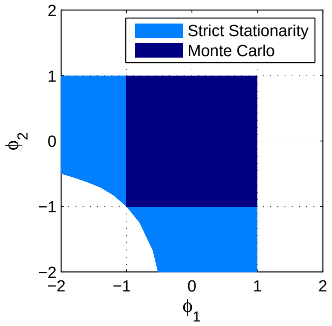
M10: NLMA model:

$$Y_t = c + \phi a_{t-1} + \psi a_t a_{t-1} + \sigma a_t$$

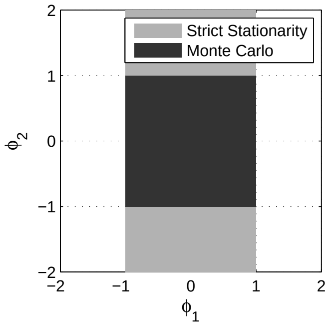
Table 2.2 Parameters of non-linear models

model	**parameters**
AR, MA	$c = 1$ $\sigma^2 = 1$ $\phi \in \{-0.8, -0.6, -0.4, -0.2, 0.2, 0.4, 0.6, 0.8\}$ $\theta \in \{-0.8, -0.6, -0.4, -0.2, 0.2, 0.4, 0.6, 0.8\}$
TAR, EXPAR	$c_1 = -0.25$, $c_2 = 0.25$ (for TAR only) $\sigma_1^2 = 3$, $\sigma_2^2 = 1$ (for TAR only) $c = 1$ (for EXPAR only) $\sigma^2 = 1$ (for EXPAR only) $(\phi_1, \phi_2) \in \left\{\begin{array}{ccccc} (-0.8,-0.8) & (-0.8,-0.5) & (-0.8,-0.2) & (-0.8,0.2) & (-0.8,0.5) \\ (-0.8,0.8) & (-0.5,-0.8) & (-0.5,-0.5) & (-0.5,0.5) & (-0.5,0.8) \\ (-0.2,-0.8) & (-0.2,0.8) & (0.2,-0.8) & (0.2,0.8) & (0.5,-0.8) \\ (0.5,-0.5) & (0.5,0.5) & (0.5,0.8) & (0.8,-0.8) & (0.8,-0.5) \\ (0.8,-0.2) & (0.8,0.2) & (0.8,0.5) & (0.8,0.8) & \end{array}\right\}$
MAR, MSAR	$c_1 = -0.25$, $c_2 = 0.25$ $\sigma_1^2 = 3$, $\sigma_2^2 = 1$ $p_{11} = 0.9$, $p_{22} = 0.7$ (for MSAR only) $\pi = 0.5$ (for MAR only) $(\phi_1, \phi_2) \in \left\{\begin{array}{ccccc} (-0.8,-0.8) & (-0.8,-0.5) & (-0.8,-0.2) & (-0.8,0.2) & (-0.8,0.5) \\ (-0.8,0.8) & (-0.5,-0.8) & (-0.5,-0.5) & (-0.5,0.5) & (-0.5,0.8) \\ (-0.2,-0.8) & (-0.2,0.8) & (0.2,-0.8) & (0.2,0.8) & (0.5,-0.8) \\ (0.5,-0.5) & (0.5,0.5) & (0.5,0.8) & (0.8,-0.8) & (0.8,-0.5) \\ (0.8,-0.2) & (0.8,0.2) & (0.8,0.5) & (0.8,0.8) & \end{array}\right\}$
GARCH	$c = 1$ $\phi = 0.5$ $\sigma^2 = 1$ $(\alpha, \beta) \in \left\{\begin{array}{ccccc} (0.05,0.2) & (0.05,0.3) & (0.05,0.4) & (0.05,0.5) & (0.05,0.6) \\ (0.05,0.7) & (0.05,0.8) & (0.05,0.9) & (0.10,0.2) & (0.10,0.3) \\ (0.10,0.4) & (0.10,0.5) & & & \end{array}\right\}$
TMA	$c_1 = -0.25$, $c_2 = 0.25$ $\sigma^2 = 1$ $(\phi_1, \phi_2) \in \left\{\begin{array}{ccccc} (-0.8,-0.8) & (-0.8,-0.5) & (-0.8,-0.2) & (-0.8,0.2) & (-0.8,0.5) \\ (-0.8,0.8) & (-0.5,-0.8) & (-0.5,-0.5) & (-0.5,0.5) & (-0.5,0.8) \\ (-0.2,-0.8) & (-0.2,0.8) & (0.2,-0.8) & (0.2,0.8) & (0.5,-0.8) \\ (0.5,-0.5) & (0.5,0.5) & (0.5,0.8) & (0.8,-0.8) & (0.8,-0.5) \\ (0.8,-0.2) & (0.8,0.2) & (0.8,0.5) & (0.8,0.8) & \end{array}\right\}$
BL, RCAR	$c = 1$ $\sigma^2 = 1$ $(\phi, \psi) \in \left\{\begin{array}{ccccc} (-0.8,-0.2) & (-0.6,-0.2) & (-0.4,-0.2) & (-0.2,-0.2) & (-0.2,0.2) \\ (-0.4,0.2) & (-0.6,0.2) & (-0.6,0.4) & (-0.8,0.2) & (0.2,-0.2) \\ (0.2,0.2) & (0.4,-0.2) & (0.4,0.2) & (0.6,-0.2) & (0.6,-0.4) \\ (0.6,0.2) & (0.8,-0.2) & (0.8,0.2) & & \end{array}\right\}$
NLMA	$c = 1$ $\sigma^2 = 4$ $(\phi, \psi) \in \left\{\begin{array}{ccccc} (-0.20,0.20) & (-0.20,0.40) & (-0.40,0.20) & (-0.40,0.40) & (-0.60,0.20) \\ (-0.60,0.40) & (0.20,0.20) & (0.20,0.40) & (0.40,0.20) & (0.40,0.40) \\ (0.60,0.20) & (0.60,0.40) & & & \end{array}\right\}$

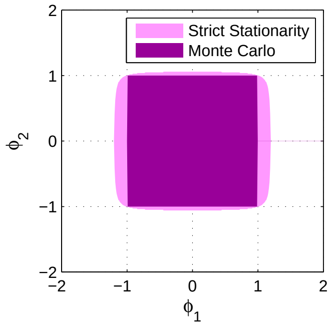
Figure 2.4 Parameter configurations of time series models



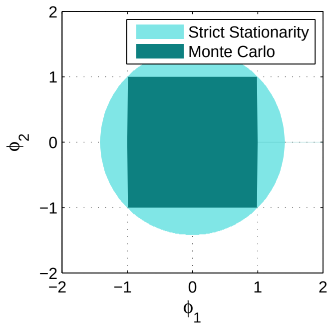
(a) TAR(2;1,1)



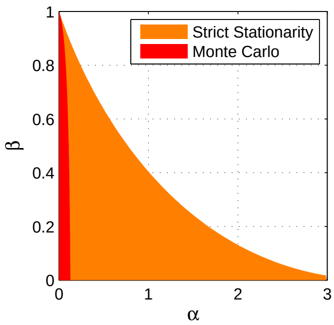
(b) EXPAR(1)



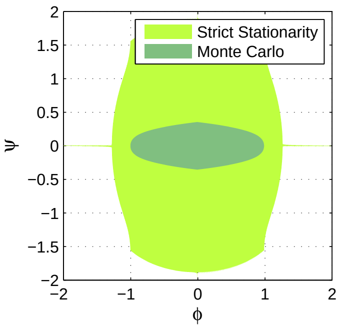
(c) MSAR(2;1,1)



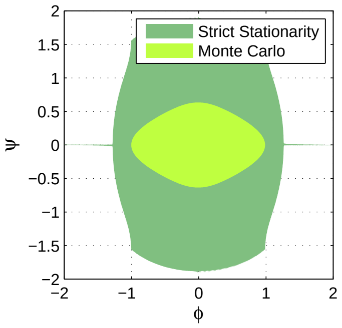
(d) MAR(2;1,1)



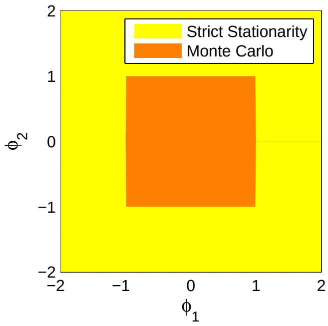
(e) GARCH(1,1)



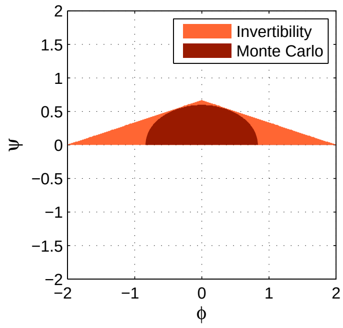
(f) BL(1,0,1,1)



(g) RCAR(1)



(h) TMA(2;1,1)



(i) NLMA(1)

Note: Strict Stationarity regions are calculated based on an assumption that $a \sim$ NID(0,1), if necessary, whereas Monte Carlo regions are calculated based on the intersection of 6th-moment stationarity and/or invertibility conditions for the following set of distributions of model innovations: NID(0,1), S7, S8, A1(+), A1(−), A2(+), A2(−). All other distributions (i.e. S3, S4, S5, S6, A3(+), A3(−)) are not considered since they do not implicitly satisfy the existence of the 6th-moment condition.

Table 2.3 Parameters of a Student t distribution

	dof	skewness	kurtosis	moment[a]
S3	3	–	–	2
S4	4	0.0	–	3
S5	5	0.0	9.0	4
S6	6	0.0	6.0	5
S7	7	0.0	5.0	6
S8	8	0.0	4.5	7

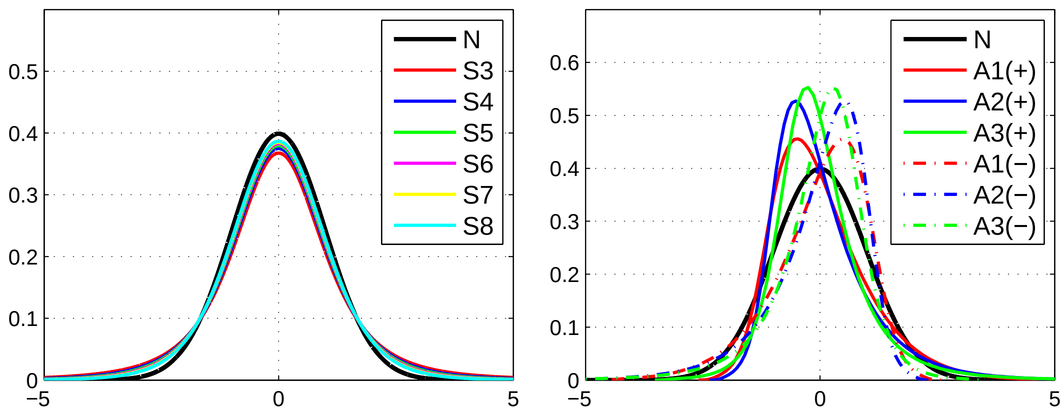
[a] The highest finite moment of a random variable drawn from a given distribution.

Table 2.4 Parameters of a generalized lambda distribution

	λ_1	λ_2	λ_3	λ_4	skewness	kurtosis	moment[a]
A1(+)	0.00000	0.04306	-0.02521	-0.09403	0.9	4.2	10
A1(−)	0.00000	-0.04306	0.02521	0.09403	-0.9	4.2	10
A2(+)	0.00000	-1.00000	-0.00750	-0.03000	1.5	7.5	33
A2(−)	0.00000	1.00000	-0.00750	-0.03000	-1.5	7.5	33
A3(+)	0.00000	-1.00000	-0.10090	-0.18020	2.0	21.1	5
A3(−)	0.00000	1.00000	-0.10090	-0.18020	-2.0	21.1	5

[a] The highest finite moment of a random variable drawn from a given distribution.

Figure 2.5 Distributions of model innovations



(a) Symmetric densities (b) Asymmetric densities

2.4 Monte Carlo Results

2.4.1 Introduction

For a given test, a given data generating process (DGP), and a given distribution of innovations, the average rejection frequency is calculated over all parameter configurations as follows

$$avg_j = \frac{1}{K}\sum_{i=1}^{K}\mathcal{P}_{i,j}, \quad j \in \{1,\dots,13\}, \tag{2.10}$$

where $\mathcal{P}_{i,j}$ is the rejection frequency of the test for a given parameter configuration $i \in \{1,\dots,K\}$ and distribution of innovations $j \in \{1,\dots,13\}$. The following set of thirteen distributions of innovations is consider in this chapter: a Gaussian distribution indexed as $j = 1$, six Student t distributions indexed from $j = 2$ to $j = 7$, and six asymmetric distributions indexed from $j = 8$ to $j = 13$. Note that we adopt a convention that $j = 1$ represents a standard normal distribution unless otherwise stated. The number of parameter configurations vary across DGPs: $K = 24$ for a TAR, EXPAR, MAR, MSAR, and TMA model, $K = 12$ for a GARCH and NLMA model, $K = 18$ for a BL model and RCAR model, see Table 2.2. The rejection frequency $\mathcal{P}_{i,j}$ is given by

$$\mathcal{P}_{i,j} = \frac{1}{R}\sum_{r=1}^{R} I(\hat{\alpha}_r \leq \alpha), \quad i \in \{1,\dots,K\}, j \in \{1,\dots,13\},$$

where R denotes the number of repetitions, $\alpha = 0.05$ is the nominal significance level, and $\hat{\alpha}$ is the estimated p-value of the test.

Variability of the size and power of the tests against the parameter configuration of DGPs is assessed using a modified coefficient of variation. For a given test, a given DGP, the following coefficient of variation is used

$$cv(N) = \frac{\max_i(\mathcal{P}_{i,1}) - \min_i(\mathcal{P}_{i,1})}{avg_1}, \tag{2.11}$$

where $\mathcal{P}_{i,1}$ denotes the rejection frequency of a given ith parameter configuration based on Gaussian innovations $j = 1$, avg_1 represents the average rejection frequency calculated over all parameter K parameter configurations of a given DGP, see (2.10).

Variability of the size and power of the tests against moment condition failure and asymmetry of innovations is assessed using the coefficient of variation as well. For a given test, a given DGP, the coefficient of variation is defined as follows

$$cv = \frac{\max_j(avg_j) - \min_j(avg_j)}{avg_1}, \tag{2.12}$$

where avg_1 represents the average rejection frequency calculated based on Gaussian innovations over all parameter K parameter configurations of a given DGP, whereas avg_j denotes the average rejection frequency calculated based on jth distribution of innovations over all parameter K parameter configurations of the DGP. Note that $j \in \{1, 2, \ldots, 7\}$ when assessing the effect of moment condition failure using Student t distributions, whereas $j \in \{1, 8, \ldots, 13\}$ when assessing the effect of asymmetry using GDL distributions. The coefficient of variation is denoted as $cv(S)$ and $cv(A)$ for symmetric and asymmetric innovations in tables below.

2.4.2 Monte Carlo Results: Parameters

Size: The size results of the tests are presented in Table 2.7. The results reveal that all non-linearity tests considered here have the size close to the nominal level $\alpha = 0.05$. The BDS and WHITE tests are the only two tests suffering from a size distortion: the BDS test is slightly oversized, whereas the WHITE test is undersized.[3] The size results of the tests improve as the sample size T increases.

Power: The power results of the non-linear tests are presented in Table 2.8. The tests can be split into two groups according to their power properties. The first group consists of the BDS, MLQ and MQ tests, which all have a very good power for MAR, MSAR, GARCH, and BL models. The second group contains the TSAY, STAR, WHITE, and NN tests, which have a very good power for TAR, EXPAR,

[3]Note that the fact that the BDS is slightly biased in small samples is well known, see Hsieh (1989) for a discussion. The size results of the WHITE test in our paper are slightly more undersized compared to those reported by Lee et al. (1993). From this we can conclude that the WHITE test is sensitive on the specification of the selection matrix $\mathbf{S}$.

TMA, and BL models.[4] It is interesting to mention that the first group of tests (i.e. the BDS and Q tests) exhibits a very good power for regime-switching models with latent exogenous switching (i.e. MAR or MSAR models), whereas the second group of tests (TSAY, STAR, WHITE, and NN tests) are powerful especially for regime-switching models with endogenous switching (i.e TAR or TMA model). It is also worth noting that a BL model is easily recognized by all the non-linearity tests, whereas all the tests have a very low power against a RCAR model, and no one from the tests exhibits power against a NLMA model.

Since both groups of tests exhibit a power for rather different types of non-linear time series models, and since the properties of the tests are homogenous in each group, a reasonable testing strategy seems to be to apply just one test from each group (e.g. the BDS and NN tests).

Power variation: Although Monte Carlo results confirm that the selected non-linearity tests can be useful, the average rejection frequencies reported in the tables do not tell us much about the robustness of the tests against parameter configurations of DGPs. For this reason, a modified coefficient of variation, calculated according to (2.11), is reported in Table 2.8 as well. The results are depicted in a graphical form in Figure 2.6(a). The figure shows the relationship between the average rejection frequency (x-axis) and the appropriate coefficient of variation (y-axis). One can logically expect the inverted U shape between the average rejection frequency and its variability: the higher the average frequency, the lower the coefficient of variation, and vice versa. The results suggest the following: (i) It is interesting to point out that the relation ship between the average rejection and its variability seems to be linear, provided that the rejection exceeds the cutoff 0.4; (ii) However, even tests with a very high average rejection frequency exceeding 0.8 can suffer from a relatively high coefficient of variation close to 1, see Figure 2.6(a). Note that this magnitude of variability (i.e. $cv(N) = 1$) means that the power of

[4]The only exception is the TSAY test, which does has a very low power against an EXPAR model.

the test drops down close to the significance level for some parameter configurations even if the average rejection frequency is high (e.g. $avg = 0.8$); (iii) Unfortunately, no clear conclusion can be made for the tests with the average rejection frequency less than 0.4.

In order to make this point clear, the individual Monte Carlo results are presented in the form of graphical images as well. Each point depicted in the graphical images represents the estimated p-value of a given non-linearity test for a given parameter configuration (x-axis) and a given Monte Carlo replication (y-axis). Moreover, for better understanding of the sensitivity results, a color range (from black to white) is used to indicate the different magnitude of the statistical significance of the non-linearity tests, see Figure 2.7 and 2.8 for the case of Gaussian innovations and the sample size $T = 1000$. For example, from the results about a BL model, it can be concluded that all estimated p-values of the TSAY, STAR, WHITE, and NN tests are less then the significance level $\alpha = 0.05$ and the results are not sensitive to any parameter configuration of a BL model (see the black color in all images). The completely opposite results (but still good) are obtained for a NLMA model, where almost all the estimated p-values of the TSAY, STAR, WHITE, and NN tests are much larger than the significance level $\alpha = 0.05$ but the results are not sensitive to the parameter configuration (see the orange color of all images). These two outcomes, although completely opposite, are in line with the inverted U shape form since they give us clear and reliable information to make correct inference. In contrast, very problematic results are obtained using, for example, the BDS test for a TMA model, where the results are extremely sensitive to the parameter configuration of a TMA model (see the annealing color of the image).

2.4.3 Monte Carlo Results: Moments and Asymmetry

The power results of the tests based on moment condition failure and asymmetry of innovations are presented in Tables 2.9 – 2.24. For better understanding of the Monte Carlo results in this section, the highest existing moment of model innovations is indicated by a color legend in tables bellow: a dark grey legend indicates moment condition failure for a given test, a light dark legend indicates that moment condition is exactly satisfied for a given test, whereas no-color legend indicates that the lowest existing moment is even higher than a given test statistic requires.

Moment condition failure: Our results suggest that the power variation of the non-linearity tests is extremely model rather than test dependent. For example, the BDS test, although it requires the existence of only the second moment of the estimated residuals, suffers from high variability of the average rejection frequency even if the second moment is satisfied: the average rejection frequency of the BDS based on Gaussian innovations for a BL model in the small sample $T = 200$ is approximately 0.47, whereas the same average rejection frequency is 0.83 for Student $t(3)$ innovations, see the top panel in Table 2.9. Very similar results can be found for other non-linearity tests and time series models. However, as in the previous case, it can be concluded that very powerful tests are usually less sensitive to moment condition failure. For example, the NN test, requiring the existence of the first six moments, exhibits almost no power variation (i.e. the coefficient of variation is close to zero) for some specific models such as TAR, EXPAR, TMA, MSAR, and BL models, but extremely large variation for MAR, GARCH or RCA models, see the last column in Table 2.16 for details. Similar results can be also observed for the Q tests. For instance, the MLQ test exhibit very small power variation (i.e. the coefficients of variation are close to 0) for MAR, MSAR, and GARCH models, but large variation for TMA and NLMA models, see the last column in Table 2.10.

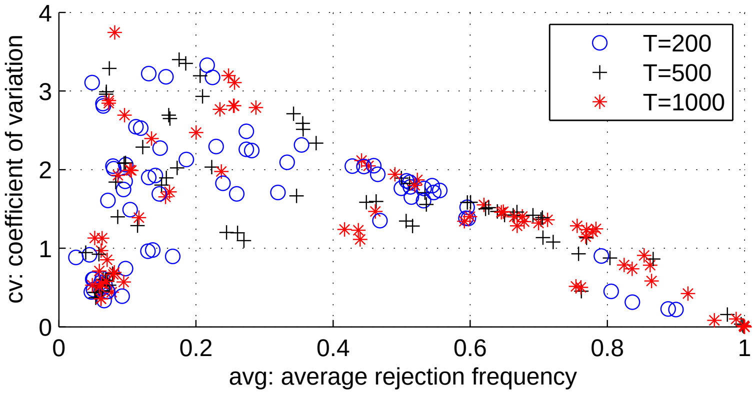
The results are depicted in a graphical form in Figure 2.6(b). The figure shows the relationship between the average rejection frequency of the tests based on Gaussian innovations ("avg_1") and the coefficients of variation under moment condition failure ("$cv(S)$"). The figure clearly reveals that the relationship between the power and its variability due to moment condition failure is significantly non-linear. Nevertheless, the power variation is significantly reduced, and can be considered as a minor problem, provided that the average rejection frequency exceeds 0.6. On the contrary, the power variation can take extremely high values when the power of the test is relatively small, say less than 0.2.

In order to make correct inference about the robustness of the non-linearity tests against different specifications of distributions of model innovations, we formally test a set of hypothesis. The null hypotheses are as follows: (i) the average rejection frequency from a particular non-Gaussian distribution (i.e. avg_j for $j \in \{2, \ldots, 13\}$), equals to the Gaussian counterpart (i.e. avg_1) for each time series model and non-linearity test considered in the paper: $H_0 : avg_j = avg_1$ against $H_1 : avg_j \neq avg_1$; (ii) the average rejection frequency from a particular non-Gaussian distribution (i.e. avg_j for $j \in \{2, \ldots, 13\}$), significantly exceeds the Gaussian counterpart (i.e. avg_1) for each time series model and non-linearity test considered in the paper: $H_0 : avg_j > avg_1$ against $H_1 : avg_j \leq avg_1$; (iii) the average rejection frequency from a particular non-Gaussian distribution (i.e. avg_j for $j \in \{2, \ldots, 13\}$), is significantly less than the Gaussian counterpart (i.e. avg_1) for each time series model and non-linearity test considered in the paper: $H_0 : avg_j < avg_1$ against $H_1 : avg_j \geq avg_1$. Since the hypothesis is about two average rejection frequencies, it means sample averages of Bernoulli random variables, we can use a Normal approximation to a Binomial distribution and apply a simple t-test for testing the null hypothesis, see Casella and Berger (2001, p. 105) for details.[5] We consider a significance level of the test $\alpha = 0.05$. Since we consider nine non-linear time series models, see Table 2.2, and six specifications from a Student t distribution, see Table 2.3, and also six specifications from a generalized lambda distribution, see Table 2.4, we obtain two sets of 54 results about the null hypothesis for each non-linear test. Therefore, the rejection frequencies of the null hypothesis for each non-linearity test is reported in Table 2.5.

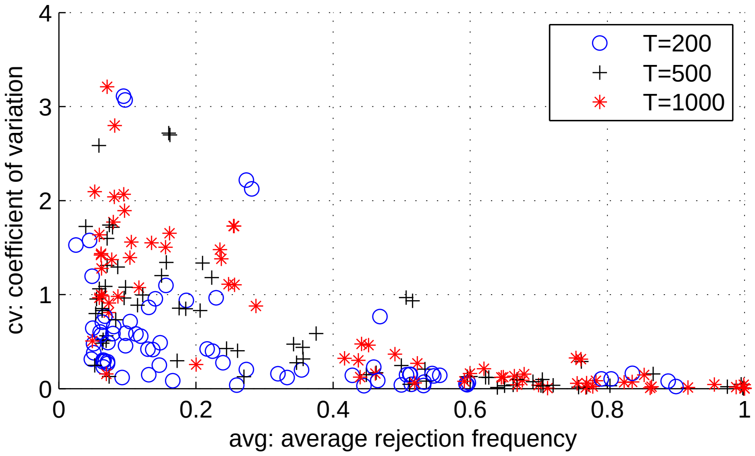
The results indicate that there seems not to be a clear cutoff between the moment requirement of a given test and the robustness of its power against moment condition failure. For example, the null hypothesis about no change (i.e. $H_0 : avg_j = avg_1$) is

[5] Moreover, since the number of replications of each experiment is set to $R = 1000$, we do not have to consider any “continuity” correction of a Normal approximation.

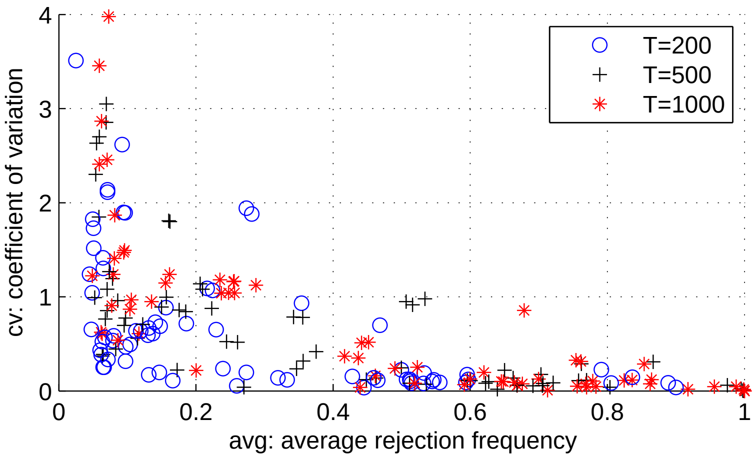
Figure 2.6 Power variation of the tests



(a) Gaussian innovations



(b) Moment failure of innovations



(c) Asymmetry of innovations

rejected in 33 % for the NN test, which requires the existence of the sixth moment, but in 57 % for the BDS test, which requires the existence of only the second moment. Some other tests, especially the Q tests, do suffer from even higher rejection of the null. For example, the null hypothesis about no change (i.e. $H_0 : avg_j = avg_1$) is rejected in 78 % for the MLQ test, which requires the existence of the forth moment. Another important finding is that that the power of the tests under moment condition failure of model innovations is predominantly inflated upwards.

Table 2.5 Summary of power results: $T = 1000$

	Frequency of rejection of the null							
Hypothesis H_0	BDS	MLQ	MQ	KEEN	TSAY	STAR	WHITE	NN
moment								
$H_0 : avg_j = avg_1$	0.57	0.78	0.74	0.48	0.57	0.61	0.57	0.33
$H_0 : avg_j > avg_1$	0.00	0.11	0.07	0.06	0.11	0.15	0.11	0.17
$H_0 : avg_j < avg_1$	0.65	0.80	0.78	0.46	0.46	0.52	0.46	0.28
asymmetry								
$H_0 : avg_j = avg_1$	0.78	0.87	0.87	0.80	0.76	0.57	0.57	0.56
$H_0 : avg_j > avg_1$	0.11	0.09	0.09	0.26	0.17	0.22	0.22	0.20
$H_0 : avg_j < avg_1$	0.70	0.81	0.80	0.57	0.59	0.46	0.39	0.37

* Note that $j \in \{2, \ldots, 7\}$ for moment condition failure, whereas $j \in \{8, \ldots, 13\}$ for asymmetry of innovations.

Much more interesting results are obtained from the robustness of the non-linearity tests against asymmetry of innovations. The results are presented in the bottom panel of Table 2.5. The results indicate that the average rejection frequencies of the tests based on asymmetric innovations suffer from even a higher variation as compared to moment condition failure. The only exception are the STAR and WHITE tests with almost identical results. The highest sensitivity is observed, rather surprisingly, for the Q tests where the null hypothesis about no change in the power (i.e. $H_0 : avg_j = avg_1$) is rejected in 87 % of all cases. As in the case of moment condition failure, the results confirm that the average rejection frequency of the tests is statistically mainly inflated upward in the case of asymmetric innovations.

2.4.4 Summary

In this chapter, the size and power properties of the standard non-linearity tests against: (a) parameter configurations of DGPs; (b) moment condition failure of innovations; and (c) asymmetry of innovations have examined using extensive Monte Carlo experiments. The aim of this section is to summarize the results and offer some conclusions for time series modelling.

The easiest way to compare the selected non-linearity tests is to order them according to their performance under different conditions: (a) the average rejection frequency, (b) robustness of the average rejection frequency against parameter configurations of DGPs; (c) against moment condition failure of innovations; (d) against asymmetry of innovations. Since eight non-linearity tests are evaluated, "1" denotes the best performance, whereas "8" the worst performance. Detailed results about the performance of the tests for the large sample $T = 1000$ are presented in Table 2.25.[6] For example, when considering the average rejection frequency of the tests itself, the results indicate that the BDS test is the most powerful test statistic for a MAR model ("1"), whereas the NN test does exhibit the lowest power across all 8 non-linearity tests ("8") for a MAR model. The same system of ordering is applied when evaluating the robustness of the tests against a parameter configuration, moment condition failure, and asymmetry of innovations. For example, when evaluating the robustness of the tests against asymmetry of innovations, the results show that the BDS test performs very badly for a TAR model ("8"), whereas the NN test does perform best in this case ("1"). Aggregated results, based on the median ordering of the results over all time series models under consideration, are presented in Table 2.6. The main reason for using median ordering is in the robustness of the results against outliers. That means, our approach penalizes tests, which perform very well for one particular time series model but completely fail for some other(s).[7]

[6] Note that similar results are obtained for other sample sizes.

[7] Another advantage of median ordering is that the results are not affected by rounding, which is not the case when using average ordering.

The results reveal that the non-linearity tests with the highest average rejection frequency across all 9 non-linear time series models are the BDS and MLQ tests. The worst test, having the lowest average rejection frequency, is the KEEN test. It is interesting to note that the overall performance of simple Q tests is better than more sophisticated non-linearity tests such as the WHITE or NN tests. Nevertheless, even very powerful tests suffer from a surprisingly high power variation. A nice example is related to the BDS test, which is the most powerful test for a given set of non-linear models, but the test also suffers from one of the highest power variation. Ordering of the non-linearity tests according to their robustness against moment failure and asymmetry of innovations gives rather different results. In the case of moment failure, the lowest variability of the average rejection frequency is obtained for the BDS and NN tests. The worst test, suffering from the highest power variation, seems to be the TSAY test. In the case of asymmetric innovations, the lowest variability of the average rejection frequency is observed for the STAR, WHITE, and NN tests. The worst test, suffering from the highest variation, seems to be the BDS test.

Table 2.6 Median ordering of non-linearity test: $T = 1000$

	median ordering			
	avg_1	cv(N)	cv(S)	cv(A)
BDS	1	6	1	6
MLQ	3	6	5	6
MQ	4	7	6	5
KEEN	6	5	5	5
TSAY	5	4	7	4
STAR	4	4	5	2
WHITE	4	5	4	3
NN	5	3	3	3

* "avg_1" stands for the average rejection frequency of the non-linearity tests based on Gaussian (N) innovations, "cv(N)" stands for the coefficient of variation calculated over all parameter configurations of a given non-linear model using Gaussian (N) innovations, "cv(S)" stands for a coefficient of variation of a given test statistic over all symmetric (S) innovations, "cv(A)" stands for a coefficient of variation of a given test statistic over all asymmetric (A) innovations.

2.5 Conclusion

In this chapter, nine non-linear models are examined using a battery of eight standard non-linearity tests. Our results suggest that one should interpret the results from the non-linearity tests with caution, since linearity does not have to be rejected only due to a particular parameter configuration of a purely non-linear process. In particular, the power of the non-linearity tests is robust (i.e. rejecting or not rejecting linearity) against DGP parameters only in less than 50 % of cases. In addition, it is demonstrated that the power of the tests is statistically significantly inflated upwards in the case of moment condition failure and/or asymmetry of innovations. Based on median ordering of the non-linearity test, it can be concluded that there is no clear link between the performance of the tests and their moments requirements. What matters is the construction of the tests rather than the moment requirement itself. All in all, since economic time series do have marginal distributions which are significantly leptokurtic rather than asymmetric, the results from a symmetric distribution might be preferred. In such a case, the best tests, according to their average rejection frequency and its variability, are the following: the BDS and NN tests, followed by the MLQ and STAR tests.

2.6 Appendix A: Tables

Table 2.7 Size of the non-linearity tests: NID(0,1) innovations

	T=200		**T=500**		**T=1000**	
	avg	cv(N)	avg	cv(N)	avg	cv(N)
BDS(n)	0.08	0.34	0.06	0.27	0.06	0.33
MLQ(m)	0.06	0.35	0.06	0.36	0.06	0.29
MQ(m)	0.06	0.38	0.06	0.35	0.06	0.29
KEENAN	0.04	0.66	0.05	0.45	0.05	0.40
TSAY(p)	0.04	0.63	0.05	0.48	0.05	0.39
STAR(p)	0.04	0.49	0.05	0.28	0.05	0.46
WHITE(p)	0.02	0.42	0.02	0.74	0.02	0.48
NN(p)	0.05	0.51	0.05	0.28	0.05	0.42

[a] The lag order p of an AR process is determined by an automatic lag order selection procedure discussed in Ng and Perron (2005). The n–history of the BDS test is set $n = 2$ for $T = 200$, $n = 3$ for $T = 500$, and $n = 4$ for $T = 1000$. The lag order m of the Q tests is determined by an automatic selection procedure developed by Escanciano and Lobato (2009).

[b] AR (#8) indicates that we evaluate $K = 8$ different parameter configurations of an AR model.

[c] avg denotes the average rejection frequency calculated over all parameter configurations of a given DGP, cv(N) represents a coefficient of variation calculated from individual rejection frequencies. The significance level is set to $\alpha = 0.05$.

Table 2.8 Power of the non-linearity tests: NID(0,1) innovations

	TAR (#24)		**EXPAR (#24)**		**MAR (#24)**		**MSAR (#24)**		**GARCH (#12)**		**TMA (#24)**		**BL (#18)**		**RCAR (#18)**		**NLMA (#12)**	
T=200	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)
BDS	0.35	2.31	0.10	0.74	0.55	1.71	0.59	1.38	0.17	0.90	0.23	2.29	0.47	1.35	0.15	2.27	0.09	0.39
MLQ(m)	0.22	3.17	0.07	0.33	0.51	1.84	0.53	1.61	0.14	0.98	0.10	1.85	0.28	2.24	0.12	2.53	0.07	0.45
MQ(m)	0.22	3.33	0.06	0.50	0.51	1.86	0.51	1.65	0.13	0.96	0.09	1.75	0.27	2.26	0.11	2.55	0.07	0.45
KEENAN	0.50	1.76	0.06	2.81	0.13	1.90	0.15	1.69	0.06	0.57	0.47	1.94	0.79	0.90	0.08	2.04	0.05	0.61
TSAY(p)	0.53	1.76	0.06	2.84	0.14	1.92	0.24	1.83	0.06	0.61	0.56	1.74	0.90	0.22	0.08	2.01	0.05	0.47
STAR(p)	0.60	1.52	0.27	2.49	0.19	2.13	0.32	1.71	0.07	0.53	0.54	1.79	0.89	0.23	0.10	2.06	0.05	0.61
WHITE(p)	0.43	2.04	0.13	3.22	0.16	3.18	0.33	2.09	0.04	0.92	0.46	2.05	0.84	0.31	0.05	3.11	0.02	0.88
NN(p)	0.60	1.38	0.44	2.04	0.10	1.49	0.26	1.69	0.06	0.48	0.51	1.78	0.81	0.45	0.07	1.61	0.05	0.44
	TAR (#24)		**EXPAR (#24)**		**MAR (#24)**		**MSAR (#24)**		**GARCH (#12)**		**TMA (#24)**		**BL (#18)**		**RCAR (#18)**		**NLMA (#12)**	
T=500	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)
BDS	0.53	1.71	0.09	1.40	0.63	1.52	0.76	0.93	0.27	1.10	0.38	2.34	0.76	0.45	0.21	3.19	0.07	0.53
MLQ(m)	0.36	2.59	0.06	0.60	0.60	1.58	0.72	1.08	0.26	1.20	0.16	2.65	0.52	1.28	0.19	3.35	0.06	0.46
MQ(m)	0.34	2.71	0.06	0.52	0.60	1.58	0.71	1.13	0.24	1.20	0.16	2.69	0.51	1.35	0.18	3.40	0.06	0.46
KEENAN	0.65	1.42	0.07	2.96	0.15	1.80	0.17	2.02	0.07	0.58	0.51	1.82	0.87	0.86	0.10	2.08	0.05	0.37
TSAY(p)	0.70	1.37	0.07	2.99	0.16	1.90	0.35	1.67	0.07	0.58	0.69	1.42	1.00	0.04	0.10	2.08	0.06	0.44
STAR(p)	0.77	1.13	0.50	1.89	0.22	2.03	0.46	1.59	0.08	0.68	0.71	1.39	1.00	0.02	0.12	2.29	0.06	0.52
WHITE(p)	0.66	1.42	0.36	2.51	0.21	2.93	0.54	1.56	0.06	0.92	0.67	1.46	1.00	0.01	0.07	3.29	0.04	0.95
NN(p)	0.80	0.87	0.64	1.47	0.11	1.29	0.45	1.58	0.07	0.59	0.62	1.50	0.98	0.16	0.08	1.84	0.05	0.44
	TAR (#24)		**EXPAR (#24)**		**MAR (#24)**		**MSAR (#24)**		**GARCH (#12)**		**TMA (#24)**		**BL (#18)**		**RCAR (#18)**		**NLMA (#12)**	
T=1000	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)	avg	cv(N)
BDS	0.68	1.34	0.10	2.69	0.68	1.41	0.86	0.58	0.44	1.11	0.49	1.94	0.96	0.08	0.29	2.79	0.07	0.58
MLQ(m)	0.45	2.05	0.06	1.13	0.65	1.46	0.84	0.74	0.44	1.23	0.26	2.81	0.76	0.50	0.26	3.11	0.06	0.36
MQ(m)	0.44	2.12	0.06	0.97	0.65	1.46	0.82	0.79	0.42	1.24	0.25	2.81	0.75	0.52	0.25	3.20	0.06	0.36
KEENAN	0.70	1.32	0.07	2.84	0.16	1.65	0.20	2.47	0.08	0.67	0.52	1.80	0.85	0.91	0.10	1.99	0.06	0.51
TSAY(p)	0.78	1.21	0.07	2.88	0.16	1.72	0.46	1.47	0.08	0.69	0.76	1.29	1.00	0.01	0.11	1.99	0.06	0.55
STAR(p)	0.86	0.78	0.62	1.55	0.24	1.97	0.59	1.34	0.09	0.57	0.78	1.25	1.00	0.00	0.14	2.40	0.06	0.71
WHITE(p)	0.77	1.15	0.52	1.87	0.24	2.77	0.67	1.29	0.07	0.85	0.77	1.24	1.00	0.00	0.08	3.74	0.05	1.13
NN(p)	0.92	0.42	0.71	1.36	0.12	1.39	0.60	1.40	0.08	0.44	0.66	1.40	0.99	0.10	0.09	1.93	0.05	0.53

[a] The lag order p of an AR process is determined by an automatic lag order selection procedure discussed in Ng and Perron (2005). The n–history of the BDS test is set $n = 2$ for $T = 200$, $n = 3$ for $T = 500$, and $n = 4$ for $T = 1000$. The lag order m of the Q tests is determined by an automatic selection procedure developed by Escanciano and Lobato (2009).

[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.

[c] avg denotes the average rejection frequency calculated over all parameter configurations of a given DGP, cv(N) represents a coefficient of variation calculated from individual rejection frequencies. The significance level is set to $\alpha = 0.05$.

Table 2.9 Power properties: BDS test

T=200	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.07	0.07	0.07	0.07	0.07	0.08	0.08	0.18
TAR (#24)	0.42	0.39	0.37	0.36	0.36	0.35	0.35	0.20
EXPAR (#24)	0.14	0.11	0.10	0.10	0.10	0.10	0.10	0.46
MAR (#24)	0.62	0.59	0.58	0.57	0.57	0.57	0.55	0.14
MSAR (#24)	0.62	0.61	0.60	0.59	0.59	0.60	0.59	0.05
GARCH (#12)	0.16	0.15	0.15	0.15	0.15	0.16	0.17	0.08
TMA (#24)	0.45	0.38	0.34	0.32	0.30	0.29	0.23	0.96
BL (#18)	0.83	0.76	0.70	0.66	0.63	0.61	0.47	0.77
RCA (#18)	0.22	0.18	0.17	0.16	0.15	0.15	0.15	0.49
NLMA (#12)	0.09	0.08	0.09	0.08	0.09	0.09	0.09	0.12
T=500	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.06	0.06	0.05	0.06	0.06	0.06	0.06	0.16
TAR (#24)	0.64	0.61	0.59	0.58	0.57	0.58	0.53	0.21
EXPAR (#24)	0.20	0.13	0.11	0.10	0.09	0.10	0.09	1.29
MAR (#24)	0.70	0.67	0.66	0.65	0.65	0.65	0.63	0.12
MSAR (#24)	0.77	0.76	0.76	0.76	0.76	0.77	0.76	0.02
GARCH (#12)	0.29	0.29	0.29	0.30	0.30	0.30	0.27	0.13
TMA (#24)	0.60	0.53	0.50	0.47	0.45	0.44	0.38	0.59
BL (#18)	0.98	0.96	0.94	0.92	0.90	0.88	0.76	0.29
RCA (#18)	0.38	0.31	0.28	0.26	0.25	0.25	0.21	0.83
NLMA (#12)	0.06	0.06	0.06	0.07	0.07	0.07	0.07	0.13
T=1000	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.06	0.05	0.06	0.05	0.05	0.05	0.06	0.10
TAR (#24)	0.78	0.75	0.74	0.74	0.73	0.73	0.68	0.15
EXPAR (#24)	0.28	0.19	0.15	0.14	0.13	0.12	0.10	1.89
MAR (#24)	0.74	0.72	0.70	0.69	0.69	0.69	0.68	0.09
MSAR (#24)	0.86	0.85	0.86	0.86	0.87	0.87	0.86	0.02
GARCH (#12)	0.45	0.46	0.48	0.49	0.48	0.49	0.44	0.12
TMA (#24)	0.67	0.61	0.58	0.56	0.54	0.53	0.49	0.37
BL (#18)	1.00	1.00	1.00	0.99	0.98	0.98	0.96	0.04
RCA (#18)	0.54	0.43	0.39	0.37	0.35	0.34	0.29	0.88
NLMA (#12)	0.06	0.06	0.06	0.06	0.07	0.07	0.07	0.16

[a] The lag order p of an AR process is determined by an automatic lag order selection procedure discussed in Ng and Perron (2005). The n–history of the BDS test is set $n = 2$ for $T = 200$, $n = 3$ for $T = 500$, and $n = 4$ for $T = 1000$.

[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.

[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(S) denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations. The significance level is set to $\alpha = 0.05$.

[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.10 Power properties: MLQ test

T=200	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.19
TAR (#24)	0.31	0.27	0.26	0.25	0.24	0.24	0.22	0.40
EXPAR (#24)	0.09	0.07	0.07	0.07	0.07	0.07	0.07	0.30
MAR (#24)	0.59	0.56	0.56	0.54	0.54	0.54	0.51	0.15
MSAR (#24)	0.55	0.55	0.55	0.54	0.54	0.54	0.53	0.03
GARCH (#12)	0.19	0.18	0.17	0.17	0.17	0.16	0.14	0.42
TMA (#24)	0.39	0.30	0.24	0.21	0.18	0.17	0.10	3.07
BL (#18)	0.88	0.77	0.67	0.61	0.55	0.52	0.28	2.12
RCA (#18)	0.19	0.16	0.15	0.14	0.13	0.13	0.12	0.56
NLMA (#12)	0.09	0.09	0.08	0.08	0.08	0.08	0.07	0.29
T=500	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.23
TAR (#24)	0.51	0.46	0.42	0.40	0.39	0.38	0.36	0.44
EXPAR (#24)	0.10	0.08	0.07	0.07	0.07	0.07	0.06	0.52
MAR (#24)	0.68	0.65	0.64	0.63	0.63	0.62	0.60	0.13
MSAR (#24)	0.70	0.70	0.70	0.71	0.71	0.71	0.72	0.04
GARCH (#12)	0.37	0.35	0.32	0.31	0.31	0.30	0.26	0.40
TMA (#24)	0.60	0.51	0.43	0.38	0.34	0.32	0.16	2.70
BL (#18)	1.00	0.99	0.96	0.92	0.88	0.85	0.52	0.93
RCA (#18)	0.34	0.28	0.25	0.23	0.22	0.22	0.19	0.85
NLMA (#12)	0.12	0.09	0.08	0.08	0.08	0.08	0.06	0.85
T=1000	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.07	0.06	0.07	0.07	0.06	0.06	0.06	0.31
TAR (#24)	0.66	0.61	0.56	0.53	0.51	0.50	0.45	0.46
EXPAR (#24)	0.13	0.09	0.08	0.08	0.07	0.07	0.06	0.99
MAR (#24)	0.73	0.71	0.69	0.68	0.67	0.67	0.65	0.12
MSAR (#24)	0.78	0.78	0.79	0.80	0.81	0.81	0.84	0.07
GARCH (#12)	0.57	0.55	0.52	0.51	0.49	0.49	0.44	0.30
TMA (#24)	0.70	0.62	0.56	0.52	0.48	0.45	0.26	1.73
BL (#18)	1.00	1.00	1.00	1.00	0.99	0.98	0.76	0.31
RCA (#18)	0.54	0.43	0.38	0.35	0.33	0.31	0.26	1.10
NLMA (#12)	0.15	0.11	0.09	0.09	0.08	0.08	0.06	1.44

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005). The lag order m of the Q tests is determined by an automatic selection procedure developed by Escanciano and Lobato (2009).

[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.

[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(S) denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations. The significance level is set to $\alpha = 0.05$.

[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.11 Power properties: MQ test

T=200	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.07	0.07	0.06	0.07	0.06	0.06	0.06	0.25
TAR (#24)	0.31	0.27	0.25	0.24	0.23	0.23	0.22	0.42
EXPAR (#24)	0.08	0.07	0.07	0.07	0.07	0.07	0.06	0.30
MAR (#24)	0.58	0.56	0.55	0.54	0.54	0.53	0.51	0.15
MSAR (#24)	0.54	0.54	0.53	0.53	0.53	0.53	0.51	0.05
GARCH (#12)	0.18	0.17	0.16	0.15	0.16	0.15	0.13	0.42
TMA (#24)	0.39	0.29	0.24	0.20	0.18	0.17	0.09	3.11
BL (#18)	0.88	0.77	0.67	0.60	0.54	0.51	0.27	2.22
RCA (#18)	0.18	0.15	0.14	0.13	0.13	0.13	0.11	0.58
NLMA (#12)	0.09	0.09	0.08	0.08	0.08	0.08	0.07	0.27
T=500	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.22
TAR (#24)	0.50	0.45	0.41	0.39	0.37	0.37	0.34	0.47
EXPAR (#24)	0.09	0.08	0.07	0.07	0.07	0.07	0.06	0.48
MAR (#24)	0.67	0.65	0.64	0.63	0.62	0.62	0.60	0.13
MSAR (#24)	0.69	0.69	0.69	0.69	0.70	0.69	0.71	0.03
GARCH (#12)	0.35	0.33	0.30	0.29	0.29	0.29	0.24	0.43
TMA (#24)	0.60	0.51	0.43	0.38	0.34	0.31	0.16	2.72
BL (#18)	1.00	0.98	0.96	0.92	0.88	0.84	0.51	0.97
RCA (#18)	0.33	0.27	0.24	0.22	0.22	0.21	0.18	0.85
NLMA (#12)	0.12	0.09	0.08	0.08	0.08	0.08	0.06	0.83
T=1000	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.07	0.07	0.07	0.07	0.06	0.06	0.06	0.31
TAR (#24)	0.65	0.60	0.55	0.52	0.50	0.49	0.44	0.48
EXPAR (#24)	0.12	0.09	0.08	0.07	0.07	0.07	0.06	1.00
MAR (#24)	0.72	0.70	0.69	0.68	0.67	0.67	0.65	0.12
MSAR (#24)	0.77	0.77	0.78	0.79	0.80	0.80	0.82	0.07
GARCH (#12)	0.55	0.53	0.50	0.49	0.47	0.46	0.42	0.32
TMA (#24)	0.70	0.62	0.56	0.52	0.48	0.45	0.25	1.73
BL (#18)	1.00	1.00	1.00	1.00	0.99	0.98	0.75	0.33
RCA (#18)	0.52	0.42	0.37	0.34	0.32	0.30	0.25	1.11
NLMA (#12)	0.15	0.11	0.09	0.09	0.08	0.08	0.06	1.42

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005). The lag order m of the Q tests is determined by an automatic selection procedure developed by Escanciano and Lobato (2009).

[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.

[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(S) denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations. The significance level is set to $\alpha = 0.05$.

[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.12 Power properties: KEEN test

T=200	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.04	0.04	0.05	0.04	0.05	0.04	0.04	0.10
TAR (#24)	0.51	0.52	0.51	0.52	0.52	0.51	0.50	0.04
EXPAR (#24)	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.23
MAR (#24)	0.24	0.20	0.19	0.17	0.16	0.16	0.13	0.86
MSAR (#24)	0.18	0.17	0.17	0.16	0.16	0.15	0.15	0.25
GARCH (#12)	0.10	0.09	0.08	0.08	0.08	0.07	0.06	0.57
TMA (#24)	0.51	0.49	0.50	0.49	0.49	0.48	0.47	0.09
BL (#18)	0.71	0.75	0.77	0.78	0.78	0.79	0.79	0.10
RCA (#18)	0.12	0.11	0.10	0.09	0.09	0.09	0.08	0.59
NLMA (#12)	0.07	0.07	0.06	0.07	0.07	0.07	0.05	0.38
T=500	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.18
TAR (#24)	0.63	0.64	0.64	0.64	0.65	0.65	0.65	0.03
EXPAR (#24)	0.10	0.08	0.07	0.07	0.06	0.06	0.07	0.51
MAR (#24)	0.33	0.27	0.23	0.21	0.19	0.18	0.15	1.20
MSAR (#24)	0.22	0.20	0.19	0.18	0.18	0.18	0.17	0.30
GARCH (#12)	0.16	0.14	0.11	0.10	0.10	0.09	0.07	1.31
TMA (#24)	0.53	0.53	0.53	0.52	0.52	0.52	0.51	0.04
BL (#18)	0.73	0.78	0.82	0.83	0.84	0.84	0.87	0.16
RCA (#18)	0.19	0.15	0.13	0.12	0.11	0.11	0.10	0.96
NLMA (#12)	0.10	0.09	0.08	0.08	0.08	0.08	0.05	0.80
T=1000	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.05	0.04	0.04	0.05	0.05	0.05	0.05	0.15
TAR (#24)	0.68	0.68	0.67	0.68	0.68	0.69	0.70	0.04
EXPAR (#24)	0.12	0.09	0.08	0.07	0.07	0.07	0.07	0.81
MAR (#24)	0.39	0.31	0.26	0.23	0.21	0.20	0.16	1.50
MSAR (#24)	0.25	0.23	0.22	0.21	0.21	0.21	0.20	0.26
GARCH (#12)	0.22	0.18	0.14	0.13	0.11	0.11	0.08	1.77
TMA (#24)	0.55	0.54	0.53	0.53	0.53	0.52	0.52	0.06
BL (#18)	0.73	0.78	0.81	0.83	0.84	0.85	0.85	0.15
RCA (#18)	0.25	0.18	0.16	0.13	0.13	0.12	0.10	1.39
NLMA (#12)	0.12	0.10	0.09	0.08	0.09	0.08	0.06	0.97

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005).

[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.

[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(S) denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations. The significance level is set to $\alpha = 0.05$.

[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.13 Power properties: TSAY test

T=200	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.04	0.04	0.04	0.04	0.05	0.04	0.04	0.09
TAR (#24)	0.57	0.57	0.56	0.56	0.56	0.55	0.53	0.07
EXPAR (#24)	0.08	0.07	0.06	0.06	0.06	0.06	0.06	0.29
MAR (#24)	0.28	0.22	0.20	0.19	0.18	0.17	0.14	0.95
MSAR (#24)	0.31	0.29	0.27	0.27	0.26	0.25	0.24	0.27
GARCH (#12)	0.11	0.10	0.09	0.08	0.08	0.08	0.06	0.71
TMA (#24)	0.63	0.61	0.60	0.59	0.59	0.58	0.56	0.14
BL (#18)	0.91	0.91	0.92	0.92	0.92	0.92	0.90	0.02
RCA (#18)	0.13	0.11	0.10	0.10	0.09	0.09	0.08	0.66
NLMA (#12)	0.07	0.07	0.07	0.07	0.07	0.07	0.05	0.48
T=500	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.14
TAR (#24)	0.73	0.72	0.72	0.72	0.72	0.72	0.70	0.03
EXPAR (#24)	0.10	0.08	0.07	0.07	0.07	0.07	0.07	0.56
MAR (#24)	0.37	0.29	0.25	0.22	0.21	0.20	0.16	1.34
MSAR (#24)	0.44	0.41	0.38	0.37	0.36	0.36	0.35	0.28
GARCH (#12)	0.18	0.15	0.12	0.11	0.10	0.10	0.07	1.60
TMA (#24)	0.74	0.73	0.72	0.71	0.71	0.71	0.69	0.07
BL (#18)	0.95	0.97	0.97	0.98	0.99	0.99	1.00	0.05
RCA (#18)	0.20	0.16	0.13	0.12	0.12	0.12	0.10	1.08
NLMA (#12)	0.11	0.10	0.09	0.09	0.09	0.09	0.06	0.95
T=1000	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.16
TAR (#24)	0.80	0.79	0.79	0.78	0.78	0.78	0.78	0.02
EXPAR (#24)	0.13	0.10	0.08	0.07	0.07	0.07	0.07	0.91
MAR (#24)	0.43	0.33	0.28	0.25	0.22	0.21	0.16	1.65
MSAR (#24)	0.53	0.49	0.47	0.46	0.46	0.46	0.46	0.16
GARCH (#12)	0.25	0.19	0.14	0.13	0.12	0.11	0.08	2.04
TMA (#24)	0.80	0.78	0.78	0.77	0.77	0.77	0.76	0.06
BL (#18)	0.96	0.97	0.98	0.99	0.99	0.99	1.00	0.04
RCA (#18)	0.27	0.19	0.16	0.13	0.13	0.13	0.11	1.56
NLMA (#12)	0.14	0.12	0.12	0.11	0.11	0.11	0.06	1.28

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005).

[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.

[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(S) denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations. The significance level is set to $\alpha = 0.05$.

[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.14 Power properties: STAR test

T=200	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.05	0.04	0.04	0.04	0.05	0.04	0.04	0.17
TAR (#24)	0.61	0.62	0.61	0.61	0.62	0.61	0.60	0.04
EXPAR (#24)	0.22	0.23	0.24	0.24	0.24	0.25	0.27	0.20
MAR (#24)	0.36	0.30	0.27	0.25	0.24	0.23	0.19	0.94
MSAR (#24)	0.37	0.36	0.34	0.34	0.33	0.33	0.32	0.16
GARCH (#12)	0.12	0.11	0.10	0.09	0.08	0.09	0.07	0.77
TMA (#24)	0.63	0.61	0.60	0.59	0.58	0.58	0.54	0.16
BL (#18)	0.96	0.95	0.94	0.94	0.94	0.93	0.89	0.08
RCA (#18)	0.15	0.13	0.13	0.11	0.11	0.11	0.10	0.59
NLMA (#12)	0.08	0.08	0.08	0.08	0.07	0.07	0.05	0.64
T=500	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
TAR (#24)	0.78	0.78	0.78	0.78	0.78	0.77	0.77	0.02
EXPAR (#24)	0.38	0.40	0.42	0.43	0.44	0.45	0.50	0.24
MAR (#24)	0.49	0.40	0.36	0.32	0.30	0.28	0.22	1.18
MSAR (#24)	0.54	0.51	0.49	0.49	0.48	0.48	0.46	0.17
GARCH (#12)	0.21	0.18	0.15	0.13	0.12	0.12	0.08	1.72
TMA (#24)	0.77	0.75	0.74	0.73	0.73	0.73	0.71	0.10
BL (#18)	0.99	0.99	0.99	0.99	1.00	1.00	1.00	0.01
RCA (#18)	0.24	0.19	0.16	0.15	0.15	0.15	0.12	1.00
NLMA (#12)	0.12	0.11	0.10	0.09	0.09	0.09	0.06	1.06
T=1000	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.13
TAR (#24)	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.01
EXPAR (#24)	0.49	0.51	0.54	0.56	0.57	0.58	0.62	0.21
MAR (#24)	0.56	0.47	0.41	0.37	0.34	0.32	0.24	1.38
MSAR (#24)	0.64	0.61	0.60	0.60	0.60	0.59	0.59	0.08
GARCH (#12)	0.29	0.24	0.19	0.17	0.15	0.13	0.09	2.07
TMA (#24)	0.85	0.83	0.82	0.81	0.80	0.80	0.78	0.08
BL (#18)	0.99	0.99	1.00	1.00	1.00	1.00	1.00	0.01
RCA (#18)	0.34	0.25	0.21	0.19	0.17	0.16	0.14	1.55
NLMA (#12)	0.16	0.14	0.13	0.11	0.12	0.11	0.06	1.63

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005).
[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.
[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(S) denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations. The significance level is set to $\alpha = 0.05$.
[e] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.15 Power properties: WHITE test

T=200	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.03	0.03	0.02	0.02	0.02	0.02	0.02	1.07
TAR (#24)	0.49	0.48	0.47	0.46	0.46	0.45	0.43	0.14
EXPAR (#24)	0.11	0.11	0.12	0.11	0.11	0.12	0.13	0.15
MAR (#24)	0.33	0.27	0.24	0.22	0.20	0.20	0.16	1.10
MSAR (#24)	0.37	0.36	0.35	0.35	0.34	0.34	0.33	0.12
GARCH (#12)	0.12	0.10	0.09	0.08	0.07	0.07	0.04	1.58
TMA (#24)	0.56	0.54	0.52	0.51	0.50	0.49	0.46	0.23
BL (#18)	0.97	0.96	0.95	0.94	0.93	0.92	0.84	0.16
RCA (#18)	0.11	0.09	0.08	0.07	0.07	0.06	0.05	1.20
NLMA (#12)	0.06	0.06	0.05	0.05	0.05	0.05	0.02	1.53
T=500	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.04	0.03	0.02	0.03	0.02	0.02	0.02	0.74
TAR (#24)	0.69	0.68	0.68	0.68	0.67	0.67	0.66	0.04
EXPAR (#24)	0.24	0.26	0.28	0.29	0.30	0.30	0.36	0.32
MAR (#24)	0.49	0.41	0.35	0.33	0.30	0.28	0.21	1.33
MSAR (#24)	0.58	0.56	0.55	0.55	0.54	0.54	0.54	0.08
GARCH (#12)	0.21	0.17	0.15	0.12	0.11	0.10	0.06	2.58
TMA (#24)	0.73	0.71	0.70	0.69	0.68	0.68	0.67	0.10
BL (#18)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
RCA (#18)	0.20	0.15	0.12	0.11	0.10	0.09	0.07	1.74
NLMA (#12)	0.11	0.10	0.09	0.09	0.09	0.08	0.04	1.73
T=1000	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.04	0.03	0.03	0.03	0.03	0.03	0.02	0.48
TAR (#24)	0.78	0.78	0.77	0.77	0.77	0.77	0.77	0.01
EXPAR (#24)	0.38	0.40	0.43	0.44	0.46	0.47	0.52	0.27
MAR (#24)	0.58	0.49	0.43	0.38	0.35	0.33	0.24	1.48
MSAR (#24)	0.68	0.66	0.66	0.66	0.66	0.65	0.67	0.04
GARCH (#12)	0.30	0.25	0.19	0.17	0.15	0.13	0.07	3.21
TMA (#24)	0.81	0.80	0.79	0.78	0.78	0.77	0.77	0.06
BL (#18)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
RCA (#18)	0.31	0.21	0.16	0.14	0.13	0.12	0.08	2.80
NLMA (#12)	0.16	0.15	0.13	0.13	0.12	0.12	0.05	2.09

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005).

[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.

[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(S) denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations. The significance level is set to $\alpha = 0.05$.

[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.16 Power properties: NN test

T=200	S3	S4	S5	S6	S7	S8	N	cv(S)
AR (#8)	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.23
TAR (#24)	0.63	0.63	0.62	0.62	0.62	0.62	0.60	0.06
EXPAR (#24)	0.46	0.46	0.45	0.45	0.45	0.44	0.44	0.03
MAR (#24)	0.18	0.15	0.14	0.13	0.12	0.12	0.10	0.71
MSAR (#24)	0.25	0.25	0.25	0.26	0.25	0.26	0.26	0.04
GARCH (#12)	0.10	0.09	0.08	0.08	0.07	0.08	0.06	0.60
TMA (#24)	0.59	0.57	0.56	0.55	0.55	0.54	0.51	0.15
BL (#18)	0.89	0.89	0.88	0.87	0.86	0.86	0.81	0.10
RCA (#18)	0.11	0.09	0.09	0.08	0.08	0.08	0.07	0.49
NLMA (#12)	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.31
T=500	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.24
TAR (#24)	0.83	0.82	0.82	0.81	0.81	0.81	0.80	0.03
EXPAR (#24)	0.64	0.64	0.64	0.63	0.63	0.64	0.64	0.01
MAR (#24)	0.22	0.18	0.16	0.14	0.14	0.13	0.11	0.89
MSAR (#24)	0.38	0.39	0.40	0.41	0.42	0.42	0.45	0.15
GARCH (#12)	0.14	0.12	0.11	0.10	0.09	0.09	0.07	1.09
TMA (#24)	0.70	0.68	0.66	0.65	0.65	0.65	0.62	0.12
BL (#18)	0.96	0.96	0.97	0.97	0.97	0.97	0.98	0.02
RCA (#18)	0.14	0.11	0.10	0.10	0.09	0.09	0.08	0.73
NLMA (#12)	0.06	0.06	0.06	0.06	0.06	0.07	0.05	0.25
T=1000	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**N**	**cv(S)**
AR (#8)	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.07
TAR (#24)	0.93	0.93	0.92	0.92	0.92	0.92	0.92	0.01
EXPAR (#24)	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.00
MAR (#24)	0.24	0.19	0.16	0.16	0.14	0.14	0.12	1.07
MSAR (#24)	0.50	0.53	0.55	0.56	0.57	0.57	0.60	0.16
GARCH (#12)	0.18	0.15	0.12	0.11	0.10	0.09	0.08	1.37
TMA (#24)	0.75	0.73	0.71	0.70	0.69	0.69	0.66	0.13
BL (#18)	0.97	0.98	0.98	0.98	0.98	0.99	0.99	0.01
RCA (#18)	0.17	0.13	0.12	0.10	0.10	0.10	0.09	0.98
NLMA (#12)	0.07	0.07	0.07	0.07	0.07	0.07	0.05	0.51

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005).
[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.
[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(S) denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations. The significance level is set to $\alpha = 0.05$.
[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.17 Power properties: BDS test

T=200	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.08	0.08	0.08	0.07	0.08	0.08	0.08	0.17
TAR (#24)	0.27	0.57	0.24	0.57	0.32	0.50	0.35	0.93
EXPAR (#24)	0.10	0.10	0.11	0.11	0.12	0.13	0.10	0.32
MAR (#24)	0.61	0.61	0.60	0.60	0.61	0.61	0.55	0.12
MSAR (#24)	0.65	0.61	0.63	0.60	0.63	0.60	0.59	0.09
GARCH (#12)	0.15	0.15	0.15	0.15	0.15	0.15	0.17	0.11
TMA (#24)	0.32	0.32	0.29	0.29	0.38	0.38	0.23	0.65
BL (#18)	0.72	0.76	0.66	0.69	0.78	0.80	0.47	0.70
RCA (#18)	0.16	0.25	0.15	0.23	0.17	0.22	0.15	0.69
NLMA (#12)	0.16	0.32	0.17	0.33	0.11	0.17	0.09	2.62
T=500	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.13
TAR (#24)	0.34	0.84	0.32	0.83	0.45	0.76	0.53	0.98
EXPAR (#24)	0.10	0.10	0.15	0.15	0.17	0.17	0.09	0.96
MAR (#24)	0.69	0.69	0.67	0.67	0.69	0.69	0.63	0.10
MSAR (#24)	0.83	0.75	0.82	0.74	0.79	0.75	0.76	0.11
GARCH (#12)	0.28	0.28	0.28	0.28	0.27	0.27	0.27	0.04
TMA (#24)	0.47	0.46	0.43	0.44	0.53	0.53	0.38	0.42
BL (#18)	0.96	0.97	0.92	0.94	0.97	0.98	0.76	0.29
RCA (#18)	0.26	0.44	0.23	0.40	0.29	0.39	0.21	1.14
NLMA (#12)	0.24	0.53	0.25	0.54	0.13	0.26	0.07	6.28
T=1000	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.05	0.06	0.05	0.05	0.05	0.06	0.06	0.08
TAR (#24)	0.38	0.96	0.38	0.95	0.56	0.90	0.68	0.86
EXPAR (#24)	0.13	0.13	0.21	0.21	0.24	0.24	0.10	1.49
MAR (#24)	0.73	0.73	0.72	0.71	0.73	0.73	0.68	0.08
MSAR (#24)	0.93	0.82	0.92	0.82	0.89	0.83	0.86	0.12
GARCH (#12)	0.44	0.45	0.46	0.45	0.44	0.45	0.44	0.04
TMA (#24)	0.54	0.54	0.51	0.52	0.61	0.61	0.49	0.24
BL (#18)	1.00	1.00	0.99	0.99	1.00	1.00	0.96	0.05
RCA (#18)	0.35	0.61	0.32	0.56	0.39	0.56	0.29	1.12
NLMA (#12)	0.37	0.66	0.39	0.68	0.18	0.36	0.07	8.87

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005). The n–history of BDS test is set $n = 2$ for $T = 200$, $n = 3$ for $T = 500$, and $n = 4$ for $T = 1000$.

[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.

[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(A) denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations. The significance level is set to $\alpha = 0.05$.

[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.18 Power properties: MLQ test

T=200	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	**N**	**cv(A)**
AR (#8)	0.07	0.06	0.07	0.07	0.07	0.07	0.06	0.19
TAR (#24)	0.15	0.38	0.14	0.37	0.21	0.34	0.22	1.07
EXPAR (#24)	0.07	0.07	0.08	0.08	0.08	0.08	0.07	0.25
MAR (#24)	0.56	0.57	0.56	0.56	0.57	0.57	0.51	0.12
MSAR (#24)	0.57	0.53	0.57	0.53	0.57	0.54	0.53	0.08
GARCH (#12)	0.19	0.22	0.18	0.21	0.19	0.21	0.14	0.61
TMA (#24)	0.23	0.23	0.20	0.20	0.28	0.28	0.10	1.89
BL (#18)	0.66	0.74	0.58	0.65	0.76	0.81	0.28	1.88
RCA (#18)	0.14	0.20	0.13	0.18	0.15	0.19	0.12	0.64
NLMA (#12)	0.18	0.22	0.18	0.21	0.10	0.16	0.07	2.14
T=500	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	**N**	**cv(A)**
AR (#8)	0.07	0.06	0.07	0.06	0.07	0.07	0.06	0.17
TAR (#24)	0.28	0.53	0.25	0.52	0.37	0.48	0.36	0.78
EXPAR (#24)	0.07	0.07	0.08	0.08	0.09	0.09	0.06	0.39
MAR (#24)	0.66	0.66	0.64	0.65	0.66	0.67	0.60	0.11
MSAR (#24)	0.74	0.68	0.74	0.68	0.72	0.68	0.72	0.09
GARCH (#12)	0.39	0.40	0.36	0.38	0.36	0.38	0.26	0.52
TMA (#24)	0.37	0.37	0.33	0.33	0.45	0.45	0.16	1.80
BL (#18)	0.94	0.96	0.88	0.92	0.98	0.99	0.52	0.91
RCA (#18)	0.25	0.34	0.22	0.31	0.27	0.34	0.19	0.84
NLMA (#12)	0.41	0.42	0.42	0.41	0.18	0.28	0.06	5.69
T=1000	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	**N**	**cv(A)**
AR (#8)	0.07	0.06	0.07	0.07	0.07	0.08	0.06	0.35
TAR (#24)	0.39	0.62	0.38	0.61	0.50	0.57	0.45	0.52
EXPAR (#24)	0.07	0.07	0.09	0.09	0.10	0.10	0.06	0.61
MAR (#24)	0.70	0.71	0.69	0.69	0.71	0.71	0.65	0.10
MSAR (#24)	0.84	0.76	0.85	0.76	0.82	0.76	0.84	0.12
GARCH (#12)	0.58	0.59	0.56	0.57	0.57	0.58	0.44	0.35
TMA (#24)	0.46	0.46	0.43	0.42	0.55	0.55	0.26	1.17
BL (#18)	1.00	1.00	0.98	0.99	1.00	1.00	0.76	0.31
RCA (#18)	0.37	0.50	0.33	0.45	0.42	0.52	0.26	1.04
NLMA (#12)	0.67	0.63	0.68	0.62	0.33	0.44	0.06	10.08

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005). The lag order m of the Q tests is determined by an automatic selection procedure developed by Escanciano and Lobato (2009).

[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.

[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(A) denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations. The significance level is set to $\alpha = 0.05$.

[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.19 Power properties: MQ test

T=200	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	**N**	**cv(A)**
AR (#8)	0.07	0.06	0.07	0.06	0.07	0.07	0.06	0.19
TAR (#24)	0.15	0.37	0.13	0.36	0.20	0.34	0.22	1.09
EXPAR (#24)	0.07	0.07	0.08	0.07	0.08	0.08	0.06	0.25
MAR (#24)	0.56	0.57	0.55	0.55	0.57	0.57	0.51	0.12
MSAR (#24)	0.56	0.52	0.55	0.52	0.55	0.53	0.51	0.08
GARCH (#12)	0.18	0.21	0.17	0.19	0.18	0.19	0.13	0.59
TMA (#24)	0.22	0.22	0.19	0.19	0.27	0.27	0.09	1.90
BL (#18)	0.66	0.73	0.57	0.65	0.76	0.80	0.27	1.94
RCA (#18)	0.14	0.18	0.13	0.17	0.15	0.18	0.11	0.64
NLMA (#12)	0.18	0.22	0.18	0.21	0.10	0.16	0.07	2.11
T=500	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	**N**	**cv(A)**
AR (#8)	0.07	0.06	0.07	0.06	0.07	0.07	0.06	0.20
TAR (#24)	0.27	0.51	0.25	0.51	0.36	0.47	0.34	0.79
EXPAR (#24)	0.07	0.07	0.08	0.08	0.09	0.09	0.06	0.37
MAR (#24)	0.65	0.66	0.64	0.64	0.66	0.66	0.60	0.11
MSAR (#24)	0.72	0.67	0.73	0.67	0.71	0.68	0.71	0.08
GARCH (#12)	0.37	0.37	0.35	0.35	0.35	0.35	0.24	0.52
TMA (#24)	0.36	0.36	0.33	0.33	0.45	0.45	0.16	1.81
BL (#18)	0.94	0.96	0.87	0.92	0.98	0.99	0.51	0.95
RCA (#18)	0.24	0.32	0.21	0.30	0.26	0.33	0.18	0.86
NLMA (#12)	0.41	0.42	0.42	0.41	0.18	0.28	0.06	5.67
T=1000	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	**N**	**cv(A)**
AR (#8)	0.07	0.06	0.07	0.07	0.07	0.08	0.06	0.35
TAR (#24)	0.39	0.61	0.38	0.60	0.49	0.56	0.44	0.51
EXPAR (#24)	0.07	0.07	0.09	0.09	0.10	0.10	0.06	0.62
MAR (#24)	0.70	0.70	0.68	0.69	0.71	0.71	0.65	0.10
MSAR (#24)	0.83	0.75	0.84	0.76	0.81	0.76	0.82	0.11
GARCH (#12)	0.57	0.57	0.54	0.55	0.55	0.56	0.42	0.37
TMA (#24)	0.46	0.46	0.42	0.42	0.55	0.55	0.25	1.16
BL (#18)	1.00	1.00	0.98	0.99	1.00	1.00	0.75	0.33
RCA (#18)	0.36	0.49	0.32	0.43	0.41	0.51	0.25	1.04
NLMA (#12)	0.67	0.63	0.68	0.62	0.33	0.43	0.06	10.09

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005). The lag order m of the Q tests is determined by an automatic selection procedure developed by Escanciano and Lobato (2009).

[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.

[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(A) denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations. The significance level is set to $\alpha = 0.05$.

[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.20 Power properties: KEEN test

T=200	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.17
TAR (#24)	0.45	0.39	0.46	0.40	0.49	0.44	0.50	0.22
EXPAR (#24)	0.11	0.10	0.15	0.15	0.11	0.11	0.06	1.30
MAR (#24)	0.20	0.21	0.18	0.19	0.22	0.22	0.13	0.67
MSAR (#24)	0.17	0.17	0.17	0.17	0.18	0.17	0.15	0.20
GARCH (#12)	0.07	0.07	0.07	0.07	0.08	0.09	0.06	0.39
TMA (#24)	0.42	0.42	0.43	0.43	0.47	0.47	0.47	0.11
BL (#18)	0.63	0.61	0.66	0.63	0.69	0.65	0.79	0.23
RCA (#18)	0.10	0.11	0.09	0.10	0.11	0.12	0.08	0.54
NLMA (#12)	0.03	0.10	0.03	0.10	0.04	0.10	0.05	1.52
T=500	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.05	0.05	0.04	0.05	0.04	0.04	0.05	0.16
TAR (#24)	0.59	0.51	0.60	0.54	0.61	0.58	0.65	0.22
EXPAR (#24)	0.19	0.18	0.26	0.27	0.19	0.19	0.07	2.85
MAR (#24)	0.26	0.27	0.22	0.23	0.27	0.28	0.15	0.89
MSAR (#24)	0.21	0.21	0.19	0.20	0.20	0.20	0.17	0.22
GARCH (#12)	0.11	0.11	0.10	0.10	0.12	0.13	0.07	0.85
TMA (#24)	0.48	0.47	0.48	0.48	0.51	0.51	0.51	0.08
BL (#18)	0.61	0.60	0.64	0.60	0.67	0.62	0.87	0.31
RCA (#18)	0.13	0.14	0.12	0.13	0.15	0.16	0.10	0.70
NLMA (#12)	0.02	0.15	0.02	0.14	0.04	0.13	0.05	2.30
T=1000	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.05	0.05	0.04	0.04	0.04	0.05	0.05	0.21
TAR (#24)	0.66	0.61	0.66	0.64	0.68	0.65	0.70	0.13
EXPAR (#24)	0.27	0.27	0.36	0.36	0.29	0.29	0.07	3.98
MAR (#24)	0.29	0.30	0.25	0.26	0.32	0.33	0.16	1.15
MSAR (#24)	0.24	0.24	0.22	0.23	0.24	0.24	0.20	0.22
GARCH (#12)	0.15	0.15	0.12	0.13	0.18	0.18	0.08	1.24
TMA (#24)	0.49	0.49	0.50	0.50	0.53	0.53	0.52	0.08
BL (#18)	0.61	0.63	0.62	0.63	0.63	0.61	0.85	0.28
RCA (#18)	0.16	0.18	0.14	0.15	0.18	0.19	0.10	0.87
NLMA (#12)	0.03	0.16	0.03	0.15	0.04	0.17	0.06	2.41

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005).
[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.
[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(A) denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations. The significance level is set to $\alpha = 0.05$.
[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.21 Power properties: TSAY test

T=200	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.20
TAR (#24)	0.50	0.44	0.51	0.44	0.54	0.50	0.53	0.19
EXPAR (#24)	0.11	0.11	0.15	0.16	0.11	0.11	0.06	1.41
MAR (#24)	0.22	0.23	0.20	0.21	0.24	0.24	0.14	0.73
MSAR (#24)	0.29	0.29	0.27	0.28	0.30	0.29	0.24	0.24
GARCH (#12)	0.08	0.08	0.07	0.08	0.09	0.10	0.06	0.52
TMA (#24)	0.56	0.56	0.55	0.55	0.61	0.60	0.56	0.09
BL (#18)	0.90	0.93	0.90	0.93	0.90	0.93	0.90	0.04
RCA (#18)	0.10	0.12	0.09	0.11	0.11	0.13	0.08	0.59
NLMA (#12)	0.02	0.11	0.03	0.11	0.04	0.11	0.05	1.73
T=500	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.05	0.05	0.04	0.05	0.04	0.04	0.05	0.15
TAR (#24)	0.67	0.58	0.68	0.61	0.70	0.66	0.70	0.17
EXPAR (#24)	0.19	0.19	0.28	0.28	0.20	0.20	0.07	3.05
MAR (#24)	0.28	0.29	0.24	0.25	0.30	0.31	0.16	1.00
MSAR (#24)	0.41	0.43	0.38	0.42	0.42	0.42	0.35	0.23
GARCH (#12)	0.12	0.12	0.11	0.10	0.14	0.15	0.07	1.08
TMA (#24)	0.71	0.71	0.71	0.71	0.73	0.73	0.69	0.05
BL (#18)	0.99	0.98	0.98	0.99	0.97	0.98	1.00	0.02
RCA (#18)	0.14	0.15	0.12	0.14	0.16	0.17	0.10	0.78
NLMA (#12)	0.02	0.17	0.02	0.16	0.04	0.15	0.06	2.63
T=1000	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.19
TAR (#24)	0.75	0.70	0.76	0.72	0.78	0.75	0.78	0.10
EXPAR (#24)	0.28	0.28	0.39	0.39	0.31	0.31	0.07	4.39
MAR (#24)	0.31	0.33	0.27	0.28	0.35	0.36	0.16	1.24
MSAR (#24)	0.51	0.54	0.47	0.53	0.51	0.51	0.46	0.16
GARCH (#12)	0.16	0.16	0.13	0.13	0.19	0.19	0.08	1.41
TMA (#24)	0.78	0.78	0.78	0.78	0.79	0.79	0.76	0.05
BL (#18)	0.99	0.99	1.00	0.99	0.99	0.98	1.00	0.02
RCA (#18)	0.16	0.19	0.14	0.15	0.19	0.21	0.11	0.97
NLMA (#12)	0.03	0.21	0.03	0.19	0.05	0.21	0.06	2.87

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005).
[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.
[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(A) denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations. The significance level is set to $\alpha = 0.05$.
[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.22 Power properties: STAR test

T=200	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.04	0.04	0.04	0.04	0.05	0.04	0.04	0.15
TAR (#24)	0.56	0.49	0.58	0.50	0.59	0.56	0.60	0.17
EXPAR (#24)	0.24	0.24	0.23	0.24	0.22	0.22	0.27	0.20
MAR (#24)	0.29	0.31	0.26	0.28	0.31	0.32	0.19	0.71
MSAR (#24)	0.35	0.35	0.34	0.34	0.36	0.36	0.32	0.14
GARCH (#12)	0.09	0.09	0.08	0.08	0.10	0.11	0.07	0.57
TMA (#24)	0.55	0.55	0.54	0.54	0.60	0.59	0.54	0.10
BL (#18)	0.94	0.96	0.93	0.95	0.95	0.97	0.89	0.09
RCA (#18)	0.12	0.14	0.11	0.12	0.13	0.14	0.10	0.47
NLMA (#12)	0.03	0.12	0.03	0.12	0.05	0.11	0.05	1.82
T=500	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.18
TAR (#24)	0.74	0.68	0.75	0.71	0.76	0.74	0.77	0.11
EXPAR (#24)	0.45	0.45	0.42	0.42	0.38	0.38	0.50	0.25
MAR (#24)	0.38	0.39	0.33	0.35	0.41	0.42	0.22	0.88
MSAR (#24)	0.52	0.51	0.50	0.50	0.53	0.50	0.46	0.14
GARCH (#12)	0.14	0.14	0.12	0.12	0.16	0.17	0.08	1.19
TMA (#24)	0.71	0.72	0.71	0.71	0.74	0.74	0.71	0.06
BL (#18)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
RCA (#18)	0.17	0.18	0.15	0.17	0.19	0.21	0.12	0.71
NLMA (#12)	0.03	0.19	0.03	0.17	0.05	0.16	0.06	2.70
T=1000	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.18
TAR (#24)	0.83	0.80	0.83	0.81	0.84	0.83	0.86	0.08
EXPAR (#24)	0.58	0.59	0.55	0.55	0.50	0.50	0.62	0.20
MAR (#24)	0.43	0.44	0.37	0.39	0.47	0.48	0.24	1.04
MSAR (#24)	0.63	0.62	0.62	0.62	0.63	0.60	0.59	0.07
GARCH (#12)	0.19	0.19	0.16	0.16	0.23	0.23	0.09	1.47
TMA (#24)	0.79	0.80	0.80	0.80	0.82	0.82	0.78	0.05
BL (#18)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
RCA (#18)	0.22	0.24	0.18	0.20	0.25	0.26	0.14	0.95
NLMA (#12)	0.03	0.24	0.04	0.22	0.06	0.22	0.06	3.45

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005).

[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.

[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(A) denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations. The significance level is set to $\alpha = 0.05$.

[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.23 Power properties: WHITE test

T=200	A1(+)	A1(−)	A2(+)	A2(−)	A3(+)	A3(−)	N	cv(A)
AR (#8)	0.02	0.02	0.02	0.02	0.03	0.03	0.02	0.85
TAR (#24)	0.40	0.38	0.41	0.39	0.45	0.43	0.43	0.15
EXPAR (#24)	0.11	0.11	0.11	0.12	0.11	0.11	0.13	0.17
MAR (#24)	0.26	0.28	0.22	0.25	0.28	0.29	0.16	0.88
MSAR (#24)	0.36	0.37	0.36	0.35	0.37	0.37	0.33	0.12
GARCH (#12)	0.09	0.09	0.07	0.08	0.10	0.10	0.04	1.24
TMA (#24)	0.45	0.45	0.44	0.45	0.51	0.51	0.46	0.14
BL (#18)	0.93	0.93	0.92	0.92	0.96	0.96	0.84	0.15
RCA (#18)	0.08	0.10	0.07	0.08	0.09	0.10	0.05	1.04
NLMA (#12)	0.02	0.10	0.02	0.09	0.03	0.09	0.02	3.51
T=500	**A1(+)**	**A1(−)**	**A2(+)**	**A2(−)**	**A3(+)**	**A3(−)**	N	**cv(A)**
AR (#8)	0.02	0.02	0.03	0.03	0.03	0.03	0.02	0.69
TAR (#24)	0.61	0.57	0.62	0.59	0.65	0.65	0.66	0.14
EXPAR (#24)	0.30	0.30	0.29	0.29	0.25	0.24	0.36	0.32
MAR (#24)	0.39	0.41	0.34	0.37	0.41	0.44	0.21	1.08
MSAR (#24)	0.57	0.56	0.57	0.55	0.57	0.56	0.54	0.07
GARCH (#12)	0.15	0.15	0.12	0.13	0.16	0.17	0.06	1.85
TMA (#24)	0.66	0.66	0.65	0.66	0.69	0.69	0.67	0.06
BL (#18)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
RCA (#18)	0.13	0.15	0.11	0.13	0.15	0.17	0.07	1.27
NLMA (#12)	0.02	0.18	0.02	0.17	0.04	0.16	0.04	4.10
T=1000	**A1(+)**	**A1(−)**	**A2(+)**	**A2(−)**	**A3(+)**	**A3(−)**	N	**cv(A)**
AR (#8)	0.02	0.02	0.03	0.03	0.03	0.03	0.02	0.56
TAR (#24)	0.72	0.71	0.73	0.73	0.75	0.76	0.77	0.08
EXPAR (#24)	0.48	0.48	0.46	0.46	0.39	0.39	0.52	0.25
MAR (#24)	0.45	0.48	0.40	0.42	0.50	0.51	0.24	1.18
MSAR (#24)	0.70	0.67	0.70	0.66	0.69	0.66	0.67	0.06
GARCH (#12)	0.20	0.20	0.16	0.16	0.24	0.24	0.07	2.46
TMA (#24)	0.75	0.75	0.75	0.75	0.78	0.78	0.77	0.04
BL (#18)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
RCA (#18)	0.17	0.20	0.14	0.16	0.21	0.23	0.08	1.87
NLMA (#12)	0.03	0.26	0.03	0.24	0.06	0.24	0.05	4.37

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005).
[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.
[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(A) denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations. The significance level is set to $\alpha = 0.05$.
[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.24 Power properties: NN test

T=200	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.15
TAR (#24)	0.59	0.54	0.59	0.54	0.61	0.60	0.60	0.12
EXPAR (#24)	0.45	0.45	0.46	0.46	0.46	0.46	0.44	0.03
MAR (#24)	0.14	0.15	0.13	0.14	0.16	0.16	0.10	0.49
MSAR (#24)	0.26	0.26	0.25	0.25	0.25	0.25	0.26	0.05
GARCH (#12)	0.07	0.07	0.07	0.07	0.09	0.08	0.06	0.43
TMA (#24)	0.51	0.51	0.50	0.51	0.55	0.55	0.51	0.10
BL (#18)	0.83	0.82	0.83	0.82	0.87	0.86	0.81	0.08
RCA (#18)	0.09	0.09	0.08	0.08	0.09	0.10	0.07	0.33
NLMA (#12)	0.04	0.07	0.05	0.07	0.05	0.07	0.05	0.65
T=500	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.12
TAR (#24)	0.79	0.78	0.79	0.78	0.81	0.80	0.80	0.04
EXPAR (#24)	0.63	0.63	0.63	0.63	0.63	0.63	0.64	0.02
MAR (#24)	0.17	0.17	0.14	0.15	0.17	0.18	0.11	0.56
MSAR (#24)	0.41	0.41	0.41	0.42	0.40	0.39	0.45	0.12
GARCH (#12)	0.11	0.10	0.09	0.09	0.12	0.12	0.07	0.76
TMA (#24)	0.64	0.64	0.63	0.63	0.67	0.67	0.62	0.08
BL (#18)	0.94	0.92	0.95	0.92	0.96	0.93	0.98	0.06
RCA (#18)	0.10	0.11	0.10	0.10	0.11	0.12	0.08	0.45
NLMA (#12)	0.04	0.09	0.04	0.09	0.05	0.08	0.05	0.99
T=1000	**A1**(+)	**A1**(−)	**A2**(+)	**A2**(−)	**A3**(+)	**A3**(−)	N	**cv(A)**
AR (#8)	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.14
TAR (#24)	0.90	0.90	0.91	0.90	0.92	0.92	0.92	0.02
EXPAR (#24)	0.71	0.71	0.71	0.72	0.71	0.71	0.71	0.01
MAR (#24)	0.17	0.17	0.15	0.16	0.19	0.19	0.12	0.61
MSAR (#24)	0.54	0.55	0.55	0.57	0.53	0.52	0.60	0.13
GARCH (#12)	0.13	0.12	0.11	0.10	0.15	0.14	0.08	0.91
TMA (#24)	0.70	0.69	0.69	0.69	0.72	0.72	0.66	0.09
BL (#18)	0.97	0.94	0.97	0.95	0.97	0.95	0.99	0.04
RCA (#18)	0.11	0.12	0.11	0.11	0.13	0.13	0.09	0.54
NLMA (#12)	0.05	0.10	0.05	0.11	0.05	0.09	0.05	1.22

[a] The lag order p of an AR process is determined by an automatic lag selection procedure discussed in Ng and Perron (2005).

[b] TAR (#24) indicates that we evaluate $K = 24$ different parameter configurations of a TAR model.

[c] Table reports the average rejection frequency (avg) calculated over K parameter configurations of a given time series model and a given distribution of innovations. cv(A) denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations. The significance level is set to $\alpha = 0.05$.

[d] A dark grey area in the legend denotes a moment condition failure of the test statistic, a grey area indicates that moment condition for the test statistic is just met, and a white (no-color) area of the legend denotes that the lowest existing moment is still larger than the test actually requires.

Table 2.25 Ordering of non-linearity test: $T = 1000$

	BDS				**MLQ**				**MQ**				**KEEN**			
	avg(N)	cv(N)	cv(S)	cv(A)	avg(N)	cv(N)	cv(S)	cv(A)	avg(N)	cv(N)	cv(S)	cv(A)	avg(N)	cv(N)	cv(S)	cv(A)
TAR	6	6	6	8	7	7	7	7	8	8	8	6	5	5	5	5
EXPAR	4	6	8	6	7	2	5	4	8	1	6	5	6	7	4	7
MAR	1	2	1	1	2	4	3	2	3	3	2	3	7	5	7	7
MSAR	1	1	1	7	2	2	3	6	3	3	2	5	8	8	8	4
GARCH	1	6	1	1	2	7	2	2	3	8	3	3	6	3	5	5
TMA	6	6	6	6	7	7	7	8	8	8	8	7	5	5	2	5
BL	5	4	4	3	7	6	5	5	8	7	6	6	6	8	8	7
RCAR	1	5	1	7	2	6	3	6	3	7	4	5	6	3	5	2
NLMA	1	6	1	8	3	1	8	6	4	2	7	7	6	3	3	2
median	1	6	1	6	3	6	5	6	4	7	6	5	6	5	5	5

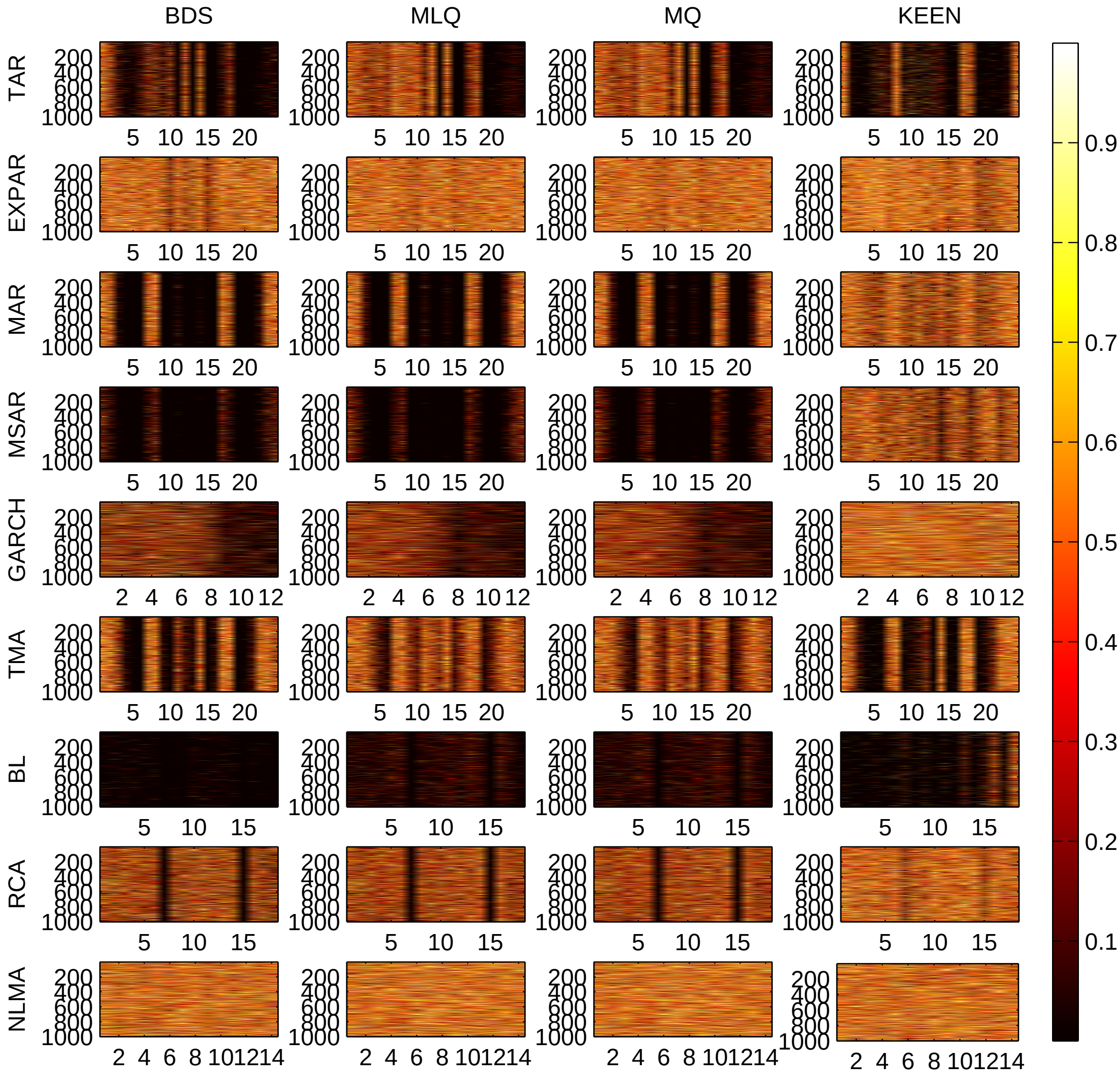
	TSAY				**STAR**				**WHITE**				**NN**			
	avg(N)	cv(N)	cv(S)	cv(A)	avg(N)	cv(N)	cv(S)	cv(A)	avg(N)	cv(N)	cv(S)	cv(A)	avg(N)	cv(N)	cv(S)	cv(A)
TAR	3	4	4	4	2	2	1	2	4	3	2	3	1	1	3	1
EXPAR	5	8	7	8	2	4	2	2	3	5	3	3	1	3	1	1
MAR	6	6	8	8	4	7	5	5	5	8	6	6	8	1	4	4
MSAR	7	7	7	8	6	5	5	1	4	4	4	2	5	6	6	3
GARCH	5	4	7	6	4	2	6	7	8	5	8	8	7	1	4	4
TMA	3	3	1	1	1	2	4	2	2	1	3	3	4	4	5	4
BL	3	3	7	4	2	2	2	2	1	1	1	1	4	5	3	8
RCAR	5	2	7	4	4	4	6	3	8	8	8	8	7	1	2	1
NLMA	2	5	4	3	5	7	6	4	7	8	5	5	8	4	2	1
median	5	4	7	4	4	4	5	2	4	5	4	3	5	3	3	3

[a] avg(N) stands for the average rejection frequency of a given non-linearity test calculated over all parameter configurations of a given non-linear model based on Gaussian (N) innovations, cv(N) stands for the coefficient of variation of a given test calculated over all parameter configurations of a given non-linear model using Gaussian (N) innovations, cv(S) stands for the coefficient of variation of a given test statistic calculated over all parameter configurations of a given non-linear model and all symmetric (S) innovations, cv(A) stands for the coefficient of variation of a given test calculated over all parameter configurations of a given non-linear model and all asymmetric (A) innovations.

[b] median represents a median ordering using individual results.

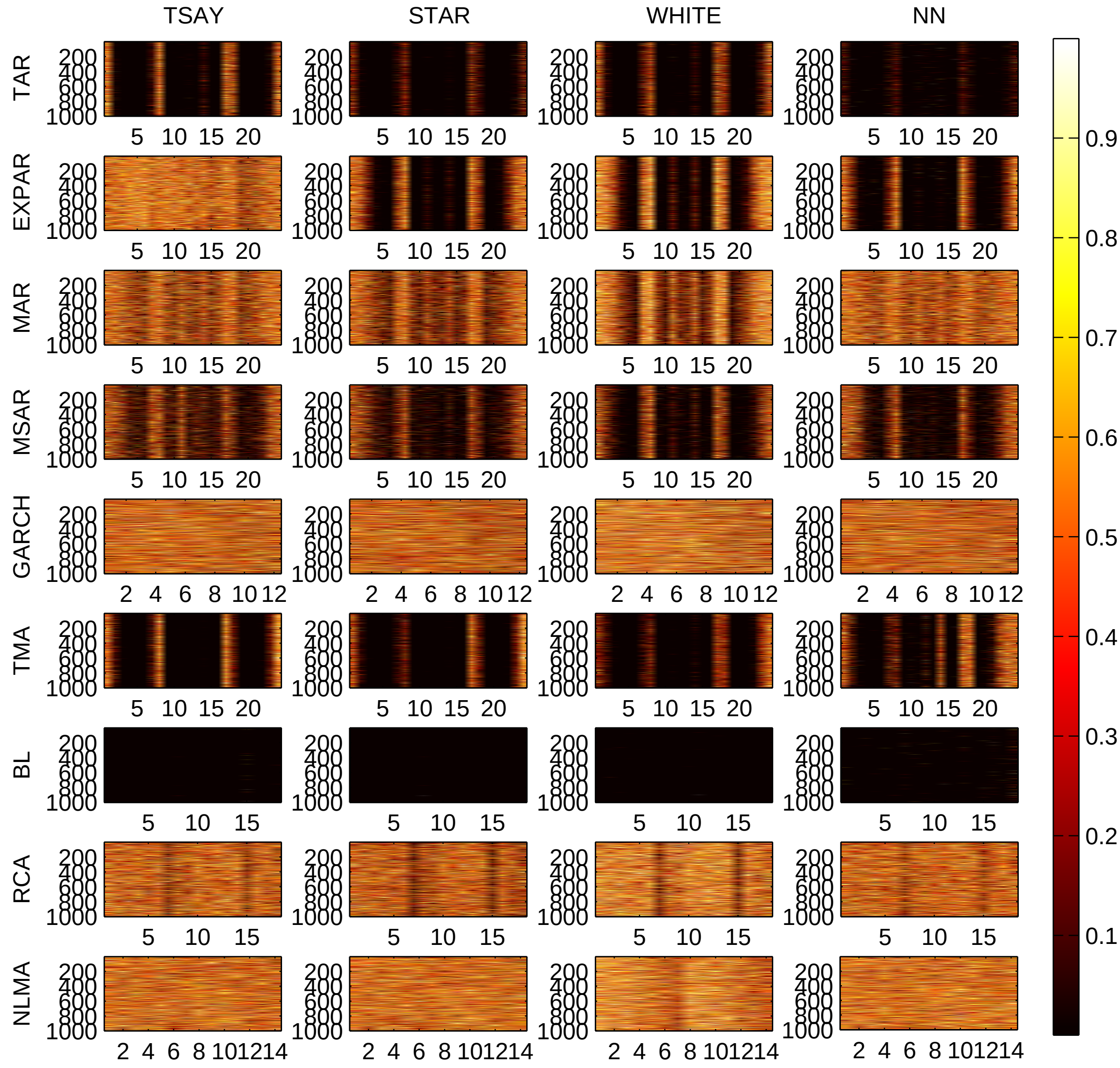
2.7 Appendix B: Figures

Figure 2.7 Power images of non-linearity tests: $T = 1000$ (part 1)



Note that each point of each image represents the estimated p-values of a given non-linearity test for a given parameter configuration (x-axis) and a given Monte Carlo repetition (y-axis).

Figure 2.8 Power images of non-linearity tests: $T = 1000$ (part 2)



Note that each point of each image represents the estimated p-values of a given non-linearity test for a given parameter configuration (x-axis) and a given Monte Carlo repetition (y-axis).

Chapter 3

Testing for Non-linearity Using a Generalized Q Test

"The Earth is round the $p - value < 0.05$."

J. Cohen, statistician

3.1 Introduction

There is a general consensus in the literature that economic time series do exhibit some form of non-linear features. Since a conditional variance has been found to be a common feature of economic variables, the portmanteau Q test proposed by McLeod and Li (1983) has become one of the most widely used tests in practice. The test is based on inspecting the correlation structure of squared residuals. Although the MLQ test has many desirable properties (e.g. the test is very intuitive, easy to calculate, follows a standard limiting distribution, and much more importantly, the test is implemented in many statistical packages), the test suffers from several shortcomings: (i) The MLQ test cannot detect interesting non-linear models (e.g. non-linear moving average models); (ii) The MLQ test exhibits a relatively low power for commonly applied non-linear time series models (e.g. threshold autoregressive and moving average models); (iii) Although the test is originally constructed as a conditional heterescedasticity test, it lacks the discrimination power against more advanced (non-linear/asymmetric) GARCH models developed recently in the literature; (iv) As shown in Chapter 2, the MLQ test suffers from the relatively high sensitivity to the configuration of model parameters and the distribution of innovations.

The main task of this chapter is to propose a generalized version of the portmanteau Q test, which fixes the aforementioned shortcomings but preserves all the desirable properties of the original MLQ test. The idea of using generalized correlations is not entirely new in the literature. For instance, Lawrance and Lewis (1985, 1987) analytically demonstrated the possible usefulness of cross-correlations for detecting non-linearity in time series analysis. However, they only use very specific models (e.g. a random coefficient model), for which the derivation of a cross-correlation structure is analytically tractable. What is more, they focused only on inspecting individual cross-correlations, whereas this chapter focuses on the portmanteau form of the test. A more efficient variance-stabilizing transformation for the Q test is implemented as well, which improves its finite sample properties.

The rest of the chapter is organized as follows. Three Q tests are discussed in Section 3.2. A description of non-linear models and Monte Carlo setup are presented in Section 3.3. Finally, the results of an extensive Monte Carlo analysis are presented in Section 3.4. Section 3.5 is devoted to an empirical application of the proposed Q tests.

3.2 Portmanteau Tests

The idea of inspecting the auto-correlation structure as a tool for detecting non-linearity in time series analysis dates back to the influential work of Granger and Andersen (1978). They show that, provided that $\{X_t : t \in \mathbb{Z}\}$ is a sequence of a linear Gaussian stationary process, it holds that

$$\rho_k(X_t^2) = \rho_k^2(X_t), \quad \text{for} \quad k \in \mathbb{Z},$$

where ρ_k denotes the k-th theoretical auto-correlation coefficient. A simple proof of this relationship can be found in Maravall (1983, p. 69). A departure from the above result might indicate some form of non-linearity and/or non-normality.

Before we proceed to a testing procedure, we state an important assumption about a stochastic process under consideration. The assumption is of the crucial importance for setting the null hypothesis of linearity and for the derivation of a limiting distribution of the test statistic.

Assumption 2 *Let us assume $\{X_t : t \in \mathbb{Z}\}$ is a zero-mean real-valued finite-order ARMA(p,q) model given by*

$$X_t = \xi_1 X_{t-1} + \cdots + \xi_p X_{t-p} + \zeta_1 a_{t-1} + \cdots + \zeta_q a_{t-q} + a_t, \tag{3.1}$$

where $\{a_t : t \in \mathbb{Z}\}$ is a sequence of IID(0,σ^2) model innovations such that $\mathbb{E}(|a_t|^8) < \infty$. Let $\boldsymbol{\beta} = (\xi_1, \ldots, \xi_p, \zeta_1, \ldots, \zeta_q, \sigma)'$ be a $(p+q+1 \times 1)$ parameter vector, which is assumed to be in the interior of the parameter space

$$\begin{aligned}\mathbf{B} = \{\boldsymbol{\beta} \in \mathbb{R}^{p+q} \times \mathbb{R}_{++} : \xi(z) &= 1 - \sum_{i=1}^{p} \xi_i z^i \neq 0 \quad \textit{for all} \quad |z| \leq 1, \\ \zeta(z) &= 1 - \sum_{i=1}^{q} \zeta_i z^i \neq 0 \quad \textit{for all} \quad |z| \leq 1, \\ \xi(z) \quad \textit{and} \quad \zeta(z) & \quad \textit{have no root in common}\}.\end{aligned}$$

$\square$

Provided that all conditions of Assumption 2 are satisfied, then a given stochastic process $\{X_t : t \in \mathbb{Z}\}$ is stationary, an appropriate model is identified and the true parameter vector $\boldsymbol{\beta}$ does not lie on the boundary of the parameter space $\mathbf{B}$. These conditions are sufficient for obtaining consistent estimates of unknown parameters and to ensure the validity of the asymptotic properties of unknown parameters. Note that some authors, for example, Box and Pierce (1970), Li (1992), and Li and Mak (1994), follow a conventional assumption about Gaussian innovations in model (3.1). The advantage of this approach is that all the moment requirements are implicitly satisfied. Another advantage is that uncorrelated Gaussian innovations immediately imply their independence, which is a very convenient property for testing the null hypothesis of linearity using a portmanteau Q test. On the other hand, this assumption might be too restrictive in practice. Therefore, we follow McLeod and Li (1983) and assume IID model innovations with a particular moment restriction.[1]

The theoretical generalized correlation function is defined as

$$\rho_{rs}(k) = \frac{\gamma_{rs}(k)}{\gamma_{rs}(0)} = \frac{\mathbb{E}\left[g_r(a_t)g_s(a_{t-k})\right]}{\sqrt{\mathbb{E}\left[g_r^2(a_t)\right]\mathbb{E}\left[g_s^2(a_t)\right]}}, \tag{3.2}$$

where $g_r(\cdot)$ and $g_s(\cdot)$ are assumed to be real-valued zero-mean continuous functions given by

$$g_r(a) = a^r - \mathbb{E}(a^r), \quad g_s(a) = a^s - \mathbb{E}(a^s), \quad \text{for} \quad r, s \in \{1, 2\}. \tag{3.3}$$

The functional form of $g_r(\cdot)$ and $g_s(\cdot)$ is used to simplify expressions about covariances and variances discussed later in this chapter.

The theoretical generalized covariance term $\gamma_{rs}(k)$ is estimated as follows

$$\hat{\gamma}_{rs}(k) = \frac{1}{T}\sum_{t=k+1}^{T} g_r(\hat{a}_t)g_s(\hat{a}_{t-k}), \quad \text{for} \quad k \in \{1, \ldots, m\}, \tag{3.4}$$

where $\hat{a}_t$ is the estimated residual after applying a linear ARMA filter to the observed series $\{X_t : -\max(p, q), \ldots, 1, \ldots, T\}$. A consistent sample analogue of the $g_r(\cdot)$ and $g_s(\cdot)$ functions is given by

$$g_r(\hat{a}_t) = \hat{a}_t^r - \frac{1}{T}\sum_{t=1}^{T}\hat{a}_t^r, \quad g_s(\hat{a}_t) = \hat{a}_t^s - \frac{1}{T}\sum_{t=1}^{T}\hat{a}_t^s, \quad \text{for} \quad r, s \in \{1, 2\}. \tag{3.5}$$

[1] Alternatively, the null hypothesis of linearity can be specified for innovations being a martingale difference sequence. That means innovations are assumed to be uncorrelated, but not independent. There are, however, at least two difficulties with testing the null in this form. First, it rules out some important non-linear processes such as conditional volatility models often used in finance. Second, the limiting distribution of the Q test can differ substantially from a χ^2 distribution since the variance-covariance matrix of the sample auto-correlations depends on parameters of a given data generating process, see Romano and Thombs (1996), Horowitz et al. (2006), Lobato (2001), Lobato et al. (2002) for details.

Recall that $\hat{\gamma}_{rs}(k)$ is slightly downward biased in small samples.[2] The sample analog of $\gamma_{rs}(0)$ takes the following form

$$\hat{\gamma}_{rs}(0) = \sqrt{\left[\frac{1}{T}\sum_{t=1}^{T} g_r^2(\hat{a}_t)\right]\left[\frac{1}{T}\sum_{t=1}^{T} g_s^2(\hat{a}_t)\right]}, \tag{3.6}$$

where functions $g_r(\hat{a}_t)$ and $g_s(\hat{a}_t)$ are defined in (3.5).

The Q test is then given by

$$Q_{rs}(m) = \sum_{k=1}^{m} (T-k-1)\hat{z}_{rs}(k)^2, \tag{3.7}$$

where $\hat{z}_{rs}(k)$ is a transformed sample generalized correlation coefficient of the form

$$\hat{z}_{rs}(k) = \frac{1}{2}\log\left(\frac{1+\hat{\rho}_{rs}(k)}{1-\hat{\rho}_{rs}(k)}\right), \tag{3.8}$$

where $\hat{\rho}_{rs}(k) = \hat{\gamma}_{rs}(k)/\hat{\gamma}_{rs}(0)$ is the k-th sample correlation coefficient calculated by combining (3.4) and (3.6). For analytical reasons, slightly modified versions of the above defined quantities are also used in this chapter: $\dot{\gamma}_{rs}(k)$ denotes the sample generalized covariance based on observed sequence of innovations $\{a_t : t = 1, \ldots, T\}$. Other quantities, such as $\dot{\gamma}_{rs}(0)$, $\dot{\rho}_{rs}(k)$ or $\dot{z}_{rs}(k)$, are defined analogically. It can be shown that, under the null hypothesis, $\mathbb{E}(\dot{z}_{rs}(k)) = O(T^{-1})$ and $\mathrm{var}(\dot{z}_{rs}(k)) = (T-k-1)^{-1} + O(T^{-2})$, see Johnson et al. (1994, Vol. 2, p. 571). This fact justifies the scaling factor $(T-k-1)$ used in (3.7). As mentioned by Anderson (2003, p. 134), an interesting property of the logarithmic transformation is that the quantity $\dot{z}$ converges to the limiting normal distribution faster than $\dot{\rho}$ in general (see Konishi (1978)). This fact implies that, under the null hypothesis and provided that we directly observe the sequence $\{a_t : t = 1, \ldots, T\}$, $\mathrm{var}(Q_{rs}(m)) = 2\sum_{k=1}^{m}(T-k-1)^2[\mathbb{E}(\dot{z}_{rs}^2(k))]^2 \approx 2m$, since $\mathbb{E}(\dot{z}_{rs}^2(k)) = (T-k-1)^{-1} + O(T^{-2})$ and $\mathrm{cov}(\dot{z}_{rs}^2(i), \dot{z}_{rs}^2(j)) \approx 0$ for integers $i, j \in \{1, \ldots, m\}$, such that $i \neq j$. Therefore, a more complicated variance-stabilizing transformation of the Q test as in Ljung and Box (1978) does not have to be considered. The efficiency of the log-transformation

[2] See also Kendall and Ord (1973, p. 79) for a textbook example. The bias, however, disappears quite quickly, see Wei (1990, p. 19).

was also confirmed in Kwan and Sim (1996a,b) by means of Monte Carlo experiments.

The test specification captures two very well known Q tests: (i) setting $r = s = 1$ leads to a test proposed by Box and Pierce (1970) and modified by Ljung and Box (1978); (ii) setting $r = s = 2$ leads to a test proposed by McLeod and Li (1983). The only difference between our specification of the Q tests and those proposed by other authors is that we use directly a more efficient variance-stabilizing transformation. The main focus of this chapter is on the following three specifications: (i) $r = 1$ and $s = 2$; (ii) $r = 2$ and $s = 1$; and (iii) $r = 2$ and $s = 2$. Note that the Q test can be theoretically defined for any integers $r, s \in \mathbb{N}$, but for high values, extremely high moment conditions must be satisfied. For instance, for $r = s = 2$, the Q_{22} test, an analogy to the MLQ test, requires the existence of the first eight moments to have a valid limiting distribution. Yet, this is in sharp contrast with empirical findings about economic time series, for which the maximum exponent, $\kappa = \sup_{k>0} \mathbb{E}(|X_t|^k) < \infty$, usually lies between 2 and 4, see Jansen and Vries (1991), Loretan and Phillips (1994), or Runde (1997). Anderson and Walker (1964) and Anderson (1991) show that, for linear time series models under the null, the moment condition can be further relaxed provided that one imposes a stronger restriction on the parameters of a data generating process. This implies that the effect of moment failure on the size of the Q test can be minimized to some extent. However, a similar moment restriction is infeasible to obtain for non-linear time series models. In this case, moment condition failure is very likely to inflate the power of the Q test as shown in Chapter 2. For this reason, some authors recommend the use of the Q test with correlations based on absolute residuals rather than squared ones to minimize the moment requirements of the test. It can be shown that the limiting distribution of the Q test is unchanged with the solely requirement of the existence of the first four moments, which is a relatively reasonable assumption, see Pérez and Ruiz (2003) for a discussion. Moreover, Ding et al. (1993) argue that for short-memory models, auto-correlation functions of absolute and squared asset returns are very similar.

Finally, we state two theorems about the limiting properties of the above discussed quantities.

Theorem 1 *Under Assumption 2, the limiting distribution of a vector of sample correlations* $\hat{\mathbf{z}}_{rs} = (\hat{z}_{rs}(1), \ldots, \hat{z}_{rs}(m))'$ *is given by*

$$\sqrt{T}(\hat{\mathbf{z}}_{rs} - \mathbf{z}_{rs}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}),$$

for integers $(r, s) \in \{(1, 2), (2, 1), (2, 2)\}$ *and some integer* $m > 0$. □

Proof. *See Appendix A for a proof.* ■

Theorem 2 *Under Assumption 2, the limiting distribution of the Q tests is given by*

$$Q_{rs}(m) \xrightarrow{d} \chi^2(m),$$

for integers $(r, s) \in \{(1, 2), (2, 1), (2, 2)\}$ *and some integer* $m > 0$. □

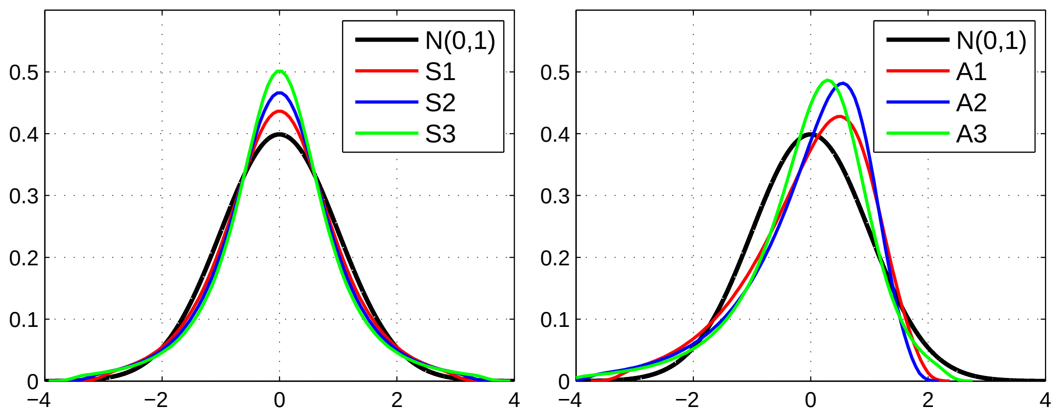
Proof. *See Appendix A for a proof.* ■

3.3 Monte Carlo Setup

3.3.1 Time Series Models

The statistical properties of the proposed generalized Q test are examined using: (i) Linear time series models: an autoregressive (AR) model and a moving average (MA) model; (ii) Non-linear time series models: a threshold (TAR) model, an exponential autoregressive (EXPAR) model, a mixture autoregressive (MAR) model, a Markov switching autoregressive (MSAR) model, generalized autoregressive conditional heteroscedasticity (GARCH) model, a non-linear autoregressive conditional heteroscedasticity (NLGARCH) model, a bilinear (BL) model, a non-linear moving average (NLMA) model, and finally, a threshold moving average (TMA) model. Although the list of non-linear time series models is not definitely exhaustive, we are strongly convinced that all the main classes of non-linear models are included. The models are summarized in Table 3.2. A complete set of model parameters can be found in Table 3.3.

Figure 3.1 Density functions



(a) Symmetric densities (b) Asymmetric densities

Table 3.1 Parameters of a generalized lambda distribution

	λ_1	λ_2	λ_3	λ_4	skewness	kurtosis	moment
S1	0.00000	-1.00000	-0.08000	-0.08000	0.0	6.0	12
S2	0.00000	-0.39791	-0.16000	-0.16000	0.0	11.6	6
S3	0.00000	-1.00000	-0.24000	-0.24000	0.0	126.0	4
A1	0.00000	-0.04306	0.02521	0.09403	-0.9	4.2	10
A2	0.00000	1.00000	-0.00750	-0.03000	-1.5	7.5	33
A3	0.00000	1.00000	-0.10090	-0.18020	-2.0	21.1	5

[a] Note that a standard normal distribution can be also approximated by a generalized lambda distribution with the following parameters: $\lambda_1 = 0$, $\lambda_2 = 0.1975$, $\lambda_3 = \lambda_4 = 0.1349$.

[b] The highest finite moment of a random variable drawn from a given distribution.

The robustness of the proposed Q tests is examined against various distributions of innovations as well. In particular, apart form a Gaussian distribution, which serves as a benchmark for comparison, various model innovations coming from a generalized lambda distributions (GLD) are considered in this chapter, see Randles et al.

(1980). This family provides a wide range of distributions that are easily generated since they are defined in terms of the inverses of the cumulative distribution functions: $F^{-1}(\nu) = \lambda_1 + [\nu^{\lambda_3} - (1-\nu)^{\lambda_4}]/\lambda_2$, for $0 \leq \nu \leq 1$. This chapter considers the following particular distributions: three distributions are symmetric, but leptokurtic, and three distributions are asymmetric, see Table 3.1 and Figure 3.1 for details. All generated innovations are normalized to have zero mean and unit variance. Note that parameters of all non-linear time series models are designed in such a way to satisfy strict stationarity, 4th-moment stationarity and/or invertibility conditions, if necessary. The only exception is S3 specification of model innovations, see Table 3.1, for which the 4th-moment stationarity cannot be reached, at least for some non-linear time series models. It is worth pointing out, however, that additional (4th-moment stationarity and/or invertibility) conditions significantly restrict the parameter space available for Monte Carlo experiments. The effect of 4th-moment stationarity and/or invertibility restrictions imposed on the model parameters is depicted in the graphical form in Figure 3.2, see "Monte Carlo" parameter regions.

3.3.2 Monte Carlo Setup

Originally, T+100 observations are generated in each experiment, but the first 100 of them are discarded in order to eliminate the effect of the initial observations. The number of repetitions of all experiments is set to $R = 1000$. In all experiments, the generated series is filtered by an AR(p) model, where the lag order p is selected by the Bayesian information criterion (BIC) developed by Schwarz (1978). Following the arguments in Ng and Perron (2005), a modified version of the information criterion is used. Ng and Perron (2005) show, based on extensive Monte Carlo experiments, that the best method to give the correct lag order is that with a fixed efficient sample size. Therefore, the selection criterion is defined as follows

$$BIC_l = \log(\hat{\sigma}_l^2) + \frac{l \log(N)}{N},$$
$$\hat{\sigma}_l^2 = \frac{1}{N} \sum_{t=L+1}^{T} \hat{a}_{lt}^2,$$

Table 3.2 List of non-linear models

M1: ARMA models:

$$Y_t = c + \phi Y_{t-1} + \sigma a_t,$$
$$Y_t = c + \theta a_{t-1} + \sigma a_t,$$

M2: A TAR model:

$$Y_t = (c_1 + \phi_1 Y_{t-1} + \sigma_1 a_t) I(Y_{t-1} \leq 0) + (c_2 + \phi_2 Y_{t-1} + \sigma_2 a_t) I(Y_{t-1} > 0),$$

M3: An EXPAR model:

$$Y_t = c + (\phi_1 + (\phi_2 - \phi_1)\exp(-Y_{t-1}^2)) Y_{t-1} + \sigma a_t,$$

M4: A MAR model:

$$Y_t = (c_1 + \phi_1 Y_{t-1} + \sigma_1 a_t) I(S_t = 1) + (c_2 + \phi_2 Y_{t-1} + \sigma_2 a_t) I(S_t = 2),$$

M5: A MSAR model:

$$Y_t = (c_1 + \phi_1 Y_{t-1} + \sigma_1 a_t) I(S_t = 1) + (c_2 + \phi_2 Y_{t-1} + \sigma_2 a_t) I(S_t = 2),$$

M6: A GARCH model:

$$Y_t = c + \phi Y_{t-1} + \epsilon_t, \quad \epsilon_t = a_t \sqrt{h_t},$$
$$h_t = \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1},$$

M7: A NLGARCH model:

$$Y_t = c + \phi Y_{t-1} + \epsilon_t, \quad \epsilon_t = a_t \sqrt{h_t},$$
$$h_t = \omega + \alpha(|\epsilon_{t-1}| + \xi \epsilon_{t-1})^2 + \beta h_{t-1},$$

M8: A TMA model:

$$Y_t = c + \phi_1 a_{t-1} I(Y_{t-1} \leq 0) + \phi_2 a_{t-1} I(Y_{t-1} > 0) + \sigma a_t,$$

M9: A BL model:

$$Y_t = c + \phi Y_{t-1} + \psi Y_{t-1} a_{t-1} + \sigma a_t,$$

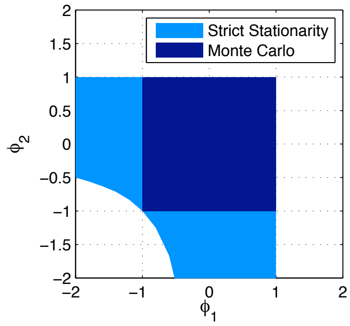
M10: A NLMA model:

$$Y_t = c + \phi a_{t-1} + \psi a_t a_{t-1} + \sigma a_t,$$

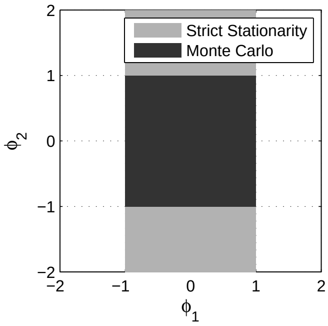
Table 3.3 Model parameters

model	**parameters**
AR, MA	$c = 1$ $\sigma^2 = 1$ $\phi \in \{-0.8, -0.6, -0.4, -0.2, 0.2, 0.4, 0.6, 0.8\}$ $\theta \in \{-0.8, -0.6, -0.4, -0.2, 0.2, 0.4, 0.6, 0.8\}$
TAR, TMA MAR, MSAR	$c_1 = -0.25$, $c_2 = 0.25$ $\sigma_1^2 = 3$, $\sigma_2^2 = 1$ $\sigma^2 = 1$ (for TMA only) $p_{11} = 0.9$, $p_{22} = 0.7$ (for MSAR only) $p_1 = 0.5$ (for MAR only) $(\phi_1, \phi_2) \in \left\{\begin{matrix} (-0.8,-0.8) & (-0.8,-0.5) & (-0.8,-0.2) & (-0.8,0.2) & (-0.8,0.5) \\ (-0.8,0.8) & (-0.5,-0.8) & (-0.5,-0.5) & (-0.5,0.5) & (-0.5,0.8) \\ (-0.2,-0.8) & (-0.2,0.8) & (0.2,-0.8) & (0.2,0.8) & (0.5,-0.8) \\ (0.5,-0.5) & (0.5,0.5) & (0.5,0.8) & (0.8,-0.8) & (0.8,-0.5) \\ (0.8,-0.2) & (0.8,0.2) & (0.8,0.5) & (0.8,0.8) & \end{matrix}\right\}$
EXPAR	$c = 1$ $\sigma^2 = 1$ $(\phi_1, \phi_2) \in \left\{\begin{matrix} (-0.8,-0.8) & (-0.8,-0.5) & (-0.8,-0.2) & (-0.8,0.2) & (-0.8,0.5) \\ (-0.8,0.8) & (-0.5,-0.8) & (-0.5,-0.5) & (-0.5,0.5) & (-0.5,0.8) \\ (-0.2,-0.8) & (-0.2,0.8) & (0.2,-0.8) & (0.2,0.8) & (0.5,-0.8) \\ (0.5,-0.5) & (0.5,0.5) & (0.5,0.8) & (0.8,-0.8) & (0.8,-0.5) \\ (0.8,-0.2) & (0.8,0.2) & (0.8,0.5) & (0.8,0.8) & \end{matrix}\right\}$
GARCH	$c = 1$ $\phi = 0.5$ $\sigma^2 = 1$ $\xi = -0.5$ (for NLGARCH only) $(\alpha, \beta) \in \left\{\begin{matrix} (0.05,0.3) & (0.05,0.4) & (0.05,0.5) & (0.05,0.6) & (0.05,0.7) \\ (0.05,0.8) & (0.10,0.3) & (0.10,0.4) & (0.10,0.5) & (0.10,0.6) \\ (0.10,0.7) & (0.10,0.8) & (0.15,0.3) & (0.15,0.4) & (0.15,0.5) \\ (0.15,0.6) & (0.15,0.7) & & & \end{matrix}\right\}$
BL	$c = 1$ $\sigma^2 = 1$ $(\phi, \psi) \in \left\{\begin{matrix} (-0.8,-0.2) & (-0.6,-0.2) & (-0.4,-0.2) & (-0.2,-0.2) & (-0.2,0.2) \\ (-0.4,0.2) & (-0.6,0.2) & (-0.6,0.4) & (-0.8,0.2) & (0.2,-0.2) \\ (0.2,0.2) & (0.4,-0.2) & (0.4,0.2) & (0.6,-0.2) & (0.6,-0.4) \\ (0.6,0.2) & (0.8,-0.2) & (0.8,0.2) & & \end{matrix}\right\}$
NLMA	$c = 1$ $\sigma^2 = 4$ $(\phi, \psi) \in \left\{\begin{matrix} (-0.2,0.2) & (-0.2,0.4) & (-0.4,0.2) & (-0.4,0.4) & (-0.6,0.2) \\ (-0.6,0.4) & (0.2,0.2) & (0.2,0.4) & (0.4,0.2) & (0.4,0.4) \\ (0.6,0.2) & (0.6,0.4) & & & \end{matrix}\right\}$

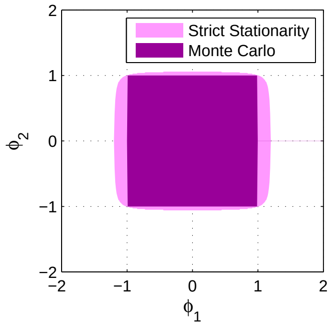
Figure 3.2 Stationarity regions of non-linear models



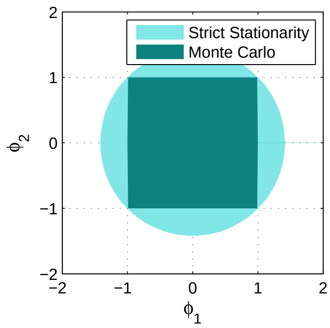
(a) TAR(2;1)



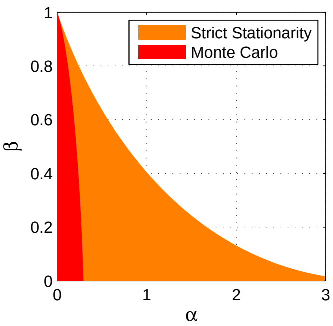
(b) EXPAR(1)



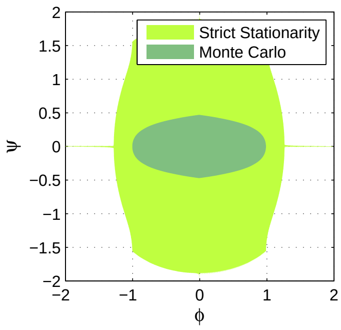
(c) MSAR(2;1)



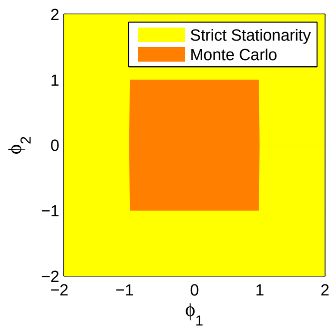
(d) MAR(2;1)



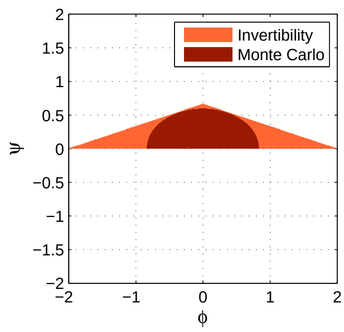
(e) GARCH(1,1), NL-GARCH(1,1)



(f) BL(1,0,1,1)



(g) TMA(2;1)



(h) NLMA(1)

* Strict Stationarity regions are calculated based on the assumption that $a \sim$ NID(0,1), whereas Monte Carlo regions are calculated based on the intersection of 4-th moment stationarity and/or invertibility conditions for S1, S2, A1, A2, A3 specifications of distributions of model innovations. The only exception is S3 specification of model innovations for which 4th-moment stationarity cannot be reached, at least for some non-linear time series models.

where $l \in \{1, \dots, L\}$, $N = T - L$, T is the sample size and the maximum lag order is constraint according to $L = [8(T/100)^{0.25}]$. The lag order for an AR(p) model is estimated by the following simple rule $\hat{p} = \min_{l \in \{1,\dots,L\}}(BIC_l)$. Finally, the sample size is set to $T \in \{200, 500, 1000\}$.

We also report the Q tests with the automatically selected lag order based on Escanciano and Lobato (2009).[3] The estimated lag order is selected by maximizing the following objective function

$$Q_{rs}(l)^* = Q_{rs}(l) - q_{rs}(l),$$
$$q_{rs}(l) = \begin{cases} p\log(N) & \text{if} \quad \max_{j\in\{1,\dots,L\}} |\hat{\rho}_{rs}(j)| \leq \sqrt{c\log(N)/N}, \\ \\ 2p & \text{if} \quad \max_{j\in\{1,\dots,L\}} |\hat{\rho}_{rs}(j)| > \sqrt{c\log(N)/N}, \end{cases}$$

where the constant $c = 2.4$ is recommended by Escanciano and Lobato (2009). Finally, the lag order of the Q tests is determined by the simple rule $\hat{m} = \max_{l\in\{1,\dots,L\}}(Q_{rs}(l)^*)$.

3.4 Monte Carlo Results

3.4.1 Size and Statistical Properties

The average rejection frequency is calculated for each Q test as follows

$$\mathcal{P}_i = \frac{1}{R}\sum_{j=1}^{R} I(\hat{\alpha}_{ij} \leq \alpha),$$

where $i \in \{1, \dots, K\}$ denotes the i-th particular parameter configuration of a given time series model, R is the number of repetitions set to $R = 1000$, $I(\cdot)$ is a standard indicator function taking 1 if $\hat{\alpha}_{ij} \leq \alpha$ and 0 otherwise, α represents the statistical significance level set to 0.05, and $\hat{\alpha}_{ij}$ is the estimated p-value of the Q test for the ith-parameter configuration and jth-Monte Carlo replication. Subsequently, three

[3]Recall that a given procedure is proposed for the realizations of stochastic processes and not the filtered ones. The additional simulations show, however, that the procedure may be adopted for filtered processes as well.

quantities for each Q test are presented in the following tables: "*avg*" stands for the average rejection frequency of a given Q test over all parameter configurations of a given model, "*min*" and "*max*" indicate the minimum and maximum of the average rejection frequencies of the test over all parameter configurations of a given model. Formally, the statistics are defined as follows

$$avg = \frac{1}{K}\sum_{i=1}^{K}\mathcal{P}_i,$$

$$min = \min_{i\in\{1,\ldots,K\}}(\mathcal{P}_i),$$

$$max = \max_{i\in\{1,\ldots,K\}}(\mathcal{P}_i),$$

where K is the number of parameter configurations for a given time series model inspected by Monte Carlo experiments: $K = 8$ for AR and MA models, $K = 24$ for a TAR, MAR, MSAR, EXPAR, TMA models, $K = 17$ for GARCH and NLGARCH models, $K = 18$ for a BL model, and $K = 12$ for a NLMA model.

Since the Q tests proposed above are new versions of a standard Q test, it is important to check how a good approximation the limiting χ^2 distribution is for the tests. Provided that a χ^2 distribution is a valid limiting distribution, then $\mathbb{E}(Q(m)) \approx m$, $\mathrm{var}(Q(m)) \approx 2m$, and the variance-mean ratio $\mathrm{var}(Q(m))/\mathbb{E}(Q(m)) \approx 2$ as the sample $T \to \infty$ and $m/T \to 0$. The Monte Carlo results of the proposed Q tests for AR(1) and MA(1) processes and fixed lag order $m \in \{5, 10, 15\}$ can be found in Table 3.5. The table shows that the finite sample properties of the Q tests are in line with the limiting distribution, even for relatively small samples and different lag orders m, the average value of the Q tests is very close to m and the variance to $2m$. Figures 3.4 – 3.6 depict the χ^2 density function accompanied by the lower and upper bound of the smoothed empirical densities of the Q tests of AR and MA processes for the sample size $T = 200$ and the lag orders $m \in \{5, 10, 15\}$.[4] The figure clearly confirms that the χ^2 distribution is a valid distribution for all the Q tests even in relatively small samples. Additionally, the figure also clearly shows that the

[4]A simple reference bandwidth is used for smoothing the empirical density functions of the Q tests.

χ^2 distribution is a better approximation for the Q tests based on cross-correlations (i.e. Q_{12} and Q_{21}) as compared to the Q test based on auto-correlations (i.e. Q_{22}).

Table 3.6 illustrates that the Q tests have good size properties for both AR and MA processes. Even for a relatively small sample $T = 200$, the average rejection frequency is close to the nominal level 0.05. In addition, other descriptive statistics (min and max) indicate that the behaviour of the Q tests is good, regardless of the specification of the lag order m or the sample size T. For instance, the minimum value of the individual average rejection frequencies for both AR and MA models, denoted as min, is not smaller than 0.03 and the maximum value, denoted as max, does not exceed 0.08.

3.4.2 Power Results

Regime switching models: It can be concluded from Tables 3.7 – 3.9 that Q_{12} and Q_{21} tests significantly outperform the results of the normally used Q_{22} test for a TAR model. From detailed records, it can be concluded that all the Q tests have very good power provided that parameters of a TAR model lie in a specific range, $|\phi_2 - \phi_1| \geq 1$, with rather opposite signs and a probability of a (lower) regime $\pi \in (0.3, 0.7)$. The second result is not very surprising since if $\pi \to 0$ or $\pi \to 1$, one regime dominates the other and the process can be relatively well approximated by a simple AR(p) model, which negatively affects the power of the Q tests. Surprising results are obtained for an EXPAR model, where none of the proposed Q tests exhibit any reasonable power even in large samples.

Surprisingly, rather different results are obtained for a MAR model. In this case, a switching mechanism is independent of any DGP parameter, and therefore fully under control. We set a probability of a lower regime to $\pi = 0.5$ only for simplicity of Monte Carlo experiments. MAR models are detected very efficiently by the Q_{22} tests but only if $|\phi_2 - \phi_1| \geq 1$. The Q_{12} test is not informative for any parameter configuration under consideration, and the Q_{21} only for just a few parameter configurations. Very similar results are obtained for a MSAR model as well.

Conditional volatility models: It can be clearly concluded from Tables 3.7 – 3.9 that the Q_{22} test is very useful in detecting conditional volatility. In the case of a simple GARCH model, the Q_{12} and Q_{21} tests are not informative if model innovations are drawn from a Gaussian (symmetric) distribution. In the case of a NLGARCH model, both the Q_{21} and Q_{22} tests exhibit a good power. Again, the Q_{12} is not at all informative. However, note that the power results of the Q_{21} test depend on a combination of asymmetry of innovations and a non-linear component. Provided that model innovations are negatively skewed and a non-linear component exhibits negative asymmetry as well (as in the Monte Carlo setup in this chapter), then the power of the Q_{21} and Q_{22} tests is around 0.7 in the sample $T = 1000$, see Table 3.12.

Other models: In the case of a NLMA model, Tables 3.7 – 3.9 show that both the Q_{12} and Q_{22} tests have no power, whereas the Q_{21} exhibits a very good power, regardless of the parameter specification of a NLMA model. Completely opposite results come from a BL model, where the only non-informative test is the Q_{21} test. The Q_{12} and Q_{22} tests have a very good power regardless the model parameters. Very similar results are obtained for a TMA model, where the only informative test is the Q_{12} test. However, it is worth noting that the above results for NLMA and BL models are to some extent model dependent since these two classes are extremely flexible.

3.4.3 Sensitivity Analysis

Parameters: Since the original MLQ test suffers from a high power variation, special attention is paid to the sensitivity of the power of the new Q tests to the parameter configuration of data generating processes. For better understanding, the individual Monte Carlo results are presented in the form of graphical images. Each point depicted in a given graphical image represents the estimated p-value of a given Q test for a given parameter configuration of a given time series model (x-axis) and a particular Monte Carlo replication (y-axis). For example, in the case of a TAR model, each graphical image summarizes the results over 24000 replications ($K = 24$

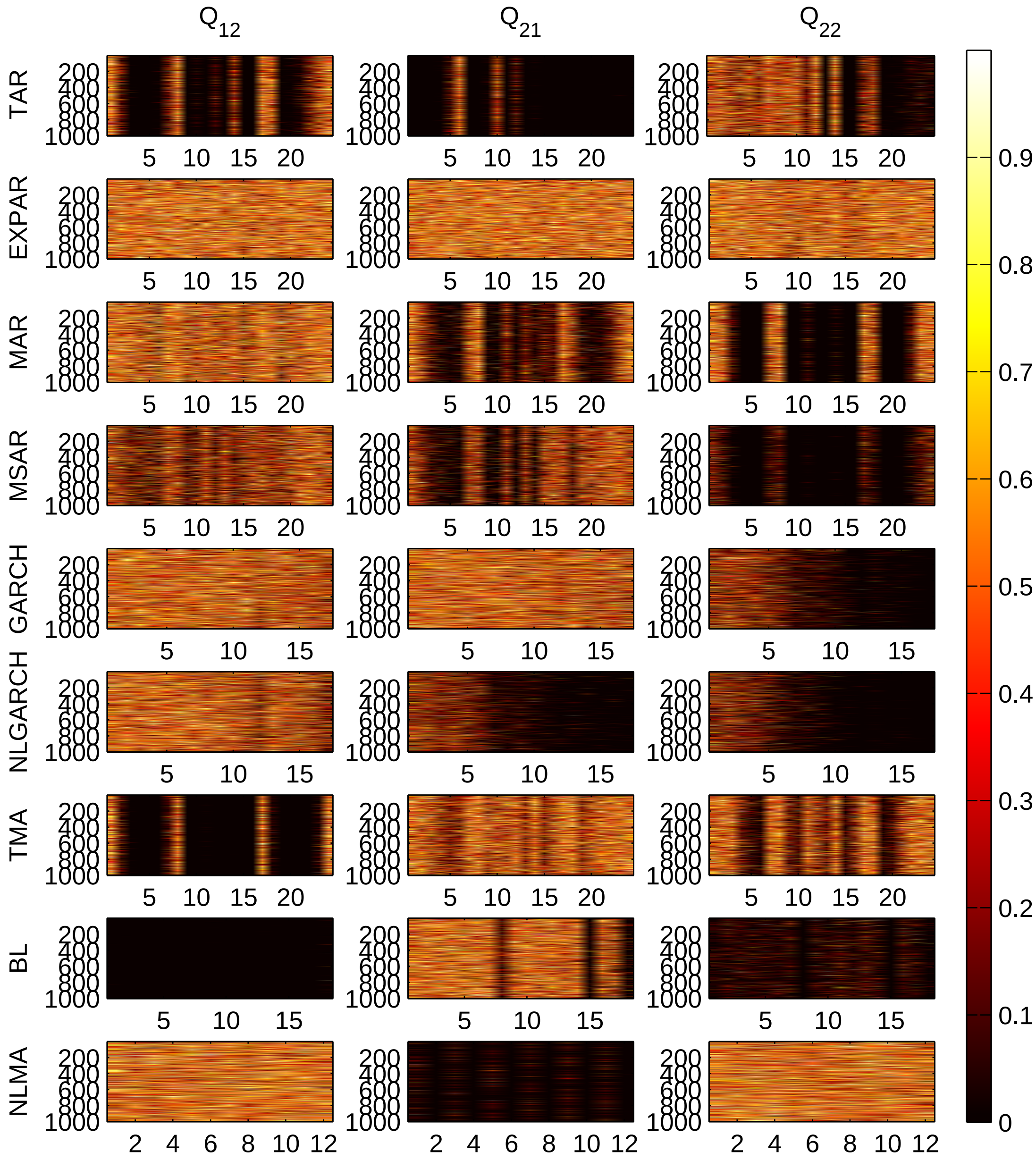
parameter configurations of a TAR model and $R = 1000$ replications). A color range (from black to white) is used to explicitly indicate the different magnitude of the statistical significance of the Q tests. The results, based on Gaussian innovations and the sample size $T = 1000$, can be found in Figure 3.3.

For example, from the results given by a NLMA model it can be seen that all the p-values of the Q_{21} test are less then the significance level 0.05 and the results are not sensitive to any parameter configuration of a NLMA model at all, whereas the results of the Q_{12} and Q_{22} tests are statistically insignificant, but not sensitive either. Another relevant example, illustrating the benefit of using the graphical images, is related to a GARCH model. The p-values of the Q_{12} and Q_{21} tests are statistically very insignificant, whereas the p-values of the standard Q_{22} test are statistically significant, but merely for the second half of parameter configuration (i.e. $\alpha \geq 0.1$ and $\beta \geq 0.3$). From the figure it can be also concluded that the power of the Q tests is quite sensitive for regime switching models such as TAR, MAR, and MSAR models, whereas for models such as BL and NLMA, the stability of the power properties of the Q test is excellent.

Innovations: The power properties of the proposed Q tests are examined against the following classes of distributions of innovations: a Gaussian distribution (N), symmetric but leptokurtic (S), and asymmetric (A) innovations. The results are summarized in a coherent way in Table 3.4. The notation is as follows: "■" indicates that a given Q test exhibits a good power for a given non-linear process (i.e. $avg \geq 0.5$), whereas "□" indicates only a reasonable power of the Q test (i.e. $0.2 < avg < 0.5$,) and "no-square" indicates almost no power of a given Q test (i.e. $avg \leq 0.2$). The table is reproduced from the Monte Carlo results based on $T = 1000$ observations.

The table concludes the following: (i) Inspecting the cross-correlation structure can be useful tool supplementing the results from the auto-correlation structure; (ii) The correlation structure of many non-linear models is rather different, which

Figure 3.3 Power images of the Q tests: $T = 1000$, $R = 1000$, NID(0,1)



* Each point depicted in the graphical image represents the estimated p-value of a given Q test for a given parameter configuration of a given time series model (x-axis) and a given Monte Carlo replication (y-axis). The results of the Q tests are based on the automatically selected lag order m.

Table 3.4 Power properties of the Q tests: $T = 1000$, $R = 1000$

	TAR			**TMA**			**MAR**			**MSAR**		
test	N	S	A	N	S	A	N	S	A	N	S	A
Q_{12}	■	■	■	■	■	■				□	□	□
Q_{21}	■	■	■		□	□	□	□	□	□	□	□
Q_{22}	□	■	□	■	■	□	■	■	■	■	■	■
	GARCH			**NLGARCH**			**BL**			**NLMA**		
test	N	S	A	N	S	A	N	S	A	N	S	A
Q_{12}		□					■	■	■			
Q_{21}			□	■	■	■	□	□	■	■	■	■
Q_{22}	■	■	■	■	■	■	■	■	■			□

enables us to use the proposed Q tests for some preliminary discrimination among various classes of non-linear time series models. For example, it is easy to see a discrimination power for TAR and GARCH models, or BL and NLMA models.

3.5 Empirical Application

In this section, the proposed Q tests are applied to a set of 22 financial time series: 5 exchange rate time series; 5 interest rate time series; 6 commodity time series; and finally 6 equity indices. We use average weakly returns from 1980 to 2010 (i.e. 1620 observations).[5] A complete description of time series can be found in Table 3.13.

In this exercise, we are particularly interested in whether or not a simple GARCH model, often applied in finance, is an adequate model. Put differently, we are interested whether or not there are other statistically significant non-linear features, apart from a conditional volatility, in selected financial returns. This question is of much practical importance in finance for asset pricing and value-at-risk management.

The results are presented in Table 3.14. Recall that a simple GARCH model might be considered as appropriate, provided that the null hypothesis of linearity is rejected

[5]Weakly time series seem to serve as a good compromise between daily returns, which are too noisy and contaminated by jumps, and monthly returns, which are too aggregated.

by only the Q_{22} test, whereas the Q_{12} and Q_{21} tests are not informative at all. The null hypothesis about linearity is clearly rejected by the Q_{22} test for all 22 asset returns at the significance level 0.01. This result is not very surprising since the conditional volatility is a common feature of asset returns. However, the null hypothesis of linearity is rejected also in 14 out of 22 cases by at least one of the cross-correlation Q tests (i.e. Q_{12} and/or Q_{21}) at the nominal level 0.05. Put differently, almost 65 % of asset returns exhibit stochastic features incompatible with a simple GARCH process.

3.6 Conclusion

It has been demonstrated that inspecting generalized residual correlations (i.e. autocorrelations and cross-correlations) can be an useful, yet very simple, tool both for testing for non-linearity. The proposed Q tests have very good size and power properties, and the limiting χ^2 distribution is a good approximation, regardless of the sample size T and the lag order m. The Monte Carlo results suggest that the proposed Q tests fix two main shortcomings of the Mcleod and Li Q (MLQ) test often used in the literature: (i) the tests are capable to capture some interesting non-linear models, for which the original MLQ test completely fails (e.g. a NLMA model). The Q tests significantly improves the power for some other non-linear models (e.g. a TAR and TMA), for which the original MLQ test does not work very well. What is more, the power of the new Q tests is even higher as compared to the BDS and NN tests, two recommended tests from Chapter 2; (ii) the proposed Q tests can be used for discrimination between simple and more complicated (non-linear/asymmetric) GARCH models as well. It can be concluded that the proposed Q test may serve as a valuable alternative to the non-linearity tests discussed in Chapter 2.

The empirical results indicate that almost 65 % of asset returns exhibit stochastic features incompatible with a simple GARCH process. Our results may be directly employed in finance for risk management. For example, according to Basel banking regulations (see Jorion (2007), among others), commercial banks are required to measure market risk of their asset portfolios and to hold capital in proportion to their risk position. As a result, banks calculating their risk position using simple GARCH models may systematically over or underestimate downside risk (the left-hand tail of a marginal distribution), which can have serious implications for stability of the financial system.

3.7 Appendix A: Proofs

3.7.1 Useful Theorems

Theorem 3 *Let Assumption 2 be satisfied, then the LS estimate $\hat{\boldsymbol{\beta}}$ has the following properties: (i) $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$; (ii) $\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{V})$.* □

Proof. *See a proof to Theorem 8.4.1 in Fuller (1996, p. 432). Yao and Brockwell (2006) obtained the same results for the ML estimator of ARMA parameters.* ■

Theorem 4 *Let $\{Z_t : t \in \mathbb{Z}\}$ be a sequence of IID(0,σ^2) innovations such that $\mathbb{E}(|Z_t|^4) < \infty$, then for some integer $m > 0$ we have that $\sqrt{T}(\dot{\boldsymbol{\rho}} - \boldsymbol{\rho}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$, where $\boldsymbol{\rho} = (\rho(1), \ldots, \rho(m))'$ denotes a vector of auto-correlations and $\dot{\boldsymbol{\rho}} = (\dot{\rho}(1), \ldots, \dot{\rho}(m))'$ denotes a vector of sample auto-correlations.* □

Proof. *See a proof to Theorem 7.2.1 in Brockwell and Davis (1991, p. 221) with a restriction $X_t = Z_t$.* ■

Theorem 5 *Let $\{X_t, Y_t : t \in \mathbb{Z}\}$ be a sequence of pairs of variables. If $|X_t - Y_t| \xrightarrow{p} 0$ and $Y_t \xrightarrow{d} Y$, then $X_t \xrightarrow{d} Y$ as well. That is, the limiting distribution of X_t exists and is the same as that of Y.* □

Proof. *See a proof to 2c.4(ix) result in Rao (1973, p. 122).* ■

Theorem 6 *Let $X_t \xrightarrow{d} X$ and $Y_t \xrightarrow{p} c$, where c is a finite constant different from 0. Then it holds that $X_t/Y_t \xrightarrow{d} X/c$.* □

Proof. *See a proof to Slutsky Theorem in Serfling (1980, p. 19).* ■

Theorem 7 *Let $\{X_t : t \in \mathbb{Z}\}$ and X be random variables defined on a probability space and let g be a Borel-measurable function defined on $\mathbb{R}$. Suppose that g is continuous with probability 1. Then $X_t \xrightarrow{p} X$ implies that $g(X_t) \xrightarrow{p} g(X)$.* □

Proof. *See a proof in Serfling (1980, p. 24).* ■

3.7.2 Proof of Theorem 1

Let $\boldsymbol{\beta} = (\xi_1, \ldots, \xi_p, \zeta_1, \ldots, \zeta_q, \sigma)'$ denote a vector of true model parameters and let $\hat{\boldsymbol{\beta}}$ denote the LS and/or ML estimates. Expanding a sample generalized covariance $\hat{\gamma}_{rs}(k)$ by a first-order Taylor expansion gives

$$\hat{\gamma}_{rs}(k) = \gamma_{rs}(k) + \sum_i (\hat{\beta}_i - \beta) \frac{\partial \dot{\gamma}_{rs}(k)}{\partial \beta_i} + O_p(T^{-1}),$$

for the lag order $k \in \{1, \ldots, m\}$ and $(r, s) = \{(1, 2), (2, 1), (2, 2)\}$. It concludes from Theorem 3 and Proposition 1 that $(\hat{\beta}_i - \beta) = O_p(T^{-1/2})$ and $\partial \dot{\gamma}_{rs}(k)/\partial \beta_i = O_p(T^{-1/2})$, which immediately implies that the product of these two stochastic components is $O_p(T^{-1})$. Then it holds that

$$\hat{\gamma}_{rs}(k) = \gamma_{rs}(k) + O_p(T^{-1}).$$

Moreover, it concludes from Proposition 2 that $\hat{\gamma}_{rs}(0) \xrightarrow{p} \gamma_{rs}(0)$ for given integers r, s. Then Theorem 6 implies that $\gamma_{rs}(0)$ can be considered as a normalizing constant having no effect on the limiting distribution of $\hat{\rho}_{rs}(k)$. Then it holds that

$$\hat{\rho}_{rs}(k) = \frac{\hat{\gamma}_{rs}(k)}{\gamma_{rs}(0)} + O_p(T^{-1}).$$

See also McLeod and Li (1983, p. 271) or Li and Mak (1994, p. 629–631) for a discussion.

Under Assumption 2 and using a slightly modified Theorem 4, it can be shown that a vector of sample correlations is given by

$$\sqrt{T}(\dot{\boldsymbol{\rho}}_{rs} - \boldsymbol{\rho}_{rs}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}).$$

Note that the modification of Theorem 4 lies in the requirement of the existence of the first eighth moments of the random variable a to ensure the validity of the above limiting result. This condition is a part of Assumption 2. Combining results from Theorem 4 and Theorem 5, it easy to show that the limiting distribution of a vector of the estimated correlations is given by

$$\sqrt{T}(\hat{\boldsymbol{\rho}}_{rs} - \boldsymbol{\rho}_{rs}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}),$$

since $|\hat{\boldsymbol{\rho}}_{rs} - \dot{\boldsymbol{\rho}}_{rs}| \xrightarrow{p} \mathbf{0}$ due to the fact that estimated residuals/parameters are consistent, which directly follows from Theorem 3.

Under Assumption 2, it can be shown that a first-order Taylor expansion of $\hat{z}_{rs}(k)$ around $\rho_{rs}(k)$ gives us the following expression

$$\hat{z}_{rs}(k) = \hat{\rho}_{rs}(k) + o_p(T^{-1/2}),$$

for the lag order $k \in \{1, \ldots, m\}$ and integer $m > 0$. It is now easy to see that vectors $\hat{\mathbf{z}}_{rs} = (z_{rs}(1), \ldots, z_{rs}(m))'$ and $\hat{\boldsymbol{\rho}}_{rs} = (\hat{\rho}_{rs}(1), \ldots, \hat{\rho}_{rs}(m))'$ have the same limiting distribution given by

$$\sqrt{T}(\hat{\mathbf{z}}_{rs} - \mathbf{z}_{rs}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}).$$

$\blacksquare$

Proposition 1 *Let Assumption 2 be satisfied. Let us define $\dot{\gamma}_{rs}(k)$ as follows*

$$\dot{\gamma}_{rs}(k) = \frac{1}{T} \sum_{t=k+1}^{T} g_r(a_t) g_s(a_{t-k}),$$

for the lag order $k \in \{1, \ldots, m\}$ and some integers $m \geq 1$, $r > 0$, $s > 0$, and $g_r(\cdot)$ and $g_s(\cdot)$ functions are defined as follows

$$g_r(a_t) = a_t^r - \frac{1}{T} \sum_{t=1}^{T} a_t^r, \quad g_s(a_t) = a_t^s - \frac{1}{T} \sum_{t=1}^{T} a_t^s.$$

Let $\boldsymbol{\beta}$ be a vector of ARMA parameters from (3.1), then it holds that $\dfrac{\partial \dot{\gamma}_{rs}(k)}{\partial \beta_i} = O_p(T^{-1/2})$, for all $\beta_i \in \boldsymbol{\beta}$. $\square$

Proof.

$$\begin{aligned}
\frac{\partial \dot{\gamma}_{rs}(k)}{\partial \beta_i} &= \frac{1}{T} \sum_{t=k+1}^{T} \left(\frac{\partial g_r(a_t)}{\partial \beta_i} \right) g_s(a_{t-k}) + \frac{1}{T} \sum_{t=k+1}^{T} g_r(a_t) \left(\frac{\partial g_s(a_{t-k})}{\partial \beta_i} \right), \\
&= O_p(T^{-1/2}) + O_p(T^{-1/2}), \\
&= O_p(T^{-1/2}),
\end{aligned}$$

since $g_r(\cdot)$ and $g_s(\cdot)$ are continuous functions in $\boldsymbol{\beta}$ and a sample average of (stationary) random variables is $O_p(T^{-1/2})$, see Jiang (2010, Ch. 3). $\blacksquare$

Proposition 2 *Let Assumption 2 with $\hat{\gamma}_{rs}(0)$ given by*

$$\hat{\gamma}_{rs}(0) = \sqrt{\left[\frac{1}{T}\sum_{t=1}^{T} g_r^2(\hat{a}_t)\right]\left[\frac{1}{T}\sum_{t=1}^{T} g_s^2(\hat{a}_t)\right]}, \tag{3.9}$$

where $\hat{a}_t$ is the estimated residual from model in (2), and functions $g_r(\hat{a}_t)$ and $g_s(\hat{a}_t)$ are defined in (3.5). Then it holds that $\hat{\gamma}_{rs}(0) \xrightarrow{p} \gamma_{rs}(0)$. □

Proof. *Theorem 7 implies that $\hat{\gamma}_{rs}(0) \xrightarrow{p} \gamma_{rs}(0)$, provided that*

$$\frac{1}{T}\sum_{t=1}^{T} g_i^2(\hat{a}_t) \xrightarrow{p} \mathbb{E}\left(g_i^2(a_t)\right), \quad \text{for} \quad i \in \{r, s\}.$$

Since both $g_r(\cdot)$ and $g_s(\cdot)$ functions are equivalent for $r = s$, it is fully sufficient to base the proof on one of these two quantities. In order to simplify the proof, the following notation is used

$$\frac{1}{T}\sum_{t=1}^{T} g_r^2(a_t) = \frac{1}{T}\sum_{t=1}^{T} a_t^{2r} - \left(\frac{1}{T}\sum_{t=1}^{T} a_t^r\right)^2 = M_2 - M_1^2.$$

Following arguments in a proof of Theorem 8.4.1 in Fuller (1996, p. 432), the proof consists of the following two steps:

(i) It directly follows from Strong Law of Large Numbers (SLLN), see Theorem B in Serfling (1980, p. 24), that both $M_1 \xrightarrow{as} \mathbb{E}(a_t^r)$ and $M_2 \xrightarrow{as} \mathbb{E}(a_t^{2r})$, which implies that

$$\frac{1}{T}\sum_{t=1}^{T} g_r^2(a_t) \xrightarrow{as} \mathbb{E}\left(g_r^2(a_t)\right),$$

for any integer $r > 0$.

(ii) It follows from Theorem 3 that $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$, which implies that $a_t(\hat{\boldsymbol{\beta}}) \equiv \hat{a}_t \xrightarrow{p} a_t$. Then, since $g_r(\cdot)$ is a continuous function in $\boldsymbol{\beta}$, it holds that $g_r(\hat{a}_t) \xrightarrow{p} g_r(a_t)$ for any integer $r > 0$.

Combining the results from (i) and (ii), it follows that

$$\frac{1}{T}\sum_{t=1}^{T} g_r^2(\hat{a}_t) \xrightarrow{p} \mathbb{E}\left(g_r^2(a_t)\right),$$

for any integer $r > 0$. This completes the proof. ■

3.7.3 Proof of Theorem 2

Note that the proposed Q tests can be written into the form of a quadratic function given by

$$Q_{rs}(m) = \sum_{k=1}^{m} (T - k - 1)\hat{z}_{rs}^2(k) = \hat{\mathbf{z}}_{rs}' \mathbf{C} \hat{\mathbf{z}}_{rs},$$

where $\hat{\mathbf{z}}_{rs} = (\hat{z}_{rs}(1), \ldots, \hat{z}_{rs}(m))'$ is an $(m \times 1)$ vector of the estimated correlations and $\mathbf{C}$ is an appropriate $(m \times m)$ symmetric matrix. It follows from the Theorem 1 that $\sqrt{T}(\hat{\mathbf{z}}_{rs} - \mathbf{z}_{rs}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$. Then the limiting $\chi^2(m)$ distribution of the Q_{rs} tests, for $(r, s) \in \{(1, 2), (2, 1), (2, 2)\}$, immediately follows from Theorem 9.8 in Schott (2005, p. 378) about the limiting distribution of a quadratic form of standard normal random variables. The degrees of freedom follow from the fact that $\mathbf{C}$ is a matrix with $\mathrm{rk}(\mathbf{C}) = m$. ■

3.8 Appendix B: Tables

Table 3.5 Statistical properties of the Q tests: NID(0,1)

			AR (#8)			**MA (#8)**		
sample	test	lag m	mean	var	var/mean	mean	var	var/mean
T=200	Q_{12}	5	5.0	9.8	1.96	5.0	10.1	2.02
		10	9.9	19.6	1.97	10.0	19.9	2.00
		15	14.9	29.4	1.97	14.9	29.8	2.00
	Q_{21}	5	5.0	9.7	1.96	5.1	10.3	2.03
		10	9.9	19.3	1.94	10.0	20.1	2.00
		15	14.9	28.2	1.90	15.0	30.4	2.02
	Q_{22}	5	5.0	9.9	1.99	5.0	10.5	2.12
		10	9.9	21.0	2.12	9.9	21.3	2.15
		15	14.9	32.8	2.20	14.9	32.9	2.21
T=500	Q_{12}	5	5.0	10.1	2.03	5.0	10.2	2.03
		10	10.0	20.6	2.06	10.1	20.4	2.03
		15	15.0	31.7	2.11	15.1	29.8	1.98
	Q_{21}	5	5.0	10.0	2.01	5.1	10.2	2.02
		10	10.0	19.2	1.93	10.0	20.3	2.03
		15	14.9	28.8	1.93	15.0	29.8	1.99
	Q_{22}	5	5.0	9.7	1.97	5.0	10.1	2.01
		10	10.0	20.2	2.02	9.9	20.0	2.02
		15	15.0	31.1	2.07	14.9	31.1	2.09
T=1000	Q_{12}	5	5.0	10.1	2.02	5.0	10.0	2.00
		10	9.9	20.4	2.06	10.0	20.3	2.02
		15	15.0	30.9	2.07	15.1	30.4	2.01
	Q_{21}	5	5.0	10.2	2.03	5.0	9.9	1.97
		10	10.0	19.5	1.95	10.0	19.9	1.99
		15	15.0	29.6	1.98	15.0	30.1	2.01
	Q_{22}	5	5.0	10.2	2.06	5.0	10.0	1.98
		10	10.0	21.0	2.10	10.0	19.9	1.99
		15	14.9	31.6	2.11	15.0	30.4	2.03

[a] AR (#8) indicates that 8 different parameter configurations of an AR model are evaluated.

[b] mean stands for a sample mean of the Q test over all replications and parameter configurations, var stands for a sample variance value of the Q test over all replications and parameter configurations, var/mean denotes a variance-mean ratio. The significance level is set to $\alpha = 0.05$.

Table 3.6 Size of the Q tests: NID(0,1)

			AR (#8)			MA (#8)		
sample	test	lag m	avg	min	max	avg	min	max
T=200	Q_{12}	5	0.050	0.037	0.063	0.049	0.043	0.054
		10	0.046	0.035	0.056	0.054	0.045	0.057
		15	0.045	0.038	0.052	0.049	0.043	0.054
		m	0.050	0.037	0.058	0.049	0.044	0.063
	Q_{21}	5	0.049	0.036	0.056	0.049	0.040	0.063
		10	0.051	0.044	0.057	0.044	0.034	0.057
		15	0.048	0.044	0.055	0.049	0.038	0.059
		m	0.049	0.035	0.062	0.047	0.039	0.055
	Q_{22}	5	0.049	0.037	0.057	0.054	0.040	0.063
		10	0.054	0.040	0.063	0.058	0.046	0.064
		15	0.056	0.039	0.062	0.058	0.053	0.071
		m	0.057	0.042	0.079	0.055	0.041	0.069
T=500	Q_{12}	5	0.050	0.045	0.055	0.046	0.040	0.057
		10	0.046	0.036	0.056	0.049	0.041	0.059
		15	0.044	0.034	0.051	0.047	0.033	0.062
		m	0.051	0.045	0.055	0.047	0.040	0.057
	Q_{21}	5	0.050	0.038	0.058	0.057	0.047	0.071
		10	0.051	0.038	0.065	0.055	0.043	0.062
		15	0.050	0.041	0.060	0.050	0.042	0.058
		m	0.050	0.033	0.060	0.051	0.039	0.072
	Q_{22}	5	0.048	0.034	0.065	0.056	0.049	0.066
		10	0.051	0.037	0.064	0.054	0.045	0.062
		15	0.053	0.042	0.066	0.051	0.038	0.065
		m	0.056	0.048	0.065	0.062	0.051	0.072
T=1000	Q_{12}	5	0.049	0.044	0.056	0.050	0.045	0.053
		10	0.048	0.037	0.056	0.050	0.033	0.062
		15	0.051	0.043	0.058	0.050	0.044	0.060
		m	0.053	0.039	0.073	0.054	0.044	0.073
	Q_{21}	5	0.049	0.036	0.060	0.050	0.039	0.063
		10	0.050	0.043	0.056	0.050	0.040	0.066
		15	0.048	0.038	0.055	0.051	0.042	0.061
		m	0.052	0.044	0.066	0.048	0.037	0.056
	Q_{22}	5	0.052	0.038	0.064	0.050	0.032	0.059
		10	0.051	0.040	0.068	0.051	0.042	0.066
		15	0.049	0.040	0.055	0.054	0.045	0.064
		m	0.054	0.041	0.067	0.057	0.050	0.068

[a] m denotes the automatically selected lag order based on Escanciano and Lobato (2009).

[b] AR (#8) indicates that 8 different parameter configurations of an AR model are evaluated.

[c] avg stands for the average rejection frequency of the Q tests over all parameter configurations of a given time series model, min and max denote a minimum and maximum of the average rejection frequencies over all parameter configurations. The significance level is set to $\alpha = 0.05$.

Table 3.7 Power properties of the Q tests: NID(0,1), $T = 200$

		TAR (#24)			**EXPAR (#24)**			**MAR (#24)**			**MSAR (#24)**			**GARCH (#17)**			**NLGARCH (#17)**			**TMA (#24)**			**BL (#18)**			**NLMA (#12)**		
test	lag	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max
Q_{12}	5	0.31	0.04	0.80	0.05	0.03	0.06	0.09	0.04	0.23	0.13	0.06	0.22	0.08	0.05	0.11	0.09	0.05	0.17	0.51	0.03	0.98	0.78	0.62	1.00	0.05	0.04	0.06
	10	0.25	0.05	0.69	0.05	0.03	0.07	0.08	0.03	0.21	0.11	0.04	0.18	0.07	0.04	0.12	0.08	0.04	0.18	0.44	0.04	0.94	0.66	0.49	0.99	0.05	0.03	0.07
	15	0.21	0.04	0.61	0.05	0.03	0.06	0.07	0.03	0.20	0.09	0.05	0.16	0.07	0.05	0.12	0.08	0.04	0.17	0.39	0.04	0.90	0.58	0.41	0.97	0.05	0.04	0.06
	m	0.36	0.04	0.93	0.05	0.04	0.10	0.10	0.03	0.22	0.15	0.06	0.27	0.08	0.06	0.12	0.10	0.05	0.18	0.59	0.04	1.00	0.85	0.54	1.00	0.05	0.04	0.07
Q_{21}	5	0.59	0.05	0.90	0.05	0.04	0.07	0.18	0.04	0.45	0.20	0.07	0.57	0.08	0.05	0.11	0.25	0.08	0.54	0.05	0.04	0.07	0.09	0.04	0.36	0.26	0.11	0.42
	10	0.50	0.05	0.80	0.05	0.03	0.06	0.15	0.03	0.39	0.17	0.06	0.52	0.07	0.05	0.11	0.21	0.06	0.50	0.05	0.04	0.06	0.08	0.04	0.32	0.19	0.08	0.32
	15	0.43	0.04	0.73	0.05	0.04	0.07	0.13	0.04	0.36	0.16	0.06	0.47	0.07	0.05	0.11	0.19	0.06	0.45	0.05	0.04	0.06	0.08	0.04	0.28	0.16	0.07	0.25
	m	0.63	0.06	0.98	0.05	0.03	0.07	0.21	0.05	0.49	0.22	0.08	0.59	0.09	0.06	0.14	0.26	0.10	0.48	0.06	0.04	0.09	0.09	0.04	0.38	0.43	0.20	0.68
Q_{22}	5	0.19	0.04	0.66	0.05	0.04	0.07	0.45	0.04	1.00	0.48	0.10	0.98	0.21	0.06	0.52	0.31	0.08	0.75	0.07	0.04	0.13	0.22	0.11	0.71	0.05	0.04	0.06
	10	0.15	0.04	0.53	0.05	0.04	0.07	0.42	0.05	1.00	0.43	0.09	0.97	0.19	0.07	0.47	0.28	0.07	0.70	0.06	0.04	0.12	0.19	0.09	0.65	0.05	0.04	0.06
	15	0.14	0.05	0.46	0.06	0.04	0.07	0.40	0.05	1.00	0.40	0.09	0.95	0.17	0.07	0.44	0.26	0.07	0.67	0.06	0.04	0.10	0.17	0.09	0.59	0.06	0.04	0.07
	m	0.22	0.05	0.82	0.05	0.04	0.08	0.51	0.04	1.00	0.53	0.11	0.98	0.24	0.07	0.52	0.33	0.09	0.72	0.09	0.05	0.19	0.27	0.16	0.81	0.07	0.06	0.09

[a] m denotes the automatically selected lag order based on Escanciano and Lobato (2009).
[b] TAR (#24) indicates that 24 different parameter configurations of a TAR model are evaluated.
[c] avg stands for the average rejection frequency of the Q tests over all parameter configurations of a given time series model, min and max denote a minimum and maximum of the average rejection frequencies over all parameter configurations. The significance level is set to $\alpha = 0.05$.

Table 3.8 Power properties of the Q tests: NID(0,1), $T = 500$

		TAR (#24)			**EXPAR (#24)**			**MAR (#24)**			**MSAR (#24)**			**GARCH (#17)**			**NLGARCH (#17)**			**TMA (#24)**			**BL (#18)**			**NLMA (#12)**		
test	lag	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max
Q_{12}	5	0.53	0.04	1.00	0.05	0.04	0.07	0.10	0.04	0.22	0.19	0.08	0.35	0.08	0.05	0.13	0.11	0.06	0.24	0.70	0.05	1.00	0.99	0.98	1.00	0.05	0.04	0.06
	10	0.47	0.04	0.99	0.05	0.04	0.07	0.09	0.04	0.22	0.16	0.06	0.29	0.08	0.04	0.13	0.10	0.05	0.26	0.67	0.04	1.00	0.98	0.93	1.00	0.05	0.04	0.06
	15	0.44	0.04	0.98	0.05	0.04	0.07	0.08	0.04	0.21	0.14	0.06	0.24	0.08	0.04	0.15	0.10	0.05	0.24	0.64	0.05	1.00	0.97	0.89	1.00	0.05	0.04	0.07
	m	0.53	0.04	1.00	0.06	0.03	0.08	0.10	0.04	0.22	0.20	0.07	0.36	0.08	0.04	0.15	0.12	0.06	0.28	0.72	0.04	1.00	0.99	0.94	1.00	0.05	0.04	0.06
Q_{21}	5	0.82	0.07	1.00	0.05	0.04	0.06	0.30	0.04	0.67	0.30	0.08	0.75	0.08	0.05	0.16	0.48	0.14	0.85	0.07	0.04	0.11	0.14	0.04	0.68	0.57	0.25	0.86
	10	0.79	0.05	1.00	0.05	0.04	0.06	0.25	0.04	0.62	0.26	0.06	0.71	0.08	0.04	0.16	0.42	0.10	0.83	0.06	0.04	0.09	0.13	0.04	0.61	0.47	0.17	0.76
	15	0.77	0.05	1.00	0.05	0.04	0.06	0.23	0.04	0.57	0.24	0.06	0.67	0.07	0.05	0.14	0.39	0.10	0.80	0.06	0.04	0.08	0.12	0.04	0.57	0.41	0.14	0.67
	m	0.83	0.06	1.00	0.05	0.04	0.06	0.33	0.05	0.72	0.30	0.07	0.76	0.09	0.05	0.16	0.49	0.16	0.83	0.09	0.04	0.18	0.15	0.04	0.68	0.74	0.49	0.98
Q_{22}	5	0.35	0.05	0.98	0.05	0.04	0.06	0.57	0.04	1.00	0.71	0.23	1.00	0.44	0.09	0.92	0.57	0.15	0.98	0.11	0.04	0.33	0.44	0.26	0.99	0.05	0.04	0.06
	10	0.31	0.05	0.95	0.05	0.04	0.07	0.55	0.05	1.00	0.67	0.18	1.00	0.39	0.07	0.91	0.52	0.13	0.97	0.10	0.05	0.25	0.37	0.20	0.98	0.05	0.04	0.07
	15	0.28	0.05	0.92	0.05	0.04	0.06	0.54	0.05	1.00	0.64	0.16	1.00	0.35	0.07	0.88	0.49	0.12	0.96	0.09	0.04	0.22	0.34	0.18	0.97	0.05	0.04	0.06
	m	0.36	0.06	0.99	0.06	0.04	0.08	0.60	0.05	1.00	0.73	0.23	1.00	0.47	0.12	0.91	0.60	0.20	0.97	0.17	0.05	0.49	0.52	0.33	1.00	0.06	0.05	0.07

[a] m denotes the automatically selected lag order based on Escanciano and Lobato (2009).
[b] TAR (#24) indicates that 24 different parameter configurations of a TAR model are evaluated.
[c] avg stands for the average rejection frequency of the Q tests over all parameter configurations of a given time series model, min and max denote a minimum and maximum of the average rejection frequencies over all parameter configurations. The significance level is set to $\alpha = 0.05$.

Table 3.9 Power properties of the Q tests: NID(0,1), $T = 1000$

		TAR (#24)			**EXPAR (#24)**			**MAR (#24)**			**MSAR (#24)**			**GARCH (#17)**			**NLGARCH (#17)**			**TMA (#24)**			**BL (#18)**			**NLMA (#12)**		
test	lag	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max
Q_{12}	5	0.65	0.04	1.00	0.05	0.04	0.08	0.10	0.04	0.25	0.26	0.10	0.49	0.08	0.05	0.16	0.11	0.05	0.29	0.78	0.04	1.00	1.00	1.00	1.00	0.05	0.04	0.06
	10	0.61	0.04	1.00	0.05	0.04	0.07	0.09	0.05	0.24	0.21	0.08	0.40	0.08	0.04	0.17	0.11	0.05	0.34	0.76	0.04	1.00	1.00	1.00	1.00	0.05	0.04	0.06
	15	0.59	0.03	1.00	0.05	0.04	0.06	0.09	0.04	0.23	0.19	0.07	0.36	0.08	0.04	0.16	0.11	0.05	0.32	0.74	0.04	1.00	1.00	1.00	1.00	0.05	0.04	0.06
	m	0.64	0.04	1.00	0.06	0.04	0.09	0.11	0.04	0.24	0.27	0.09	0.50	0.09	0.06	0.17	0.13	0.06	0.35	0.79	0.04	1.00	1.00	1.00	1.00	0.06	0.04	0.07
Q_{21}	5	0.89	0.07	1.00	0.05	0.04	0.07	0.46	0.03	0.88	0.41	0.09	0.89	0.09	0.05	0.17	0.68	0.20	0.99	0.08	0.04	0.16	0.19	0.05	0.90	0.77	0.49	1.00
	10	0.87	0.07	1.00	0.05	0.04	0.06	0.39	0.05	0.85	0.36	0.08	0.87	0.08	0.06	0.17	0.63	0.15	0.99	0.07	0.04	0.12	0.18	0.05	0.85	0.69	0.36	0.99
	15	0.86	0.07	1.00	0.05	0.04	0.06	0.36	0.04	0.81	0.33	0.08	0.84	0.08	0.05	0.17	0.58	0.14	0.99	0.07	0.04	0.11	0.17	0.04	0.83	0.65	0.30	0.97
	m	0.89	0.07	1.00	0.05	0.04	0.07	0.49	0.04	0.89	0.41	0.09	0.88	0.09	0.06	0.17	0.69	0.25	0.99	0.13	0.05	0.28	0.19	0.05	0.90	0.89	0.77	1.00
Q_{22}	5	0.44	0.05	1.00	0.05	0.03	0.06	0.63	0.04	1.00	0.84	0.44	1.00	0.63	0.18	1.00	0.75	0.24	1.00	0.19	0.04	0.59	0.69	0.52	1.00	0.05	0.04	0.06
	10	0.42	0.06	1.00	0.05	0.04	0.07	0.61	0.04	1.00	0.80	0.36	1.00	0.58	0.14	1.00	0.70	0.19	1.00	0.15	0.05	0.47	0.60	0.41	1.00	0.05	0.04	0.07
	15	0.40	0.05	1.00	0.05	0.04	0.06	0.60	0.04	1.00	0.78	0.29	1.00	0.54	0.13	1.00	0.67	0.17	1.00	0.13	0.04	0.40	0.55	0.34	1.00	0.05	0.04	0.07
	m	0.45	0.06	1.00	0.06	0.04	0.11	0.65	0.05	1.00	0.84	0.41	1.00	0.66	0.24	1.00	0.77	0.30	1.00	0.28	0.06	0.76	0.77	0.63	1.00	0.06	0.05	0.09

[a] m denotes the automatically selected lag order based on Escanciano and Lobato (2009).
[b] TAR (#24) indicates that 24 different parameter configurations of a TAR model are evaluated.
[c] avg stands for the average rejection frequency of the Q tests over all parameter configurations of a given time series model, min and max denote a minimum and maximum of the average rejection frequencies over all parameter configurations. The significance level is set to $\alpha = 0.05$.

Table 3.10 Power properties of the Q tests: $T = 200$

				TAR (#24)			**EXPAR (#24)**			**MAR (#24)**			**MSAR (#24)**			**GARCH (#17)**			**NLGARCH (#17)**			**TMA (#24)**			**BL (#18)**			**NLMA (#12)**		
test	distr.	skew.	kurt.	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max
Q_{12}	N	0.0	3.0	0.36	0.04	0.93	0.05	0.04	0.10	0.10	0.03	0.22	0.15	0.06	0.27	0.08	0.06	0.12	0.10	0.05	0.18	0.59	0.04	1.00	0.85	0.54	1.00	0.05	0.04	0.07
	S1	0.0	6.0	0.42	0.04	0.95	0.06	0.04	0.08	0.13	0.03	0.29	0.16	0.07	0.28	0.11	0.06	0.20	0.13	0.07	0.27	0.63	0.04	1.00	0.94	0.69	0.98	0.06	0.04	0.07
	S2	0.0	11.6	0.44	0.05	0.95	0.06	0.04	0.10	0.15	0.05	0.28	0.17	0.06	0.33	0.13	0.08	0.20	0.15	0.07	0.27	0.64	0.04	1.00	0.95	0.70	0.99	0.06	0.05	0.08
	S3	0.0	126.0	0.46	0.06	0.94	0.06	0.03	0.10	0.18	0.04	0.34	0.19	0.05	0.32	0.15	0.09	0.24	0.17	0.10	0.28	0.65	0.06	1.00	0.95	0.76	0.99	0.06	0.04	0.07
	A1	-0.9	4.2	0.35	0.03	0.94	0.07	0.03	0.24	0.10	0.04	0.21	0.16	0.07	0.30	0.07	0.04	0.12	0.11	0.04	0.28	0.57	0.03	1.00	0.83	0.45	0.97	0.03	0.03	0.04
	A2	-1.5	7.5	0.35	0.03	0.96	0.09	0.03	0.37	0.12	0.04	0.28	0.18	0.07	0.33	0.07	0.04	0.11	0.13	0.04	0.32	0.56	0.03	0.99	0.83	0.44	0.96	0.03	0.02	0.03
	A3	-2.0	21.2	0.40	0.03	0.98	0.07	0.03	0.22	0.15	0.04	0.33	0.19	0.06	0.35	0.10	0.06	0.20	0.16	0.06	0.36	0.61	0.04	1.00	0.91	0.66	0.98	0.04	0.03	0.05
Q_{21}	N	0.0	3.0	0.63	0.06	0.98	0.05	0.03	0.07	0.21	0.05	0.49	0.22	0.08	0.59	0.09	0.06	0.14	0.26	0.10	0.48	0.06	0.04	0.09	0.09	0.04	0.38	0.43	0.20	0.68
	S1	0.0	6.0	0.48	0.07	0.93	0.05	0.04	0.07	0.21	0.04	0.48	0.21	0.07	0.56	0.10	0.06	0.16	0.23	0.11	0.40	0.08	0.05	0.13	0.15	0.08	0.45	0.30	0.13	0.48
	S2	0.0	11.6	0.42	0.07	0.87	0.05	0.04	0.07	0.22	0.04	0.49	0.22	0.07	0.59	0.13	0.07	0.21	0.23	0.10	0.38	0.09	0.05	0.15	0.21	0.11	0.43	0.26	0.11	0.41
	S3	0.0	126.0	0.36	0.07	0.80	0.05	0.04	0.08	0.23	0.05	0.51	0.22	0.07	0.58	0.14	0.08	0.21	0.22	0.11	0.37	0.11	0.05	0.17	0.27	0.19	0.44	0.21	0.10	0.32
	A1	-0.9	4.2	0.56	0.08	0.96	0.06	0.04	0.08	0.28	0.05	0.76	0.21	0.07	0.51	0.14	0.08	0.23	0.41	0.18	0.67	0.09	0.06	0.20	0.19	0.09	0.85	0.42	0.21	0.64
	A2	-1.5	7.5	0.47	0.08	0.89	0.07	0.06	0.10	0.32	0.05	0.89	0.23	0.08	0.50	0.18	0.09	0.31	0.44	0.22	0.67	0.12	0.06	0.24	0.34	0.15	0.95	0.38	0.18	0.58
	A3	-2.0	21.2	0.39	0.09	0.81	0.06	0.05	0.09	0.28	0.05	0.74	0.22	0.07	0.51	0.16	0.08	0.26	0.37	0.17	0.60	0.11	0.06	0.23	0.35	0.16	0.89	0.27	0.13	0.42
Q_{22}	N	0.0	3.0	0.22	0.05	0.82	0.05	0.04	0.08	0.51	0.04	1.00	0.53	0.11	0.98	0.24	0.07	0.52	0.33	0.09	0.72	0.09	0.05	0.19	0.27	0.16	0.81	0.07	0.06	0.09
	S1	0.0	6.0	0.25	0.05	0.82	0.06	0.04	0.09	0.55	0.04	1.00	0.54	0.08	0.98	0.27	0.11	0.54	0.34	0.11	0.69	0.23	0.05	0.64	0.66	0.54	0.97	0.07	0.04	0.08
	S2	0.0	11.6	0.27	0.05	0.83	0.06	0.04	0.09	0.56	0.04	1.00	0.55	0.09	0.98	0.28	0.11	0.54	0.34	0.12	0.65	0.32	0.07	0.80	0.79	0.68	0.99	0.07	0.05	0.08
	S3	0.0	126.0	0.30	0.07	0.85	0.06	0.04	0.10	0.58	0.05	1.00	0.54	0.07	0.98	0.28	0.13	0.52	0.33	0.14	0.58	0.40	0.08	0.91	0.89	0.82	1.00	0.07	0.05	0.09
	A1	-0.9	4.2	0.11	0.03	0.30	0.06	0.04	0.07	0.52	0.04	1.00	0.55	0.11	0.98	0.26	0.09	0.57	0.51	0.20	0.90	0.14	0.05	0.45	0.38	0.19	0.95	0.11	0.08	0.14
	A2	-1.5	7.5	0.12	0.02	0.46	0.07	0.04	0.10	0.55	0.04	1.00	0.56	0.10	0.98	0.28	0.12	0.54	0.54	0.24	0.88	0.19	0.06	0.64	0.56	0.37	0.98	0.11	0.08	0.14
	A3	-2.0	21.2	0.20	0.04	0.60	0.06	0.04	0.11	0.56	0.04	1.00	0.55	0.08	0.99	0.27	0.11	0.56	0.48	0.19	0.83	0.27	0.07	0.81	0.76	0.60	0.99	0.07	0.06	0.08

[a] m denotes the automatically selected lag order based on Escanciano and Lobato (2009).
[b] TAR (#24) indicates that 24 different parameter configurations of a TAR model are evaluated.
[c] avg stands for the average rejection frequency of the Q tests over all parameter configurations of a given time series model, min and max denote a minimum and maximum of the average rejection frequencies over all parameter configurations. The significance level is set to $\alpha = 0.05$.

Table 3.11 Power properties of the Q tests: $T = 500$

				TAR (#24)			**EXPAR (#24)**			**MAR (#24)**			**MSAR (#24)**			**GARCH (#17)**			**NLGARCH (#17)**			**TMA (#24)**			**BL (#18)**			**NLMA (#12)**		
test	distr.	skew.	kurt.	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max
Q_{12}	N	0.0	3.0	0.53	0.04	1.00	0.06	0.03	0.08	0.10	0.04	0.22	0.20	0.07	0.36	0.08	0.04	0.15	0.12	0.06	0.28	0.72	0.04	1.00	0.99	0.94	1.00	0.05	0.04	0.06
	S1	0.0	6.0	0.60	0.05	1.00	0.05	0.04	0.08	0.14	0.05	0.28	0.21	0.09	0.39	0.14	0.09	0.27	0.18	0.09	0.38	0.75	0.05	1.00	1.00	0.98	1.00	0.06	0.04	0.07
	S2	0.0	11.6	0.62	0.05	1.00	0.06	0.04	0.10	0.18	0.05	0.33	0.23	0.07	0.41	0.17	0.10	0.32	0.22	0.11	0.43	0.76	0.05	1.00	0.99	0.97	1.00	0.06	0.04	0.08
	S3	0.0	126.0	0.64	0.05	1.00	0.07	0.04	0.12	0.24	0.05	0.44	0.26	0.09	0.45	0.22	0.11	0.39	0.26	0.12	0.46	0.77	0.05	1.00	0.98	0.94	1.00	0.06	0.04	0.10
	A1	-0.9	4.2	0.53	0.04	1.00	0.12	0.04	0.53	0.11	0.04	0.24	0.23	0.10	0.44	0.09	0.05	0.18	0.18	0.05	0.49	0.71	0.05	1.00	0.99	0.88	1.00	0.03	0.02	0.04
	A2	-1.5	7.5	0.52	0.02	1.00	0.17	0.03	0.76	0.14	0.04	0.34	0.25	0.11	0.50	0.10	0.04	0.23	0.21	0.06	0.58	0.71	0.04	1.00	0.98	0.88	1.00	0.02	0.02	0.03
	A3	-2.0	21.2	0.58	0.04	1.00	0.12	0.04	0.41	0.19	0.04	0.37	0.27	0.10	0.48	0.16	0.09	0.31	0.28	0.11	0.59	0.74	0.04	1.00	0.98	0.84	1.00	0.04	0.03	0.04
Q_{21}	N	0.0	3.0	0.83	0.06	1.00	0.05	0.04	0.06	0.33	0.05	0.72	0.30	0.07	0.76	0.09	0.05	0.16	0.49	0.16	0.83	0.09	0.04	0.18	0.15	0.04	0.68	0.74	0.49	0.98
	S1	0.0	6.0	0.72	0.10	1.00	0.05	0.03	0.07	0.34	0.04	0.70	0.31	0.08	0.75	0.13	0.07	0.23	0.44	0.18	0.79	0.12	0.04	0.24	0.23	0.12	0.68	0.61	0.33	0.90
	S2	0.0	11.6	0.67	0.11	0.99	0.05	0.04	0.07	0.34	0.05	0.69	0.32	0.07	0.74	0.17	0.09	0.29	0.42	0.17	0.71	0.14	0.05	0.29	0.32	0.20	0.70	0.53	0.27	0.79
	S3	0.0	126.0	0.60	0.10	0.98	0.06	0.05	0.09	0.35	0.05	0.67	0.32	0.07	0.73	0.20	0.10	0.35	0.41	0.18	0.69	0.17	0.06	0.35	0.40	0.29	0.70	0.44	0.20	0.68
	A1	-0.9	4.2	0.78	0.10	1.00	0.07	0.05	0.10	0.42	0.05	0.99	0.31	0.07	0.62	0.22	0.12	0.45	0.70	0.35	0.97	0.12	0.07	0.36	0.33	0.12	1.00	0.71	0.44	0.96
	A2	-1.5	7.5	0.71	0.07	1.00	0.08	0.06	0.12	0.43	0.05	1.00	0.36	0.08	0.69	0.32	0.12	0.57	0.73	0.40	0.96	0.17	0.07	0.48	0.60	0.29	1.00	0.65	0.38	0.92
	A3	-2.0	21.2	0.64	0.12	0.99	0.07	0.04	0.15	0.43	0.05	0.98	0.34	0.09	0.63	0.28	0.14	0.48	0.67	0.32	0.94	0.18	0.08	0.48	0.60	0.32	0.99	0.54	0.27	0.82
Q_{22}	N	0.0	3.0	0.36	0.06	0.99	0.06	0.04	0.08	0.60	0.05	1.00	0.73	0.23	1.00	0.47	0.12	0.91	0.60	0.20	0.97	0.17	0.05	0.49	0.52	0.33	1.00	0.06	0.05	0.07
	S1	0.0	6.0	0.41	0.06	0.99	0.07	0.03	0.12	0.64	0.04	1.00	0.71	0.16	1.00	0.52	0.19	0.93	0.60	0.19	0.97	0.43	0.06	0.96	0.96	0.91	1.00	0.07	0.06	0.08
	S2	0.0	11.6	0.47	0.06	1.00	0.08	0.04	0.14	0.66	0.05	1.00	0.70	0.12	1.00	0.53	0.20	0.90	0.60	0.21	0.96	0.53	0.07	1.00	0.99	0.98	1.00	0.07	0.05	0.09
	S3	0.0	126.0	0.52	0.05	0.99	0.08	0.04	0.17	0.68	0.06	1.00	0.70	0.11	1.00	0.53	0.20	0.90	0.59	0.26	0.92	0.61	0.09	1.00	1.00	1.00	1.00	0.08	0.06	0.11
	A1	-0.9	4.2	0.22	0.04	0.64	0.06	0.04	0.10	0.61	0.05	1.00	0.75	0.26	1.00	0.51	0.16	0.93	0.79	0.39	1.00	0.25	0.06	0.85	0.67	0.40	1.00	0.18	0.10	0.27
	A2	-1.5	7.5	0.25	0.04	0.86	0.07	0.05	0.14	0.64	0.05	1.00	0.74	0.20	1.00	0.54	0.19	0.92	0.82	0.47	1.00	0.33	0.08	0.96	0.88	0.75	1.00	0.24	0.12	0.39
	A3	-2.0	21.2	0.36	0.03	0.96	0.08	0.04	0.17	0.66	0.06	1.00	0.72	0.15	1.00	0.54	0.23	0.91	0.79	0.40	1.00	0.46	0.08	1.00	0.98	0.96	1.00	0.12	0.06	0.19

[a] m denotes the automatically selected lag order based on Escanciano and Lobato (2009).
[b] TAR (#24) indicates that 24 different parameter configurations of a TAR model are evaluated.
[c] avg stands for the average rejection frequency of the Q tests over all parameter configurations of a given time series model, min and max denote a minimum and maximum of the average rejection frequencies over all parameter configurations. The significance level is set to $\alpha = 0.05$.

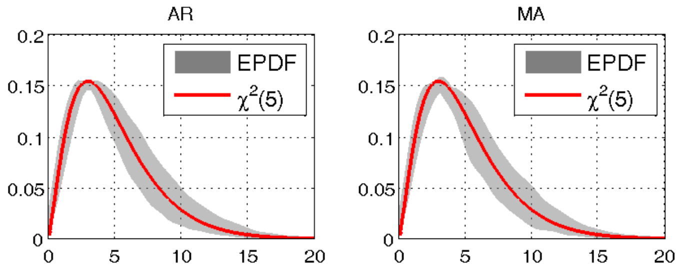
Table 3.12 Power properties of the Q tests: $T = 1000$

				TAR (#24)			**EXPAR (#24)**			**MAR (#24)**			**MSAR (#24)**			**GARCH (#17)**			**NLGARCH (#17)**			**TMA (#24)**			**BL (#18)**			**NLMA (#12)**		
test	distr.	skew.	kurt.	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max
Q_{12}	N	0.0	3.0	0.64	0.04	1.00	0.06	0.04	0.09	0.11	0.04	0.24	0.27	0.09	0.50	0.09	0.06	0.17	0.13	0.06	0.35	0.79	0.04	1.00	1.00	1.00	1.00	0.06	0.04	0.07
	S1	0.0	6.0	0.69	0.04	1.00	0.06	0.04	0.08	0.15	0.04	0.32	0.26	0.10	0.47	0.16	0.07	0.30	0.20	0.08	0.48	0.80	0.05	1.00	1.00	1.00	1.00	0.06	0.05	0.07
	S2	0.0	11.6	0.72	0.06	1.00	0.06	0.05	0.10	0.20	0.04	0.40	0.28	0.12	0.49	0.21	0.10	0.41	0.27	0.12	0.57	0.81	0.05	1.00	1.00	0.99	1.00	0.06	0.04	0.08
	S3	0.0	126.0	0.74	0.04	1.00	0.07	0.05	0.12	0.27	0.06	0.47	0.32	0.09	0.54	0.28	0.15	0.50	0.33	0.16	0.60	0.82	0.06	1.00	0.99	0.97	1.00	0.07	0.05	0.10
	A1	-0.9	4.2	0.64	0.04	1.00	0.19	0.03	0.77	0.11	0.04	0.28	0.31	0.15	0.60	0.09	0.05	0.20	0.21	0.05	0.61	0.78	0.03	1.00	1.00	0.99	1.00	0.03	0.03	0.04
	A2	-1.5	7.5	0.63	0.04	1.00	0.27	0.03	0.96	0.14	0.04	0.34	0.33	0.14	0.64	0.12	0.05	0.29	0.28	0.08	0.72	0.79	0.03	1.00	1.00	0.95	1.00	0.02	0.02	0.03
	A3	-2.0	21.2	0.69	0.05	1.00	0.19	0.04	0.70	0.21	0.05	0.42	0.35	0.11	0.63	0.20	0.09	0.43	0.36	0.12	0.73	0.81	0.06	1.00	0.99	0.89	1.00	0.04	0.03	0.05
Q_{21}	N	0.0	3.0	0.89	0.07	1.00	0.05	0.04	0.07	0.49	0.04	0.89	0.41	0.09	0.88	0.09	0.06	0.17	0.69	0.25	0.99	0.13	0.05	0.28	0.19	0.05	0.90	0.89	0.77	1.00
	S1	0.0	6.0	0.80	0.12	1.00	0.05	0.03	0.07	0.48	0.04	0.87	0.41	0.09	0.88	0.15	0.08	0.28	0.65	0.27	0.96	0.17	0.05	0.43	0.30	0.13	0.89	0.81	0.59	1.00
	S2	0.0	11.6	0.77	0.13	1.00	0.06	0.04	0.08	0.47	0.04	0.84	0.42	0.10	0.87	0.19	0.09	0.36	0.62	0.24	0.93	0.20	0.05	0.49	0.41	0.25	0.87	0.75	0.50	0.98
	S3	0.0	126.0	0.73	0.13	1.00	0.06	0.05	0.10	0.46	0.06	0.79	0.43	0.09	0.85	0.25	0.12	0.51	0.60	0.27	0.90	0.25	0.07	0.56	0.51	0.38	0.86	0.66	0.39	0.92
	A1	-0.9	4.2	0.84	0.13	1.00	0.07	0.05	0.10	0.51	0.05	1.00	0.44	0.09	0.76	0.34	0.13	0.68	0.87	0.57	1.00	0.17	0.07	0.59	0.50	0.18	1.00	0.87	0.71	1.00
	A2	-1.5	7.5	0.80	0.10	1.00	0.09	0.05	0.19	0.48	0.06	1.00	0.47	0.09	0.95	0.49	0.21	0.81	0.90	0.66	1.00	0.23	0.10	0.75	0.78	0.52	1.00	0.82	0.60	1.00
	A3	-2.0	21.2	0.77	0.13	1.00	0.09	0.06	0.20	0.49	0.06	1.00	0.46	0.09	0.89	0.42	0.19	0.70	0.86	0.57	1.00	0.27	0.10	0.77	0.78	0.47	1.00	0.75	0.48	0.97
Q_{22}	N	0.0	3.0	0.45	0.06	1.00	0.06	0.04	0.11	0.65	0.05	1.00	0.84	0.41	1.00	0.66	0.24	1.00	0.77	0.30	1.00	0.28	0.06	0.76	0.77	0.63	1.00	0.06	0.05	0.09
	S1	0.0	6.0	0.55	0.07	1.00	0.07	0.04	0.17	0.69	0.04	1.00	0.80	0.22	1.00	0.70	0.25	1.00	0.79	0.36	1.00	0.56	0.08	1.00	1.00	1.00	1.00	0.07	0.05	0.08
	S2	0.0	11.6	0.62	0.07	1.00	0.09	0.04	0.19	0.71	0.05	1.00	0.79	0.19	1.00	0.72	0.31	1.00	0.78	0.35	1.00	0.63	0.08	1.00	1.00	1.00	1.00	0.08	0.06	0.10
	S3	0.0	126.0	0.67	0.07	1.00	0.10	0.04	0.26	0.72	0.06	1.00	0.78	0.16	1.00	0.72	0.33	1.00	0.78	0.40	0.99	0.70	0.09	1.00	1.00	1.00	1.00	0.09	0.06	0.12
	A1	-0.9	4.2	0.37	0.05	0.92	0.07	0.04	0.15	0.66	0.05	1.00	0.87	0.46	1.00	0.70	0.26	1.00	0.92	0.64	1.00	0.34	0.07	0.99	0.87	0.68	1.00	0.33	0.15	0.52
	A2	-1.5	7.5	0.38	0.04	1.00	0.08	0.05	0.21	0.69	0.04	1.00	0.86	0.38	1.00	0.74	0.32	1.00	0.94	0.73	1.00	0.42	0.08	1.00	0.99	0.96	1.00	0.45	0.20	0.71
	A3	-2.0	21.2	0.50	0.04	1.00	0.09	0.04	0.28	0.71	0.06	1.00	0.82	0.24	1.00	0.74	0.33	0.99	0.93	0.68	1.00	0.56	0.08	1.00	1.00	1.00	1.00	0.22	0.10	0.35

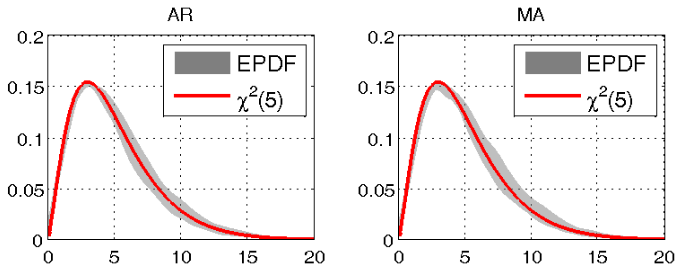
[a] m denotes the automatically selected lag order based on Escanciano and Lobato (2009).
[b] TAR (#24) indicates that 24 different parameter configurations of a TAR model are evaluated.
[c] avg stands for the average rejection frequency of the Q tests over all parameter configurations of a given time series model, min and max denote a minimum and maximum of the average rejection frequencies over all parameter configurations. The significance level is set to $\alpha = 0.05$.

3.9 Appendix C: Figures

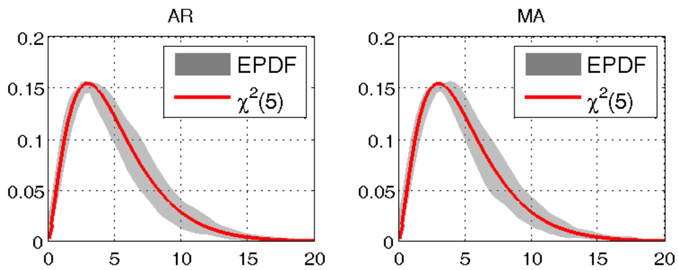
Figure 3.4 Smoothed empirical density functions of the Q(5) tests and $\chi^2(5)$: $T = 200,\ R = 5000$



(a) $Q_{12}(5)$



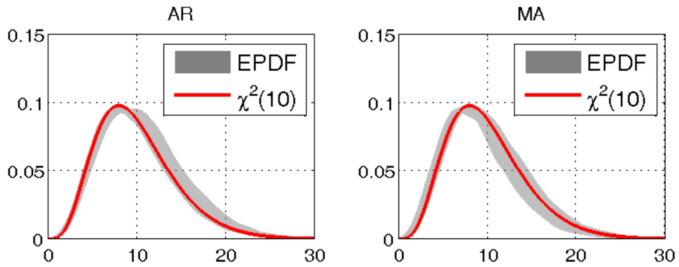
(b) $Q_{21}(5)$



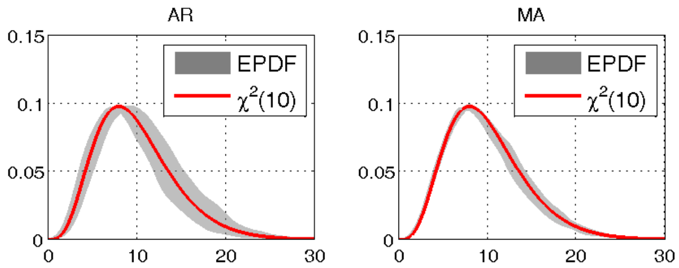
(c) $Q_{22}(5)$

* The empirical densities are smoothed by a kernel smoothing procedure with a simple reference bandwidth for all parameters of AR and MA models $\phi, \theta \in \{-0.8, -0.6, -0.4, -0.2, 0.2, 0.4, 0.6, 0.8\}$. Figures depict bands calculated from the highest and lowest smoothed empirical density functions in order to explicitly show parameter uncertainty of the finite sample distributions.

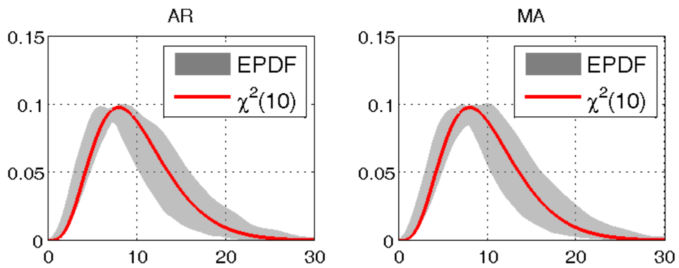
Figure 3.5 Smoothed empirical density functions of the Q(10) tests and $\chi^2(10)$: $T = 200,\ R = 5000$



(a) $Q_{12}(10)$



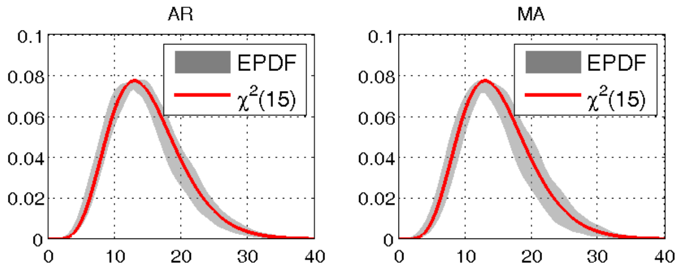
(b) $Q_{21}(10)$



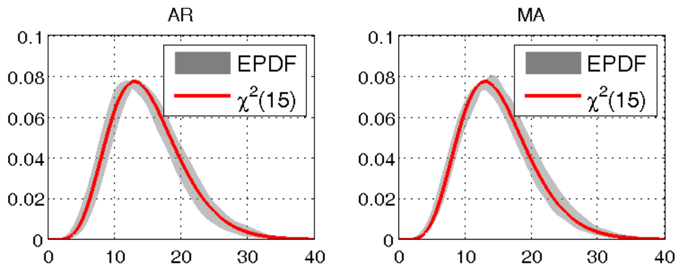
(c) $Q_{22}(10)$

* The empirical densities are smoothed by a kernel smoothing procedure with a simple reference bandwidth for all parameters of AR and MA models $\phi, \theta \in \{-0.8, -0.6, -0.4, -0.2, 0.2, 0.4, 0.6, 0.8\}$. Figures depict bands calculated from the highest and lowest smoothed empirical density functions in order to explicitly show parameter uncertainty of the finite sample distributions.

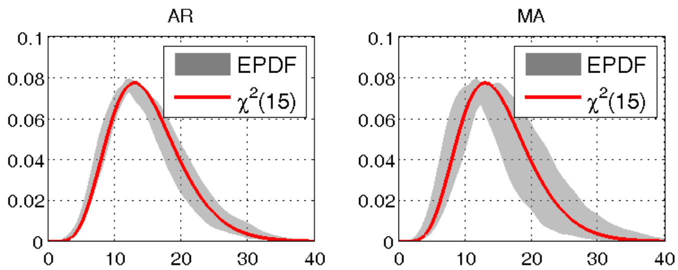
Figure 3.6 Smoothed empirical density functions of the Q(15) tests and $\chi^2(15)$: $T = 200$, $R = 5000$



(a) $Q_{12}(15)$



(b) $Q_{21}(15)$



(c) $Q_{22}(15)$

* The empirical densities are smoothed by a kernel smoothing procedure with a simple reference bandwidth for all parameters of AR and MA models $\phi, \theta \in \{-0.8, -0.6, -0.4, -0.2, 0.2, 0.4, 0.6, 0.8\}$. Figures depict bands calculated from the highest and lowest smoothed empirical density functions in order to explicitly show parameter uncertainty of the finite sample distributions.

3.10 Appendix D: Data

Table 3.13 Description of weakly time series: sample 07/01/1980 – 31/12/2010

variable	description	transformation
Exchange rates		
USDGBP	US dollar to British pound exchange rate	$\Delta \log$
USDJPY	US dollar to Japanese yen exchange rate	$\Delta \log$
USDCAD	US dollar to Canadian dollar exchange rate	$\Delta \log$
USDAUD	US dollar to Australian dollar exchange rate	$\Delta \log$
USDCHF	US dollar to Swiss frank exchange rate	$\Delta \log$
Interest rates		
USIR3M	US interbank interest rates, 3M	Δ
UKIR3M	UK interbank interest rates, 3M	Δ
CAIR3M	Canadian interbank interest rates, 3M	Δ
AUIR3M	Australian interbank interest rates, 3M	Δ
CHFIR3M	Swiss interbank interest rates, 3M	Δ
Equity indices		
DJIA	US Dow Jones Industrials Share Index	$\Delta \log$
FTSE	UK FT All Shares Index	$\Delta \log$
TOPIX	Tokyo Stock Exchange Index	$\Delta \log$
TSE	Toronto Stock Exchange Index	$\Delta \log$
AUSE	Australian Stock Exchange Index	$\Delta \log$
CHSE	Swiss Stock Exchange Index	$\Delta \log$
Commodities		
WHEAT	Kansas wheat, hard, cents/bushel	$\Delta \log$
SOYBEAN	soybeans, yellow, cents/bushel	$\Delta \log$
COFFEE	Brazilian coffee beans, cents/pound	$\Delta \log$
COTTON	cotton, cents/pound	$\Delta \log$
FUEL	fuel oil, cents/gallon	$\Delta \log$
GOLD	gold bullion, USD/troy ounce	$\Delta \log$

[a] Source: Thomson Reuters.

[b] Δ denotes a first difference of a given series, $\Delta \log$ is an approximation to the growth rate of a given time series.

Table 3.14 Application of the Q tests

	$\mathbf{Q_{12}}$				$\mathbf{Q_{21}}$				$\mathbf{Q_{22}}$			
variable/lag	5	10	15	m	5	10	15	m	5	10	15	m
Exchange rates												
USDGBP	0.37	0.46	0.43	0.12	0.08	0.11	0.03	0.03	0.00	0.00	0.00	0.00
USDJPY	0.97	0.94	0.66	0.00	0.10	0.03	0.07	0.21	0.00	0.00	0.00	0.00
USDCAD	0.37	0.66	0.80	0.31	0.89	0.92	0.96	0.96	0.00	0.00	0.00	0.00
USDAUD	0.88	0.70	0.81	0.48	0.05	0.19	0.24	0.33	0.00	0.00	0.00	0.00
USDCHF	0.85	0.99	1.00	0.98	0.43	0.25	0.24	0.25	0.00	0.00	0.00	0.00
Interest rates												
USIR3M	0.62	0.93	0.83	0.89	0.99	1.00	0.89	0.97	0.00	0.00	0.00	0.00
UKIR3M	0.11	0.08	0.02	0.01	0.62	0.53	0.80	0.86	0.00	0.00	0.00	0.00
AUIR3M	0.74	0.47	0.72	0.89	0.90	0.84	0.97	0.95	0.00	0.00	0.00	0.00
CAIR3M	0.13	0.01	0.01	0.05	0.83	0.34	0.35	0.48	0.00	0.00	0.00	0.00
CHIR3M	0.88	0.93	0.97	0.98	0.31	0.28	0.50	0.82	0.00	0.00	0.00	0.00
Equity indices												
DJIA	0.53	0.50	0.12	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FTSE	0.10	0.24	0.13	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AUSE	0.49	0.16	0.13	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TSE	0.66	0.69	0.50	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TOPIX	0.17	0.02	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CHSE	0.59	0.70	0.59	0.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Commodities												
WHEAT	0.61	0.26	0.19	0.31	0.87	0.98	1.00	0.93	0.00	0.00	0.00	0.00
SOYBEAN	0.55	0.01	0.02	0.10	0.23	0.63	0.50	0.02	0.00	0.00	0.00	0.00
COFFEE	0.45	0.32	0.23	0.31	0.01	0.01	0.01	0.09	0.00	0.00	0.00	0.00
COTTON	0.26	0.36	0.62	0.67	0.41	0.50	0.26	0.25	0.00	0.00	0.00	0.00
FUEL	0.14	0.05	0.03	0.07	0.31	0.68	0.88	0.98	0.00	0.00	0.00	0.00
GOLD	0.60	0.56	0.38	0.62	0.41	0.70	0.53	0.81	0.00	0.00	0.00	0.00

[a] m denotes the automatically selected lag order based on Escanciano and Lobato (2009).

Chapter 4

Testing for Marginal Asymmetry in Time Series

"Normality is a myth, there never was,
and never will be, a normal distribution."

R. C. Geary, statistician

4.1 Introduction

Symmetry of the marginal law of stochastic processes plays a key role in finance and economics. The following examples illustrate our point.

Asset pricing – A commonly applied option pricing model in finance is a famous Black-Scholes formula developed by Black and Scholes (1973). The pricing model explicitly assumes that equity returns are marginally normally distributed, and, therefore, symmetric. If this underlying assumption fails, for instance, due to the presence of asymmetry in equity returns, then the BS formula very likely misprices options. Therefore, some modifications of the BS formula have been developed in the literature. The methods usually rely on a Gram-Charlier or Edgeworth expansion of some flexible probability distribution, see Jarrow and Rudd (1982) and Corrado and Su (1996), among others. However, Vahamaa (2003) shows that more sophisticated pricing formulae can produce even larger pricing errors as compared to the original

BS formula, mainly due to the numerical difficulties related to more sophisticated methods.

Portfolio management – Standard portfolio diversification in a mean-variance setup, assuming a perfect capital market with risk averse investors, concludes that adding randomly selected and equally weighted assets to a portfolio leads to a risk reduction without any effect on returns, see Markowitz (1968). However, the presence of marginal asymmetry in asset returns can fundamentally change portfolio diversification (i.e. the number of selected assets and their optimal weights), see Conine and Tamarkin (1981) for a theoretical treatment and Prakash et al. (2003) for empirical evidence.

Business cycle fluctuations – Since the Great depression in 1930's, the causes and consequences of business cycle fluctuations have attracted much interest in theoretical and empirical macroeconomics. The presence of marginal asymmetry in economic indicators has implications both for developing new theoretical models in economics and applied macroeconomics. For example, Acemoglu and Scott (1997) build a model where business cycle fluctuations are based on intertemporal increasing returns in the economy. They show that this model specification is helpful in explaining business fluctuations even in the case of IID shocks because individuals respond differently to shocks depending on their past investment activity. Boldin (1999) shows that the effect of monetary policy measures varies over the business cycle. As a result, ignoring business cycle asymmetry can lead to misleading economic policy conclusions in practice.[1]

However, testing for asymmetry in economic time series is by no means easy in practice. Two problems immediately arise. First, economic time series do exhibit some form of weak dependence, which invalidates critical values of standard symmetry tests originally derived for independently and identically distributed (IID)

[1]The interested reader is referred to Psaradakis and Sola (2003) for further details about the implications of marginal asymmetry in economic time series.

random variables. Only quite recently, some symmetry tests have been developed for weakly dependent (WD) stochastic processes as well. Bai and Ng (2005) proposed a symmetry test based on sample skewness, whereas Psaradakis (2003) developed a Kolmogorov-Smirnov test based on a sieve bootstrap. However, it is important to point out that both tests suffer from some shortcomings and their application to economic time series might be problematic. The main drawbacks of the tests are related to either the estimation of some key quantities or strong assumptions about the underlying stochastic process under consideration. For example, it is well known that the estimation of higher-order moments (e.g. the variance-covariance matrix) is more involved for weakly dependent processes. As shown in Andrews (1991), standard variance-covariance estimators perform very poorly for (persistent) WD stochastic processes. As a result, the Bai and Ng (2005) symmetry test suffers from a significant power loss, which, in turn, can lead to misleading inference. The problem with the Kolmogorov-Smirnov test proposed in Psaradakis (2003) is that a sieve bootstrap is theoretically valid only for a limited class of linear models allowing for an AR(∞) representation. Moreover, the procedure might not give satisfactory results for non-IID innovations (e.g. martingale difference sequences) in general. Although the bootstrap-based test gives very satisfactory results even in small samples for models with IID innovations, and clearly outperforms the Bai and Ng test, it is computationally intensive.

Second, economic time series are often contaminated by outliers, see Balke and Fomby (1994) for empirical evidence. The problem is that standard symmetry tests are not robust against outliers, which means that a sufficiently large aberrant observation biases the measure of asymmetry (e.g. the coefficient of skewness), and, thus, can lead to misleading inference, see Bowman and Shenton (1975) for a discussion, and Peiró (1999) or Premaratne and Bera (2005) for Monte Carlo evidence.

The main task of this chapter is to propose a modified test of symmetry based on sample quantiles. It will be shown that the test has an intuitive interpretation, it is easy to calculate, it has a standard limiting distribution, and it is robust against weak dependence of observations. Especially the last feature is very useful for applied research. It will be demonstrated later on in the chapter that the quantile-based specification of the test makes the computation of some key quantities (e.g. the variance-covariance matrix) almost insensitive to dependence of observations. This fact significantly reduces the possibility of inferential errors caused by the incorrect configuration and implementation of the test.

The chapter is organized as follows. Two symmetry tests are discussed in Section 4.2. Monte Carlo setup and results are discussed in Section 4.3 and 4.4. Finally, testing symmetry of financial time series is discussed in Section 4.5.

4.2 Test for Symmetry

4.2.1 Robust Measure of Skewness

Let $\{Y_t : t \in \mathbb{Z}\}$ be a strictly stationary sequence of random variables with the marginal distribution function $F(y) = \mathbb{P}(Y \leq y)$, for any $y \in \mathbb{R}$. The problem of interest is to test the hypothesis that F is symmetric about ζ, the centre of symmetry, that is

$$\mathbb{F}(\zeta + y) = 1 - \mathbb{F}(\zeta - y), \quad \text{for every} \quad y \in \mathbb{R}, \zeta \in \mathbb{R}. \tag{4.1}$$

The test for symmetry explored here relies on a measure of skewness based on selected quantiles of F. Specifically, letting $\xi_p = \inf\{y : F(y) \geq p\}$, $p \in (0,1)$, denote the p-th quantile of F, it is easy to see that $\xi_{1/2} - \xi_p = \xi_{1-p} - \xi_{1/2}$ when (4.1) holds. Motivated by this observation, we consider the following measure of skewness

$$S = \boldsymbol{\delta}'\boldsymbol{\xi}, \tag{4.2}$$

where $\boldsymbol{\xi} = (\xi_{p_1}, \ldots, \xi_{p_k}, \xi_{1/2}, \xi_{1-p_k}, \ldots, \xi_{1-p_1})'$ for some fixed integer $k \geq 1$ and constants $0 < p_1 \ldots p_k < 1$, and $\boldsymbol{\delta} \neq 0$ is a $(2k+1 \times 1)$ is fixed selection vector such that $S = 0$ when F is symmetric.[2] Note that for $k = 1$ and $\boldsymbol{\delta} = (1, -2, 1)'$, S becomes

[2]For example, $\boldsymbol{\delta}_{k+1} = -2$ and $\boldsymbol{\delta}_i = 1/k$ for $i \neq k+1$, with $\boldsymbol{\delta}_i$ denoting the i-th component of a vector $\boldsymbol{\delta}$.

an unscaled version of the measure of skewness considered in Hinkley (1975).

It is easy to show that the proposed measure of symmetry S satisfies basic properties discussed in Groeneveld and Meeden (1984). In particular, the index of skewness $S \equiv S(F)$ of a distribution F defined as in (4.2) satisfies the following properties: (i) $S(F) = S(aF+b)$ for any fixed $b \in \mathbb{R}$ and $a > 0$; (ii) $S(F) = 0$ is F is symmetric; (iii) $S(-F) = -S(F)$; (iv) $S(F) \leq S(G)$ for any distribution G that is at least as skewed to the right as F (i.e. any G distribution such that $G^{-1}(F(y))$ is convex). The verification of the first three properties is quite straightforward, while the verification of the last property ("skew dominance") is more complex, see Groeneveld and Meeden (1984) for a discussion.

Before we proceed to the testing procedure, let us state some necessary assumptions.

Assumption 3 *The process $\{Y_t : t \in \mathbb{Z}\}$ is assumed to be strictly stationary real-valued α-mixing such that $\alpha(n) = O(n^{-\varphi})$, where $\varphi > 3$, and some integer n such that $n \to \infty$.* $\square$

The strong-mixing coefficients $\alpha(n)$, for some integer n such that $n \to \infty$, of a strictly stationary random sequence $\{Y_t : t \in \mathbb{Z}\}$ are defined as follows

$$\alpha(n) = \sup_{\mathcal{A} \in \mathcal{F}_{-\infty}^{0}, \mathcal{B} \in \mathcal{F}_{n}^{\infty}} |\mathbb{P}(\mathcal{A} \cap \mathcal{B}) - \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B})|, \quad \text{for} \quad n \in \mathbb{N},$$

where, for $-\infty \leq r \leq s \leq \infty$, $\mathcal{F}_r^s$ denotes σ-algebra generated by $\{Y_t : r \leq t \leq s\}$. If $\alpha(n) \to 0$ as $n \to \infty$, then the sequence $\{Y_t : t \in \mathbb{Z}\}$ is said to be α-mixing. The α-mixing assumption is an important condition for the validity of the central limit theorem of quantities calculated from weakly dependent observations, see Lehmann (1999, Ch. 2.8) for details. The α-mixing condition is fairly mild and is satisfied by a wide variety of linear and non-linear random processes. Examples of α-mixing processes include, among many others, Markov chains satisfying mild regularity conditions, non-linear processes admitting a Haris ergodic Markovian representation, ARCH type and stochastic volatility processes, q-dependent processes, and linear processes driven by innovations having a continuous distribution which satisfies suitable smoothness conditions, see Doukhan (1994).

Given a sample $\{Y_1, \ldots, Y_T\}$, for $T \geq 1$, a natural estimator of ξ_p is the p-th sample quantile $\hat{\xi}_p = \inf\{y : \hat{F}(y) \geq p\}$, $p \in (0,1)$, where $\hat{F}(y) = (1/T)\sum_t^T I(Y_t \leq y)$, $y \in \mathbb{R}$, is the sample distribution function and $I(\cdot)$ denotes the indicator function. Under α-mixing condition, it is straightforward to show that $\boldsymbol{\xi}$ is consistently estimated by the vector of sample quantiles $\hat{\boldsymbol{\xi}} = (\hat{\xi}_{p_1}, \ldots, \hat{\xi}_{p_k}, \hat{\xi}_{1/2}, \hat{\xi}_{1-p_k} \ldots, \hat{\xi}_{1-p_1})'$. More specifically, putting $\mathcal{P}_k = \{p_1, \ldots, p_k, 1-p_1, \ldots, 1-p_k, 1/2\}$, we have the following limiting result.

Theorem 8 *Suppose that $\alpha(n) \to 0$ as $n \to \infty$ and that, for every $p \in \mathcal{P}_k$, ξ_p is the unique p-th quantile of F. Then, $\hat{\boldsymbol{\xi}} - \boldsymbol{\xi} \xrightarrow{a.s.} 0$ as $T \to \infty$.* □

Proof. *See Appendix A for the proof.* ■

By strengthening the mixing condition somewhat and imposing some smoothness on the marginal distribution of $\{Y_t : t \in \mathbb{Z}\}$, the limiting distribution of $\hat{\boldsymbol{\xi}}$ can be obtained.

Theorem 9 *Suppose that $\sum_{n=1}^{\infty} \alpha(n) < \infty$ and that, for every $p \in \mathcal{P}_k$, F is differentiable at ξ_p with $F'(\xi_p) = f(\xi_p)$. Then $\sqrt{T}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$ as $T \to \infty$, where $\boldsymbol{\Sigma} = [\sigma_{i,j}]_{i,j=1}^{2k+1}$ with*

$$\sigma_{i,j} = \frac{1}{f(\boldsymbol{\xi}_i) f(\boldsymbol{\xi}_j)} \left\{ \gamma_{i,j}(0) + \sum_{h=1}^{\infty} [\gamma_{i,j}(h) + \gamma_{j,i}(h)] \right\}, \tag{4.3}$$

where $\gamma_{i,j}(h) = cov\left[I(Y_1 \leq \boldsymbol{\xi}_i), I(Y_{1+h} \leq \boldsymbol{\xi}_j)\right]$, for $h \geq 0$. □

Proof. *See Appendix A for a proof.* ■

The differentiability condition on F in Theorem 9 is standard in the literature on sample quantiles and not overly restrictive. In fact, asymptotic normality does not hold if F is not differentiable at the quantiles of interest, see Theorem 2 in Sharipov and Wendler (2013).[3] Note also that the summability condition is the best currently available condition for the central limit theorem for bounded random variables.[4]

4.2.2 Test for Symmetry

Since the measure of skewness S given in (4.2) is 0 when F satisfies (4.1), our statistic for testing the hypothesis of marginal symmetry of $\{Y_t\}$ is defined as

$$QS = T\left(\frac{(\hat{S}-S)^2}{\mathrm{var}(\hat{S})}\right) = T(\boldsymbol{\delta}'\hat{\boldsymbol{\xi}})^2/\boldsymbol{\delta}'\hat{\boldsymbol{\Sigma}}\boldsymbol{\delta}, \tag{4.4}$$

where $\hat{\boldsymbol{\Sigma}}$ is a suitable estimator of $\boldsymbol{\Sigma}$. The following result is an immediate consequence of Theorem 9, the continuous mapping theorem, and Slutsky's theorem.

Theorem 10 *Suppose that the conditions in Theorem 9 hold and let $\hat{\boldsymbol{\Sigma}}$ be a consistent estimator of $\boldsymbol{\Sigma}$. Then, $QS \xrightarrow{d} \chi^2(1)$ as $T \to \infty$ under (4.1).* □

Proof. *See Appendix A for a proof.* ■

To make the test operational, a consistent estimator of the asymptotic variance-covariance matrix $\boldsymbol{\Sigma}$ is required in (4.4). Note that the variance-covariance matrix $\boldsymbol{\Sigma}$ is not a diagonal matrix even when we deal with IID observations. Following the literature on the estimation of asymptotic covariance matrices in the presence of weak dependence, we consider here an estimator $\boldsymbol{\Sigma} = [\sigma_{i,j}]_{i,j=1}^{2k+1}$ with

$$\hat{\sigma}_{i,j} = \frac{1}{\hat{f}(\hat{\boldsymbol{\xi}}_i)\hat{f}(\hat{\boldsymbol{\xi}}_j)}\left\{\hat{\gamma}_{i,j}(0) + \sum_{h=1}^{T-1} w(h/m)\left[\hat{\gamma}_{i,j}(h) + \hat{\gamma}_{j,i}(h)\right]\right\}, \tag{4.5}$$

[3] Non-Gaussian weak limits are assumed to be expected for extreme sample quantiles ξ_p with $p \to 0$ or $p \to 1$, see Beirlant (2004).

[4] It is interesting to note that, as in the case of IID data, the error of approximation in the central limit theorem for sample quantiles is of order $O(1/\sqrt{T})$ under suitable polynomial α-mixing condition, see Lahiri and Sun (2009).

with

$$\hat{\gamma}_{i,j}(h) = \frac{1}{T}\sum_{t=1}^{T-h} I(Y_t \leq \hat{\boldsymbol{\xi}}_i)I(Y_{t+h} \leq \hat{\boldsymbol{\xi}}_j) - \frac{1}{T^2}\sum_{t=1}^{T-h} I(Y_t \leq \hat{\boldsymbol{\xi}}_i)\sum_{t=1}^{T-h} I(Y_{t+h} \leq \hat{\boldsymbol{\xi}}_j), \quad (4.6)$$

where $0 \leq h < T$, $w(\cdot)$ are kernel weights, m is a real-valued bandwidth such that $m \to \infty$ and $m/T \to 0$ as $T \to \infty$, and $\hat{f}$ is a consistent estimator of f. Assuming f is a bounded density for F, it is estimated by means of a standard Parzen-Rosenblatt estimator

$$\hat{f}(y) = \frac{1}{bT}\sum_{t=1}^{T} K\left(\frac{y - Y_t}{b}\right), \quad y \in \mathbb{R}, \quad (4.7)$$

where $K(\cdot)$ is a kernel function and $b > 0$ is a bandwidth such that $b \to 0$ and $Tb \to \infty$ as $T \to \infty$.

In view of Theorems 8 and 9, consistency of the estimator in (4.5) follows from well-known resuts on covariance matrix estimation (e.g. Andrews (1991, Theorem 1), Hansen (1992, Theorem 2), and De Jong (2000, Theorem 2), among others), combined with uniform consistency of the kernel estimator $\hat{f}$. Since $I(Y_t \leq \boldsymbol{\xi}_i)$ is bounded, a mixing rate $\alpha(n) = O(n^{-\beta})$ for some $\beta > r/2$ and $r \in (2, 4]$, coupled with $m = o(T^{1/2-1/r})$, is sufficient for $\hat{\gamma}_{i,j}(0)+\sum_{h=1}^{T-1} w(h/m)\,[\hat{\gamma}_{i,j}(h) + \hat{\gamma}_{j,i}(h)]$ to converge in probability to $\gamma_{i,j}(0) + \sum_{h=1}^{\infty} [\gamma_{i,j}(h) + \gamma_{j,i}(h)]$, see De Jong (2000). Regular conditions which ensure uniform consistency of the kernel estimator in (4.7) can be found in Cai and Roussas (1992, Theorem 4.1), Liebscher (1996, Theorem 4.2), and Hansen (2008, Theorem 6), among others. A polynomial mixing rate $\alpha(n) = O(n^{\varphi})$ with $\varphi > 3$ and some smoothness conditions for f are typically sufficient for such results to hold. A set of additional conditions under which $\hat{f}(\hat{\boldsymbol{\xi}}_i)$ converges almost surely to $f(\boldsymbol{\xi}_i)$ is stated in the following assumption.

Assumption 4 *The kernel function $K(\cdot)$ is assumed to possess the following properties:*

(a) $K(u) \geq 0$ *for any* $u \in \mathbb{R}$ *and differentiable,*

(b) $\int_{\mathbb{R}} K(u)\mathrm{d}u = 1,$

(c) $\int_{\mathbb{R}} |u|K(u)\mathrm{d}u < \infty$,

(d) $\int_{\mathbb{R}} |K'(u)|\mathrm{d}u < \infty$,

(e) $\lim_{|u|\to\infty} K(u) = 0$,

(f) as $T \to \infty$*, the bandwidth parameter* b *is assumed to be a positive quantity such that: (i)* $b \to 0$*; (ii)* $b^2T/\log\log T \to \infty$*.* □

Under Assumptions 3 – 4, it holds that $\hat{f}(y) - f(y) \xrightarrow{a.s.} 0$ as $T \to \infty$, uniformly in $y \in \mathbb{R}$, by Theorem 4.1 of Cai and Roussas (1992). Together with the result in Theorem 8, this implies that $\hat{f}(\hat{\xi}_p) - f(\xi_p) \xrightarrow{a.s.} 0$ as $T \to \infty$ for every $p \in \mathcal{P}_k$.

Needless to say, there are many choices available for suitable kernels $w(\cdot)$ and $K(\cdot)$ that may be used in (4.5) and (4.7). The differences in the resulting estimators are not generally substantial in finite samples (see, e.g., Andrews (1991), Silverman (1986)), and we will take $w(\cdot)$ and $K(\cdot)$ in the remainder of this paper to be the Bartlett kernel and the Gaussian kernel, respectively, i.e., $w(x) = (1-|x|)I(|x| \leq 1)$ and $K(x) = \exp(-x^2/2)/\sqrt{2\pi}$. With regard to the bandwidth parameters m and b, the former will be selected by means of automatic data-dependent method of Newey and West (1994). For the latter, we will use the popular normal reference bandwidth $b = 0.79(\xi_{3/4} - \xi_{1/4})T^{-1/5}$ discussed in Silverman (1986, Section 3.4.2).[5]

Finally, we note that, instead of relying on covariance estimators of the type given in (4.5), a bootstrap estimator of the asymptotic covariance matrix $\mathbf{\Sigma}$ could be used. Sun and Lahiri (2006) showed that, under a polynomial α-mixing rate and mild smoothness conditions on F, the asymptotic variance of a sample quantile can be

[5] Of course, many other methods for selecting the bandwidth b are available in the literature (see, e.g., Jones et al. (1996). We found in our simulations that the finite-sample properties of the test for symmetry based on the QS test are fairly robust with respect to different data-dependent bandwith selection methods, and so we focus here on the computationally simple normal reference bandwidth selector. It is also worth noting that bandwidth selectors designed for IID data often work equally well under dependence (see, e.g., Hall et al. (1995).

consistently estimated by a blockwise bootstrap method. The blockwise bootstrap may also be used to obtain a consistent estimator of the distribution of a sample quantile, see Sun and Lahiri (2006) and Sharipov and Wendler (2013), among others. Such techniques could be adapted to the problem of testing symmetry using the statistic in (4.4), but bootstrap-based versions of our test will not be investigated here.

4.2.3 Simple QS Test

One of the main tasks of this paper is to show that the proposed quantile-based symmetry test is robust against weak dependence of observations observed in economic time series. An intuitive explanation can be found in Figure 4.1. The figure depicts selected sample quantiles from the empirical distribution function (EDF) and the corresponding points in the realization of a given stochastic process (i.e. the DGP is an AR(1) process with standard Gaussian innovations). The robustness of the test follows from the fact that individual sample quantiles are well separated over time, and, thus, might be considered as quantities from an "almost" uncorrelated stochastic process. This example helps to understand why it might be possible to approximate the variance-covariance matrix for WD processes by the variance-covariance matrix for IID processes. The figure also clearly shows the trade-off between the number of quantiles and their dependence: the more quantiles the higher the dependence.[6] In order to explicitly demonstrate the robustness of the test, a modified symmetry test, denoted as QS*, is considered as well. The test statistic is given by

$$QS^* = T(\boldsymbol{\delta}'\hat{\boldsymbol{\xi}})^2/\boldsymbol{\delta}'\hat{\boldsymbol{\Sigma}}^*\boldsymbol{\delta}, \tag{4.8}$$

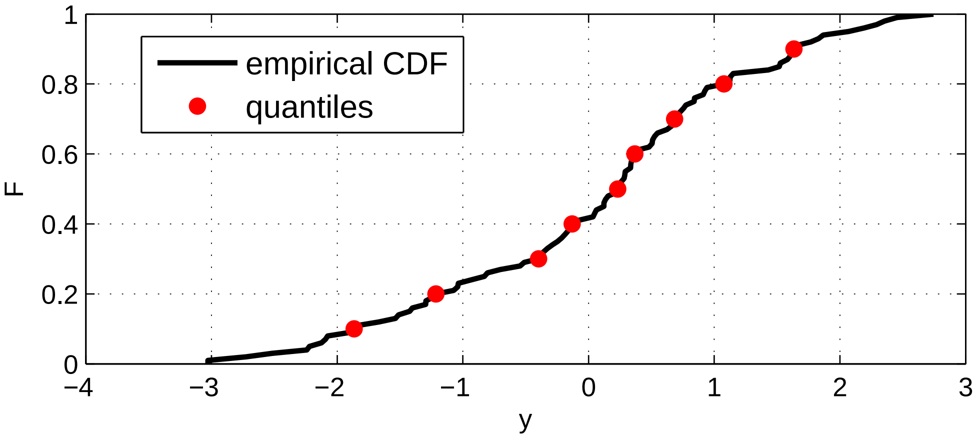
where $\hat{\boldsymbol{\Sigma}}^* = [\sigma^*_{i,j}]_{i,j=1}^{2k+1}$ is the estimated variance-covariance matrix for IID observations given by

$$\hat{\sigma}^*_{i,j} = \frac{p_i(1-p_j)}{\hat{f}(\hat{\boldsymbol{\xi}}_i)\hat{f}(\hat{\boldsymbol{\xi}}_j)}, \quad \text{for} \quad i \leq j. \tag{4.9}$$

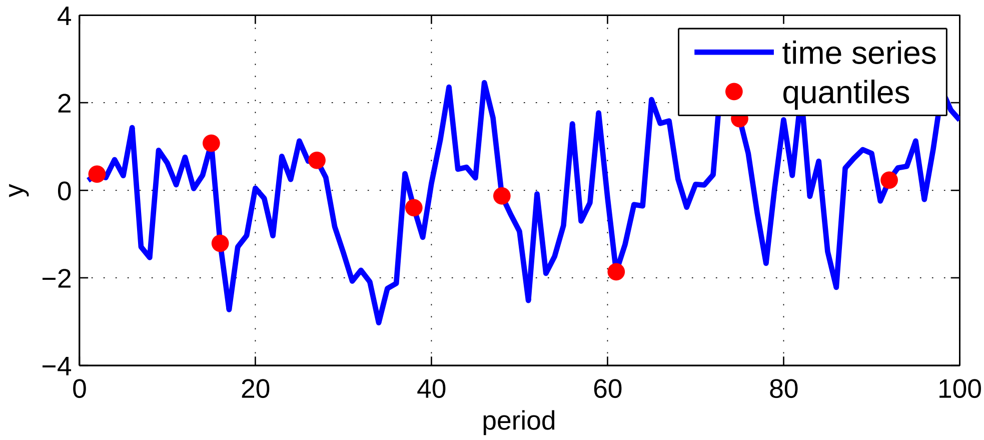
The estimation of $\hat{f}$ is described in the previous section. Recall that the modified variance-covariance matrix is strictly correct for IID observations and only an

[6] We leave the issue of optimal number of quantiles for further research.

Figure 4.1 Empirical quantiles



(a) empirical CDF and estimated quantiles



(b) time series and estimated quantiles

Note: the data generating process is $y_t = 0.5y_{t-1} + a_t$, where $a \sim NID(0, 1)$.

approximation for WD observations.

4.3 Monte Carlo Setup

The size and power properties of the proposed quantile symmetry test are assessed using a list of time series models usually used in applied economics/finance: (i) models M1–M3 represent autoregressive and moving average (ARMA) models; (ii) models H1–H3 denote autoregressive conditional heteroscedasticity (GARCH) models with a different functional form; (ii) models N1–N3 denote mixture autoregressive (MAR) models with a different functional form. A complete set of models can be found in Table 4.1.

Table 4.1 Time series models for comparison

M1 model: $y_t = a_t$

M2 model: $y_t = 0.5y_{t-1} + a_t$

M3 model: $y_t = 0.8y_{t-1} - 0.5a_{t-1} + a_t$

H1 model: $y_t = 1 + 0.5y_{t-1} + \epsilon_t$, $\epsilon_t = a_t\sqrt{h_t}$, $h_t = 0.4 + 0.1\epsilon_{t-1}^2 + 0.5h_{t-1}$

H2 model: $y_t = 1 + 0.5y_{t-1} + \epsilon_t$, $\epsilon_t = a_t\sqrt{h_t}$, $\log h_t = 0.4 + 0.1a_{t-1}^2 + 0.5\log h_{t-1}$

H3 model: $y_t = 1 + 0.5y_{t-1} + \epsilon_t$, $\epsilon_t = a_t\sqrt{h_t}$, $h_t = 0.4 + 0.1y_{t-1}^2 + 0.5h_{t-1}$

N1 model: $y_t = (-2.0 + 0.5y_{t-1} + a_t)I(S_t = 1) + (0.5y_{t-1} + a_t)I(S_t = 2)$

N2 model: $y_t = (-2.0 + 0.75y_{t-1} + a_t)I(S_t = 1) + (0.25y_{t-1} + a_t)I(S_t = 2)$

N3 model: $y_t = (-2.0 + 0.5y_{t-1} + \sqrt{2}a_t)I(S_t = 1) + (0.5y_{t-1} + a_t)I(S_t = 2)$

Since various DGPs are considered in this paper, some comments are in order. The marginal distribution of ARMA models is asymmetric, provided that the distribution of innovations is asymmetric. Asymmetry of the marginal distribution of GARCH processes is caused by asymmetry of model innovations.[7] In contrast, asymmetry in the marginal law of MAR models can be caused by a combination of different model parameters such as regime constants, autoregressive parameters, and regime variances, see Amendola et al. (2006). Note that all the above mentioned time series models satisfy the necessary strong-mixing condition under relatively mild assumptions, see Chanda (1974, p. 403) for ARMA models and Francq and Zakoïan

[7]Note, however, that the closed form solution of the marginal distribution is not known for GARCH processes in general. The closed-form distribution is available only for specific GARCH-type models such as moving average conditional heteroscedastic processes, see Yang and Bewley (1995) for details.

(2006, p. 822) for GARCH models, Liebscher (2005, p. 680) for NLAR models (and therefore MAR models as well). In order to meet the moment condition of the test, restrictions on some parameters of DGPs must be imposed. In particular, we set the model parameters and the distributions of innovations in such a way that $\mathbb{E}(|Y_t|^4) < \infty$.

The power properties of the proposed robust symmetry tests are examined on various distributions of innovations. In particular, apart from a Gaussian distribution, which serves as a benchmark for comparison, we consider model innovations coming from a generalized lambda distributions (GLD), see Randles et al. (1980). This family provides a wide range of distributions that are easily generated, since they are defined in terms of the inverse of the cumulative distribution functions: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, for $0 \le u \le 1$ and $\lambda_j \in \mathbb{R}$ for $j \in \{1, \ldots, 4\}$. Particular parameters of a generalized lambda family used in Monte Carlo experiments come from Bai and Ng (2005) and can be found in Table 4.2.

Table 4.2 Parameters of a generalized lambda distribution

	λ_1	λ_2	λ_3	λ_4	skewness	kurtosis	moment
S1	0.000000	-1.000000	-0.080000	-0.080000	0.0	6.0	12
S2	0.000000	-0.397912	-0.160000	-0.160000	0.0	11.6	6
A1	0.000000	1.000000	-0.007500	-0.030000	-1.5	7.5	10
A2	0.000000	1.000000	-0.100900	-0.180200	-2.0	21.1	5

[a] Note that a standard normal distribution can also be approximated by a generalized lambda distribution with the following parameters: $\lambda_1 = 0$, $\lambda_2 = 0.1975$, $\lambda_3 = \lambda_4 = 0.1349$.

[b] The highest possible moment available for a random variable drawn from a given distribution.

Originally, T+100 observations in each experiment are generated, but first 100 of them are discarded in order to eliminate the effect of the initial observations. The number of repetitions of all experiments is set to $R = 1000$ and the number of observations is set to $T \in \{200, 500, 1000\}$. A particular specification of the quantiles is $p_1 = 0.05$, $p_2 = 0.15$, $p_3 = 0.25$, and $p_4 = 0.35$. The configuration of the sample quantiles reflects: (i) 5 % trimming of extreme values from each tail; and (ii) the number of equally spaced sample quantiles ($k = 4$)

is set based on preliminary Monte Carlo experiments.[8] The selection vector is $\boldsymbol{\delta} = (1/4, 1/4, 1/4, 1/4, -2, 1/4, 1/4, 1/4, 1/4)'$. The probability of switching of MAR models is set to $\mathbb{P}(S_t = 1) = 0.25$ and $\mathbb{P}(S_t = 2) = 0.75$.

4.4 Monte Carlo Results

The Monte Carlo results are presented in Tables 4.5 – 4.7. For each DGP and the configuration of the distribution of model innovations, the average rejection frequency is reported as follows

$$avg = \frac{1}{R}\sum_{j=1}^{R} I(\hat{\alpha} \leq \alpha), \tag{4.10}$$

where R denotes the number of repetitions, $I(\cdot)$ is a standard indicator function, α is the statistical significance of the test set to 0.05, and $\hat{\alpha}$ is the estimated p-value of the test. The results reveal the following: (i) The proposed tests have good size properties, regardless of the sample size (i.e. M1, H1, H2, H3 models with N, S1, S2 innovations). Nevertheless, a small size distortion is observed for more complex ARMA specifications (i.e. M2 and M3 models) with heavy tailed symmetric innovations (i.e. S1 and S2). However, it is important to emphasize that the distortion is not of the magnitude to make the tests unattractive; (ii) The tests have very good power properties, provided that kurtosis does not dominate the stochastic properties of innovations. This fact can be illustrated using, for example, M2 model. In this case, the average rejection frequency of the QS test is 0.98 for A1 configuration of innovations and the sample size $T = 200$. However, once kurtosis dominates the stochastic properties of innovations, a power loss of the tests is observed in small samples. In this case, the average rejection drops from 0.98 to 0.50. However, it is worth pointing out that the power of the test quickly improves as the sample size increases. The average rejection frequency increases from 0.50 in the sample $T = 200$ to 0.88 in the sample $T = 500$; (iii) Significant differences are observed in the behaviour of the tests for two non-linear models (i.e. GARCH and MAR models). As for the GARCH models, asymmetry generated by

[8] Note that we leave the issue of optimal number of quantiles for further research.

the functional form of the volatility component is, as might be expected, negligible as compared to the effect of model innovations. In contrast, asymmetry in MAR models can be generated by many different parameter configurations. The results indicate that the tests have very good power for most of the configurations even in small samples, provided that asymmetry is generated by at least two parameters of a MAR model (e.g. regime constants and AR parameters). Of course, the power of the tests significantly improves with asymmetry of innovations. For example, the average rejection frequency of N2 model ranges from 0.86 to 1.00 in the sample $T = 200$, regardless of the distribution of innovations; (iv) Finally, no significant differences in the size and power properties are found between the QS test, a test with the correctly estimated variance-covariance matrix $\mathbf{\Sigma}$, and the QS* test, a test with the approximated variance-covariance matrix $\mathbf{\Sigma}^*$. The fact that the researcher does not have to consider the correct estimation of the variance-covariance matrix and can easily use the approximated one makes the QS* test very attractive for applied research.

4.5 Comparison With Other Tests

The performance of the proposed QS test is compared with two other test statistics mentioned earlier: (i) a skewness-based symmetry test developed by Bai and Ng (2005), denoted as NBS; and (ii) a bootstrap-based Kolmogorov-Smirnov test developed by Psaradakis (2003), denoted as BKS. The data generating processes considered for comparison can be found in Table 4.1. Models M1 and M2 are considered in Bai and Ng (2005), whereas model M3 in Psaradakis (2003). All results are based on $R = 1000$ replications.

The Monte Carlo results are presented in Table 4.3. The average rejection frequency is reported in the table. The results reveal the following: (i) No differences are noticed in the size properties of all tests (i.e. DGP configurations M1, M2, M3 and symmetric innovations NID(0,1), S2, $t(5)$). This means that under the null hypothesis of symmetry, all tests perform similarly; (ii) The power results of the QS test are significantly better as compared to the NBS test (i.e. DGP configurations M1, M2 and asymmetric innovations A2), and approximately similar as compared to the BKS test (i.e. DGP configuration M3 with asymmetric innovations $\chi^2(4)$). Our results clearly suggest that, in situations where the user is not proficient in bootstrap or it is not clear which bootstrap method should be implemented, the quantile-based symmetry test may serve as a valuable alternative when testing for marginal asymmetry. Moreover, the computation of the quantile test is considerably easier and faster as compared to any bootstrap test.

Table 4.3 Comparison of marginal symmetry tests: $T = 200$

	NID(0,1)		S2		A2	
model	NBS	QS	NBS	QS	NBS	QS
M1	0.05	0.06	0.04	0.06	0.43	0.68
M2	0.04	0.04	0.04	0.07	0.37	0.50
	NID(0,1)		$t(5)$		$\chi^2(4)$	
model	BKS	QS	BKS	QS	BKS	QS
M3	0.04	0.06	0.04	0.06	0.86	0.90

[a] BKS denotes a bootstrap Kolmogorov-Smirnov test discussed in Psaradakis (2003), NBS denotes a symmetry test based on a coefficient of skewness discussed in Bai and Ng (2005), QS denotes a test for marginal symmetry based on quantiles.
[b] The significance level is set to $\alpha = 0.05$.

4.6 Empirical Example

In this section, the quantile marginal symmetry tests QS and QS* (i.e. the correct one and its approximation) are applied to a set of 22 monthly economic time series spanning the period January 1980 and December 2010. A detailed description of time series can be found in Table 4.8. There are two main tasks of this exercise. First, we are interested in how consistent results are obtained from both robust symmetry tests in practice. That means whether both marginal symmetry tests lead to the same conclusion about rejecting and/or not rejecting the null hypothesis. Second, we are interested in answering a question whether asymmetry is a characteristic feature of some types of economic indicators (e.g. equity returns), or whether it is a purely series-dependent feature. This issue is of much practical importance in finance for both asset pricing and/or risk management when constructing and evaluating financial portfolios.

Table 4.4 Testing marginal symmetry: period 1980M1 – 2010M12

variable	QS	QS*	variable	QS	QS*
Exchange rates			**Interest rates**		
USDGBP	0.64	0.65	USIR3M	0.07	0.02
USDJPY	0.01	0.01	UKIR3M	0.60	0.51
USDCAD	0.59	0.52	CAIR3M	0.09	0.07
USDAUD	0.09	0.08	AUIR3M	0.17	0.08
USDCHF	0.07	0.06	CHIR3M	0.94	0.92
Equities			**Commodities**		
DJIA	0.58	0.56	WHEAT	0.55	0.55
TOPIX	0.25	0.27	SOYBN	0.79	0.80
FTUK	0.81	0.79	COFFEE	0.40	0.37
TSE	0.01	0.01	COTTON	0.94	0.93
AUSE	0.42	0.39	FUEL	0.86	0.86
CHSE	0.00	0.01	GOLD	0.08	0.10

The estimated p-values of the QS tests are presented in Table 4.4. The results suggest the following: (i) Both tests lead to the same conclusion (i.e. rejecting or not rejecting the null of symmetry) in 21 out of 22 cases at the significance level 0.10. This finding fully supports the Monte Carlo results that the QS* test, based on the approximated variance-covariance matrix $\mathbf{\Sigma}^*$, produces almost identical results as compared to the QS test, based on the correctly estimated variance-covariance matrix $\mathbf{\Sigma}$; (ii) The null hypothesis of symmetry is rejected in 8 of 22 cases by the QS test (i.e. in 36 % of cases); (iii) However, noticeable differences are observed among various classes of financial assets. For example, the null hypothesis is rejected for 3 out of 5 exchange rate returns (i.e. in 60 % of cases), whereas for just 1 out of 6 commodities (i.e. in 17 % of cases). Much more importantly, the difference in the rejection frequencies of exchange rate and commodity returns is statistically significant at the nominal level 0.10.[9] Therefore, it can be concluded that the degree

[9] Any other differences in rejection frequencies are not statistically significant at the nominal

of asymmetry varies across asset classes.

4.7 Conclusion

A modified quantile-based symmetry test of the marginal law of stationary stochastic processes has been introduced in this chapter. It has been shown that the test is intuitive, easy to calculate, follows standard limiting distribution, and much more importantly, it is robust against weak dependence of observations. Especially the last feature makes the test very attractive for applied research since it reduces the inferential errors coming from the incorrect estimation of the key quantities of the test. Monte Carlo results suggest that the finite sample properties of the QS test significantly outperforms the skewness-based symmetry test and compares favorably with the bootstrap-based Kolmogorov-Smirnov test. So, it can be concluded that in situations where the user is not proficient in bootstrap, or it is not clear which bootstrap method should be implemented, the QS test may serve as a valuable alternative when testing for marginal asymmetry. The Monte Carlo and empirical results confirm that both the original QS and the simplified QS* tests produce very similar results.

Both tests indicate that a marginal distribution of approximately one third of all asset returns considered here is statistically significantly asymmetric at the nominal level 0.10. Our results may be directly employed in finance for portfolio and risk management. For example, according to Basel banking regulations (see Jorion (2007), among others), commercial banks are required to measure market risk of their asset portfolios and to hold capital in proportion to their risk position. As a result, banks constructing portfolios from highly negatively skewed assets can be systematically exposed to higher downside risk and required to hold more (cash) reserves, which would reduce overall profitability of banks.

level 0.10.

4.8 Appendix A: Proofs

It is assumed that all conditions in Assumptions 3 – 4 are implicitly satisfied.

4.8.1 Proof of Theorem 8

Since $\xi_p \in \boldsymbol{\xi}$ for any $p \in \mathcal{P}_k$, it is fully sufficient to show that $\hat{\xi}_p - \xi_p \xrightarrow{a.s.} 0$ as $T \to \infty$. Noting that the process $\{Y_t\}$ is ergodic when $\lim \alpha(n) = 0$ as $n \to \infty$, see White (2001, Proposition 3.44), we have $\hat{F}(y) - F(y) \xrightarrow{a.s.} 0$ as $T \to \infty$, uniformly in $y \in \mathbb{R}$, on account of Glivenko-Cantelli theorem for stationary and ergodic processes (e.g. Dehling and Philipp (2002, Theorem 1.1)). The assertion then follows by a standard argument about mapping between a distribution function F and a quantile ξ_p (e.g. Serfling (1980, p. 75)).

4.8.2 Proof of Theorem 9

Under the conditions of the theorem, for every $p \in \mathcal{P}_k$, ξ_p admits the following Bahadur representation

$$\hat{\xi}_p - \xi_p = \frac{p - \hat{\mathbb{F}}(\xi_p)}{f(\xi_p)} + R = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{p - I(Y_t \leq \xi_p)}{f\xi_p)}\right) + R, \tag{4.11}$$

where $R = o(1/\sqrt{T})$ as $T \to \infty$, see Theorem 1 in Sharipov and Wendler (2013). Hence, we have

$$\sqrt{T}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{Z}_t + o_p(\mathbf{1}), \tag{4.12}$$

where

$$\mathbf{Z}_t = \left(\frac{p_1 - I\left(Y_t \leq \xi_{p_1}\right)}{f(\xi_{p_1})}, \ldots, \frac{1 - p_1 - I\left(Y_t \leq \xi_{1-p_1}\right)}{f(\xi_{1-p_1})}\right)'.$$

So it remains to establish asymptotic normality of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{Z}_t$. By the properties of $\{Y_t\}$, for any $p \in \mathcal{P}_k$, $p - I(Y_t \leq \xi_p)$ is a strictly stationary sequence of bounded random variables with α-mixing coefficients of the same size as $\alpha(n)$, see Theorem 3.49 in White (2001), and $\mathbb{E}(p - I(Y_t \leq \xi_p)) = p - F(\xi_p) = 0$ on account of the assumed continuity of F at ξ_p. Thus, by an application of the central limit theorem for stationary α-mixing processes, see Theorem 18.5.4 in Ibragimov and Linnik (1971), we may conclude that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[p - I(Y_t \leq \xi_p)]/f(\xi_p)$ has a limiting normal distribution with mean zero and variance $f^{-2}(\xi_p)\sum_{h=-\infty}^{\infty}\text{cov}[I(Y_1 \leq \xi_p), I(Y_{1+h} \leq \xi_p)]$. By considering arbitrary linear combination of the components of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{Z}_t$ and applying the central limit theorem the required result follows via the Cramér-Wold device.

4.8.3 Proof of Theorem 10

It follows from Theorem 9 that $\sqrt{T}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\hat{\boldsymbol{\xi}}$ is a $(2k+1 \times 1)$ vector of sample quantiles and $\boldsymbol{\Sigma}$ is a positive definite variance-covariance matrix. It can be shown that there exists a lower triangular matrix $\mathbf{P}$ such that $\boldsymbol{\Sigma} = \mathbf{P}\mathbf{P}'$, see Theorem 4.3 in Schott (2005, p. 139). Then, the standardized vector $\hat{\mathbf{z}} = \mathbf{P}^{-1}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})$ is distributed as $\sqrt{T}\hat{\mathbf{z}} \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$. Using the standardized vector of quantiles $\hat{\mathbf{z}}$, the QS test statistic can be formally written in the following quadratic form

$$QS = T\left(\hat{\mathbf{z}}'\mathbf{A}\hat{\mathbf{z}}\right), \tag{4.13}$$

where $\mathbf{A} = \boldsymbol{\delta}\boldsymbol{\delta}'$. Then the limiting $\chi^2(1)$ distribution immediately follows from Theorem 9.8 in Schott (2005, p. 378) about the limiting distribution of a quadratic form of standard normal random variables. The degrees of freedom follow from the fact that $\mathbf{A}$ is a matrix with $\text{rk}(\mathbf{A}) = 1$.

4.9 Appendix B: Tables

Table 4.5 Statistical properties: M models

		T=200		T=500		T=1000	
DGP	distr.	QS	QS*	QS	QS*	QS	QS*
M1	N	0.06	0.05	0.05	0.05	0.05	0.05
	S1	0.05	0.05	0.05	0.05	0.05	0.05
	S2	0.06	0.05	0.07	0.07	0.05	0.05
	A1	0.99	0.99	1.00	1.00	1.00	1.00
	A2	0.68	0.68	0.98	0.98	1.00	1.00
M2	N	0.04	0.05	0.06	0.07	0.05	0.06
	S1	0.07	0.08	0.05	0.07	0.06	0.08
	S2	0.07	0.09	0.07	0.09	0.06	0.09
	A1	0.98	0.99	1.00	1.00	1.00	1.00
	A2	0.50	0.56	0.88	0.92	0.99	1.00
M3	N	0.06	0.06	0.05	0.06	0.05	0.06
	S1	0.06	0.07	0.06	0.07	0.06	0.08
	S2	0.06	0.07	0.06	0.07	0.06	0.08
	A1	0.98	0.98	1.00	1.00	1.00	1.00
	A2	0.51	0.53	0.86	0.89	0.99	1.00

[a] QS denotes a marginal symmetry test with the correct variance-covariance matrix $\boldsymbol{\Sigma}$, whereas QS* denotes a marginal symmetry test with the approximated variance-covariance matrix $\boldsymbol{\Sigma}^*$.

Table 4.6 Statistical properties: H models

		T=200		T=500		T=1000	
DGP	distr.	QS	QS*	QS	QS*	QS	QS*
H1	N	0.06	0.06	0.06	0.06	0.04	0.05
	S1	0.06	0.06	0.05	0.08	0.05	0.07
	S2	0.08	0.09	0.05	0.08	0.05	0.08
	A1	0.97	0.98	1.00	1.00	1.00	1.00
	A2	0.50	0.49	0.86	0.86	0.99	0.99
H2	N	0.06	0.07	0.04	0.05	0.06	0.07
	S1	0.06	0.07	0.06	0.07	0.05	0.08
	S2	0.08	0.10	0.06	0.07	0.07	0.09
	A1	0.98	0.99	1.00	1.00	1.00	1.00
	A2	0.52	0.52	0.86	0.85	0.99	0.99
H3	N	0.06	0.07	0.07	0.08	0.05	0.07
	S1	0.06	0.08	0.06	0.08	0.06	0.09
	S2	0.06	0.07	0.06	0.09	0.04	0.08
	A1	0.95	0.98	1.00	1.00	1.00	1.00
	A2	0.44	0.48	0.78	0.84	0.97	0.99

[a] QS denotes a marginal symmetry test with the correct variance-covariance matrix $\mathbf{\Sigma}$, whereas QS* denotes a marginal symmetry test with the approximated variance-covariance matrix $\mathbf{\Sigma}^*$.

Table 4.7 Statistical properties: N models

		T=200		T=500		T=1000	
DGP	distr.	QS	QS*	QS	QS*	QS	QS*
N1	N	0.14	0.15	0.30	0.32	0.53	0.55
	S1	0.22	0.23	0.44	0.47	0.72	0.75
	S2	0.25	0.27	0.53	0.55	0.79	0.82
	A1	0.67	0.63	0.97	0.96	1.00	1.00
	A2	0.49	0.49	0.86	0.87	0.99	0.99
N2	N	0.86	0.86	1.00	1.00	1.00	1.00
	S1	0.95	0.96	1.00	1.00	1.00	1.00
	S2	0.97	0.98	1.00	1.00	1.00	1.00
	A1	1.00	1.00	1.00	1.00	1.00	1.00
	A2	1.00	1.00	1.00	1.00	1.00	1.00
N3	N	0.42	0.45	0.79	0.81	0.98	0.98
	S1	0.44	0.47	0.81	0.83	0.98	0.98
	S2	0.45	0.47	0.84	0.86	0.99	0.99
	A1	0.88	0.89	1.00	1.00	1.00	1.00
	A2	0.73	0.75	0.99	0.99	1.00	1.00

[a] QS denotes a marginal symmetry test with the correct variance-covariance matrix $\mathbf{\Sigma}$, whereas QS* denotes a marginal symmetry test with the approximated variance-covariance matrix $\mathbf{\Sigma}^*$.

Table 4.8 A description of monthly time series: sample 1980M1 – 2010M12

variable	description	transformation
Exchange rates		
USDGBP	the US dollar to British pound exchange rate	$\Delta \log$
USDJPY	the US dollar to Japanese yen exchange rate	$\Delta \log$
USDCAD	the US dollar to Canadian dollar exchange rate	$\Delta \log$
USDAUD	the US dollar to Australian dollar exchange rate	$\Delta \log$
USDCHF	the US dollar to Swiss frank exchange rate	$\Delta \log$
Equities		
DJIA	the US Dow Jones Industrials Share Index	$\Delta \log$
UKFT	the UK FT All Shares Index	$\Delta \log$
TOPIX	Tokyo Stock Exchange Index	$\Delta \log$
TSE	Toronto Stock Exchange Index	$\Delta \log$
AUSE	Australian Stock Exchange Index	$\Delta \log$
CHSE	Swiss Stock Exchange Index	$\Delta \log$
Commodities		
WHEAT	Kansas wheat, hard, cents/bushel	$\Delta \log$
SOYBEAN	soybeans, yellow, cents/bushel	$\Delta \log$
COFFEE	Brazilian coffee beans, cents/pound	$\Delta \log$
COTTON	cotton, cents/pound	$\Delta \log$
FUEL	fuel oil, cents/gallon	$\Delta \log$
GOLD	gold bullion, USD/troy ounce	$\Delta \log$
Interest rates		
USIR	the US interest rates, 3M	Δ
UKIR	the UK interest rates, 3M	Δ
CAIR	the Canadian interest rates, 3M	Δ
AUIR	the Australian interest rates, 3M	Δ
CHIR	the Swiss interest rates, 3M	Δ

[a] Monthly averages of daily observations.
[b] Δ denotes a first difference of a given series, $\Delta \log$ is an approximation for the growth rate of a given time series.

Chapter 5

Testing for Non-linearity in Multivariate Time Series

"Another criticism of standard significance tests is that in most applications it is known beforehand that the null hypothesis cannot be exactly true."

W. Kruskal, statistician

5.1 Introduction

Non-linear time series analysis has become a progressively growing part of statistics in the last decades. The main reason for its popularity lies in the fact that non-linear models are capable to capture characteristic features observed in stochastic processes (e.g. regime switching, time-varying volatility, etc.) which cannot be adequately accounted for by any linear models, see Tong (1990) and Granger and Teräsvirta (1993), among others. Although the mainstream literature has focused mainly on univariate non-linear models, there are situations where a set of variables are dependent in a non-linear way. Therefore, increasing attention has been paid to non-linear multivariate models. Many examples can be found in economics and finance, see Sims and Zha (2006), Liu et al. (2009), Schmitt-Grohe and Uribe (2004), Rudebusch and Swanson (2008), Engle and Kroner (1995), or Bollerslev (1990), among others. Despite recent advances in the computer science, the identification, estimation, and forecasting from multivariate non-linear models is still very computationally intensive.[1] Therefore, it is desirable to test for non-linearity in the first place.

However, it is by no means easy to test for non-linearity in multivariate time series models. Two problems immediately arise. First, although there exist many univariate non-linearity tests in the literature, see Chapter 2 for a survey, there is no guarantee that the univariate tests can adequately capture non-linearity in multivariate processes. Therefore, the use of multivariate tests seems to be more appropriate. Second, unfortunately, there are only a few multivariate tests available in the literature. And those existing tests often suffer from a dimensionality problem. It means that, due to the construction of multivariate tests, they require a large number of observations, which is not feasible to get in practice.[2]

There are two main tasks of this chapter. First, it will be shown that the dimensionality issue of two selected multivariate tests (the TSAY and ARCH tests) can be easily bypassed by means of a principal component analysis. Although principal components can reduce, or even completely eliminate, a dimensionality problem, there is still an ultimate question of how many components to retain for a test. Therefore, special attention is paid to the finite sample properties of new principal component-based multivariate tests under different stopping rules. Second, we show, by means of Monte Carlo experiments, that univariate tests can completely fail when testing for non-linearity in multivariate time series systems.

The chapter is organized as follows. A brief description of two multivariate nonlinearity tests is given in Section 5.2. The Monte Carlo setup and results are discussed in Sections 5.3 and 5.4. Two empirical examples are provided in Section 5.5. Section 5.6 concludes and summarizes our results.

[1] The interested reader is referred to Bauwens et al. (2006) for details about the computational issues of some multivariate time series models.

[2] Examples are provided in the next section.

5.2 Non-linearity Testing

5.2.1 Why Multivariate Tests?

Many routinely applied non-linearity tests are the so called neglected non-linearity tests. It means that non-linearity is inspected from residuals obtained from a linear filter. The most often applied filter in the time series literature is an ARMA model. Although this type of models may be perfectly reasonable for exogenous stochastic processes (e.g. sunspots), it might be questionable for economic time series, which are dependent (co-integrated/correlated) in nature. The problem is that applying linear ARMA models to individual components of a multivariate time series process is subject to misspecification (due to omitting some explanatory variables). This fact can lead to size and/or power distortions of the standard non-linearity tests discussed in Chapter 2, which in turn, may lead to misleading inference. In order to make this point clear, we consider the following two examples.

A size distortion: Let us consider a bivariate stationary VAR(1) model $\mathbf{x}_t = \boldsymbol{\xi}_1 \mathbf{x}_{t-1} + \mathbf{a}_t$, where $\mathbf{x}_t = (X_{1t}, X_{2t})'$ and $\mathbf{a}_t = (a_{1t}, a_{2t})'$ is a vector of model innovations such that $a_{it} \sim IID(0, \sigma_i^2)$, for $i \in \{1, 2\}$, and innovations are independent each other, regardless of the time index. In addition to that, let us consider a simple, yet very general, definition of (conditional) linearity used in Lee et al. (1993). According to their definition, the process $\mathbf{x}_t$ is called linear in the conditional mean if and only if

$$\mathbb{P}(\mathbb{E}(\mathbf{x}_t|\mathbf{x}_{t-1}) = \boldsymbol{\Upsilon}_0 + \boldsymbol{\Upsilon}_1 \mathbf{x}_{t-1}) = 1.$$

The alternative hypothesis is that $\mathbf{x}_t$ is not conditionally linear

$$\mathbb{P}(\mathbb{E}(\mathbf{x}_t|\mathbf{x}_{t-1}) = \boldsymbol{\Upsilon}_0 + \boldsymbol{\Upsilon}_1 \mathbf{x}_{t-1}) < 1.$$

Obviously, the system is linear since $\boldsymbol{\Upsilon}_0 = \mathbf{0}$ and $\boldsymbol{\Upsilon}_1 = \boldsymbol{\xi}_1$ in our case. Now suppose that non-linearity is not tested for the whole vector $\mathbf{x}_t$ but all variables are treated individually using an univariate version of the definition above. For example, let us consider the variable X_{1t} and suppose that an appropriate filter is, only for simplicity of an explanation, an AR(1) process. Although it holds that $\mathbb{E}(X_{1t}|X_{1t-1}) \neq$

$\mathbb{E}(X_{1t}|\mathbf{x}_{t-1})$, it also holds that $\mathbb{E}(X_{1t}|X_{1t-1}) = \xi_{11}X_{1t-1} + \xi_{12}\mathbb{E}(X_{2t-1}|X_{1t-1}) = (\xi_{11} + \xi_{12}\rho)X_{1t-1}$. Therefore, the incorrect conditioning does not have to lead to a serious size distortion under the null hypothesis of linearity.[3]

A power distortion: Now let us consider a specific non-linear VAR(1) model $\mathbf{x}_t = \mathbf{g}_1(\mathbf{x}_{t-1}) + \mathbf{a}_t$, where $\mathbf{x}_t = (X_{1t}, X_{2t})'$ and $\mathbf{a}_t = (a_{1t}, a_{2t})'$ is a vector of model innovations such that $a_{it} \sim IID(0, \sigma_i^2)$, for $i \in \{1, 2\}$, and innovations are independent each other, regardless of the time index. The functional form of the model is given by

$$X_{1t} = \xi_{12}X_{2t-1}^2 + a_{1t},$$
$$X_{2t} = \xi_{22}X_{2t-1} + a_{2t}.$$

Obviously, the system is non-linear since $\mathbb{E}(\mathbf{x}_t|\mathbf{x}_{t-1}) \neq \boldsymbol{\Upsilon}_0 + \boldsymbol{\Upsilon}_1\mathbf{x}_{t-1}$ in our case. Now suppose that non-linearity is not tested for the whole vector $\mathbf{x}_t$ but all variables are treated individually using an univariate version of the definition above. Only for simplicity of exposition, suppose that an appropriate filter is an AR(1) process for both variables. It is clear that $\mathbb{E}(X_{2t}|X_{2t-1}) = \xi_{22}X_{2t-1}$ and the linearity condition is satisfied. It holds that $\mathbb{E}(X_{1t}|X_{1t-1}) = \xi_{12}\sigma_2^2 + \xi_{22}^2X_{1t-1} - \xi_{22}^2a_{1t-1}$ and the linearity condition is not formally satisfied in this case. However, the condition fails only due to some IID error term. As a results, we might expect a significant power loss of standard univariate non-linearity tests. This example shows how easily the probability of type II error may arise due to incorrect filtration (conditioning) under the alternative hypothesis.

Before we proceed to a formal testing procedure, we state an assumption about a stochastic process under consideration. The assumption is of the crucial importance for setting the null hypothesis of linearity.

[3] However, it is worth noting that a size distortion can be expected to increase especially in large-dimensional time series models. The author is very grateful to Professor Timo Teräsvirta from the Aarhus University for making this point.

Assumption 5 *Let us assume the following stationary real-valued finite-order linear VAR model under the null hypothesis*

$$\mathbf{x}_t = \boldsymbol{\xi}_0 + \sum_{i=1}^{P} \boldsymbol{\xi}_i \mathbf{x}_{t-i} + \mathbf{a}_t, \tag{5.1}$$

where $\mathbf{x}_t$ *denotes a* $(k \times 1)$ *vector,* $\{\mathbf{a}_t : t \in \mathbb{Z}\}$ *is a sequence of multivariate WN(*$\mathbf{0},\boldsymbol{\Sigma}$*) innovations with zero means and the variance-covariance matrix* $\boldsymbol{\Sigma}$*, which is symmetric and positive definite, such that* $\mathbb{E}(\|\mathbf{a}_t\|^8) < \infty$*. Let* $\boldsymbol{\beta} = (\boldsymbol{\xi}_0', \mathrm{vec}(\boldsymbol{\xi}_1)', \ldots, \mathrm{vec}(\boldsymbol{\xi}_p)')'$ *be a* $(k^2P + k \times 1)$ *parameter vector, which is assumed to lie in the interior of the parameter space given by*

$$\mathbf{B} = \{\boldsymbol{\beta} \in \mathbb{R}^{k^2P+k} : \det(\mathbf{I} - \sum_{i=1}^{P} \boldsymbol{\xi}_i z^i) \neq 0 \quad \textit{for all} \quad |z| \leq 1\}.$$

□

The assumption ensures that a given linear process is stationary, parameters do not lie on the boundary, and all moment conditions are satisfied. These conditions are sufficient to ensure consistency of the estimated parameters in $\boldsymbol{\beta}$, the estimated residuals, and subsequently, the non-linearity test statistics. Note that the null hypothesis of linearity can be extended to VARMA models as well.[4]

Although there are many different non-linearity tests in the literature, special attention is paid to a multivariate version of the ARCH test proposed by Engle (1982) and the TSAY test proposed by Tsay (1986). There are three good reasons for considering this couple of tests: (i) It is shown in Chapter 2 that both test statistics capture rather different types of non-linear features. The TSAY test is a simple test for non-linearity in the conditional mean, whereas the ARCH test for non-linearity in the conditional variance; (ii) Both tests suffer from a dimensionality problem in a similar way; (iii) Both tests are very well known, they are easy to construct and follow standard limiting distributions.

[4] Recall that identifying and estimating VARMA models is computationally expensive. For this reason, only a VAR model is considered under the null hypothesis.

5.2.2 Multivariate TSAY Test

Harvill and Ray (1999) proposed a multivariate version of the TSAY test. The test is based on running the following auxiliary equation

$$\hat{\mathbf{a}}_t = \mathbf{b}_0 + \mathbf{B}_1\mathbf{z}_t + \mathbf{B}_2\mathbf{v}_t + \mathbf{u}_t, \tag{5.2}$$

where $\hat{\mathbf{a}}_t$ is a $(k \times 1)$ vector of residuals from a particular model under the null hypothesis (e.g. a VAR model), $\mathbf{z}_t = (\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-P})'$ denotes a $(kP \times 1)$ vector of aggregated predetermined variables, $\mathbf{v}_t = \mathrm{vech}(\mathbf{z}_t \otimes \mathbf{z}'_t)$ represents an $(s \times 1)$ final vector of predetermined variables consisting of all square and cross-product elements $(s = kP(kP+1)/2)$. $\mathbf{b}_0$ denotes a $(k\times 1)$ vector of constants, $\mathbf{B}_1$ represents a $(k\times kP)$ matrix of parameters, and finally, $\mathbf{B}_2$ is a $(k \times s)$ matrix of parameters. The null hypothesis of linearity of the vector $\mathbf{x}_t$ is given by: $H_0 : \mathbf{B}_2 = \mathbf{0}$ versus $H_1 : \mathbf{B}_2 \neq \mathbf{0}$. The null hypothesis can be tested both by the LM- and LR-based test statistics.[5] The appropriate LR-type test statistic is given by[6]

$$MTSAY(org) = (T - \tau)(\log(|\hat{\mathbf{\Sigma}}_r|) - \log(|\hat{\mathbf{\Sigma}}_u|)) \xrightarrow{d} \chi^2(sk), \tag{5.3}$$

where $|\cdot|$ denotes the determinant of a square matrix, $\hat{\mathbf{\Sigma}}_r$ and $\hat{\mathbf{\Sigma}}_u$ represent the estimated variance-covariance matrix of the restricted model, unrestricted model respectively, T is the sample size, and $\tau = (k+s+1)/2$ is a small sample correction term recommended by Anderson (2003, p. 321-3), where k is the number of variables, and s represents the number of additional variables. He argues that the small sample correction works well, provided that $k^2 + s^2 < T/3$. Note that the model in (5.2) can be easily estimated by a multivariate LS method, see Lütkepohl (2005, Ch. 3) for details. The proof of the limiting distribution can be found in Anderson (2003, Ch. 8.5).

[5] Godfrey (1988, Chapter 2) shows that for a linear regression model such as (5.2), the LR-based test is slightly more powerful as compared to the LM-based test. In addition, both the LR- and LM-based tests are computed in the same way (using the auxiliary equation) in this particular case and have the same limiting distribution, see Davidson and MacKinnon (1999, p. 423–428). The LM-based test for multivariate systems is discussed in Deschamps (1993). Monte Carlo comparison of both LM- and LR-based tests for multivariate systems can be found in Deschamps (1996).

[6] Note that the authors use originally the F-test based on the Wilks lambda statistic.

It is worth mentioning, however, that testing for non-linearity using the above defined MTSAY(org) test statistic displays at least two shortcomings: (i) Due to a large number of terms in the vector $\mathbf{v}_t$, the original test requires a large number of observations; (ii) In addition, terms in the vector $\mathbf{v}_t$ are highly collinear, which significantly increases the degrees of freedom of the test, but actually does not improve the fit in (5.2). As a result, multicollinearity in the vector $\mathbf{v}_t$ can reduce the power of the multivariate tests. It is clear from the dimension of the $\mathbf{v}_t$ vector that, for a given set of k variables, the test requires $T > s$ observations. For example, consider a small model consisting of $k = 5$ different economic variables and the moderate lag order $P = 4$ of a VAR model under the null. Then, the MTSAY(org) test requires $T > 210$ observations but the degrees of freedom surge to 1050. Table 5.1 depicts some other examples of the degrees of freedom and the number of observations required by the original TSAY(org) test.

Table 5.1 Requirements of the original multivariate TSAY test

	number of observations					degrees of freedom				
variables k/lags P	1	2	3	4	5	1	2	3	4	5
2	3	10	21	36	55	6	20	42	72	110
3	6	21	45	78	120	18	63	135	234	360
4	10	36	78	136	210	40	144	312	544	840
5	15	55	120	210	325	75	275	600	1050	1625

A dimensionality problem of the MTSAY test can be efficiently diminished using a principal component analysis (PCA). Details about PCA are discussed in the next section. The modified multivariate TSAY test is based on running the following auxiliary equation

$$\hat{\mathbf{a}}_t = \mathbf{c}_0 + \mathbf{C}_1\mathbf{z}_t + \mathbf{C}_2\mathbf{w}_t + \mathbf{u}_t, \tag{5.4}$$

where $\hat{\mathbf{a}}_t$ is a $(k \times 1)$ vector of residuals from a particular filter, $\mathbf{w}_t$ is an $(n \times 1)$ vector of principal components, such that $k \leq n \leq s$.[7], $\mathbf{c}_0$ is a $(k \times 1)$ vector of constants,

[7] Note that the upper bound of the number of components might be restricted to $s = [T/2]$, where $[\cdot]$ denotes an integer part.

$\mathbf{C}_1$ and $\mathbf{C}_2$ are $(k \times kP)$ and $(k \times n)$ coefficient matrices. The null hypothesis of linearity of the vector $\mathbf{x}_t$, is given by: $H_0 : \mathbf{C}_2 = \mathbf{0}$ versus $H_1 : \mathbf{C}_2 \neq \mathbf{0}$. The appropriate LR-based test statistic is given by

$$TSAY = (T - \tau)(\log(|\hat{\mathbf{\Sigma}}_r|) - \log(|\hat{\mathbf{\Sigma}}_u|)) \xrightarrow{d} \chi^2(nk). \tag{5.5}$$

Although principal component analysis can reduce a dimensionality problem, in practice, there is still an ultimate question of how many components to retain. Details about a principal component analysis and various stopping rules for determining the number of principal components are discussed in Section 5.2.4.

5.2.3 Multivariate ARCH Test

Lütkepohl (2005, Ch. 16) considers a multivariate version of the ARCH test. The test is based on running the following auxiliary equation

$$\text{vech}(\hat{\mathbf{a}}_t \otimes \hat{\mathbf{a}}_t') = \mathbf{b}_0 + \mathbf{B}\mathbf{v}_t + \mathbf{u}_t, \tag{5.6}$$

where vech$(\cdot)$ is a half-stacking operator, $\hat{\mathbf{a}}_t$ is a $(k \times 1)$ vector of residuals from a particular VAR model under the null hypothesis, $\mathbf{v}_t = (\text{vech}(\hat{\mathbf{a}}_{t-1} \otimes \hat{\mathbf{a}}_{t-1}')', \ldots, \text{vech}(\hat{\mathbf{a}}_{t-Q} \otimes \hat{\mathbf{a}}_{t-Q}')')'$ denotes an $(s \times 1)$ vector of the predetermined variables of the test, $\mathbf{b}_0$ is an $(m \times 1)$ vector of constants, and $\mathbf{B}$ is an $(m \times s)$ matrix of parameters ($m = k(k+1)/2$ and $s = mQ$) matrix of parameters. The null hypothesis of homoscedasticity of the vector $\mathbf{a}_t$ is given by: $H_0 : \mathbf{B} = \mathbf{0}$ versus $H_1 : \mathbf{B} \neq \mathbf{0}$. The null hypothesis can be tested both by the LM- and LR-based test statistics in the multivariate setup. In order to be consistent with the multivariate TSAY test discussed earlier, the LR-based test is implemented. The appropriate LR-type test statistic is given by[8]

$$MARCH(org) = (T - \tau)(\log(|\hat{\mathbf{\Sigma}}_r|) - \log(|\hat{\mathbf{\Sigma}}_u|)) \xrightarrow{d} \chi^2(sm), \tag{5.7}$$

where $|\cdot|$ denotes the determinant of a square matrix, $\hat{\mathbf{\Sigma}}_r$ and $\hat{\mathbf{\Sigma}}_u$ represent the estimated variance-covariance matrix of the restricted model, unrestricted model

[8]Note that the LR-based multivariate ARCH test is considered in other studies as well, see Hacker and Hatemi-J (2005), among others. The LM version of the test can be found in Lütkepohl (2005, Ch. 16).

respectively, T is the sample size, and $\tau = (m + s + 1)/2$ is a small sample correction recommended by Anderson (2003, p. 321-3), where T stands for the sample size, k is the number of variables, and n represents the number of principal components. He argues that the small sample correction works very well, provided that $m^2 + s^2 < T/3$. Note that the model in (5.6) can be easily estimated by a multivariate LS method, see Lütkepohl (2005, Ch. 3) for details. The proof of the limiting distribution can be found in Anderson (2003, Ch. 8.5).

It is worth mentioning, however, that testing the conditional variance using the above defined MARCH(org) test statistic displays similar shortcomings like the MTSAY test. The main problem does not lie in the number of observations directly required for running the test but in the rapidly increasing degrees of freedom of the test. It is clear from the dimension of the $\mathbf{v}_t$ vector that, for a given set of k variables, the test requires $T > s$ observations but sm degrees of freedom. For example, consider a small model consisting of $k = 5$ different variables and the moderate lag order $Q = 4$ of the MARCH test. As a result, the test requires $T > 60$ observations but the degrees of freedom surge to 900. Table 5.2 depicts some other examples of the degrees of freedom and number of observations required by the original MARCH test.

Table 5.2 Requirements of the original multivariate ARCH test

	number of observations					degrees of freedom				
variables k/lags Q	1	2	3	4	5	1	2	3	4	5
2	3	6	9	12	15	9	18	27	36	45
3	6	12	18	24	30	36	72	108	144	180
4	10	20	30	40	50	100	200	300	400	500
5	15	30	45	60	75	225	450	675	900	1125

Another problem is related to the specification of the alternative hypothesis H_1. Lee and King (1993) show that the alternative hypothesis should be configured as one-sided hypothesis (i.e. $H_1 : \mathbf{B} > \mathbf{0}$) in order to reflect the fact that the conditional variance is supposed to be a non-negative quantity. For this reason, the authors proposed a locally most mean powerful test for (G)ARCH models. Nevertheless, their results indicate that, although the size and power properties of the locally most mean powerful test are better as compared to standard LM-based ARCH test, the standard LM-based ARCH test can still be considered as a conservative test with good size and power properties. Since a multivariate extension of their results is not known, at least to the best of our knowledge, the standard two-sided form of the alternative hypothesis is used in this paper. We leave this interesting issue to further research.

Two modifications of the MARCH test are introduced. First, note that the original test is based on running the auxiliary equation for $\text{vech}(\hat{\mathbf{a}}_t \otimes \hat{\mathbf{a}}_t')$, which contains many cross products. For example, in a bivariate case (i.e. $k = 2$), $\text{vech}(\hat{\mathbf{a}}_t \otimes \hat{\mathbf{a}}_t') = (\hat{a}_{1t}^2, \hat{a}_{1t}\hat{a}_{2t}, \hat{a}_{1t}^2)'$. The cross-elements might be important for the modelling purposes but not necessarily for testing heteroscedasticity itself. The main argument is that provided $\text{diag}(\hat{\mathbf{a}}_t \otimes \hat{\mathbf{a}}_t')$ is homoscedastic, then the cross elements are very likely to be homoscedastic as well. This useful property holds for many multivariate (G)ARCH models (e.g. VEC, BEKK, CCC-GARCH models), but not all (e.g. a DCC-GARCH model). This diagonal modification immediately reduces the required degrees of freedom from Qm^2 to Qkm. Using the same example as above, the required degrees of freedom would decrease from 900 to 300. Second, a principal component analysis is implemented in order to further diminish the dimensionality problem. The modified multivariate ARCH test is based on running the following auxiliary equation

$$\text{diag}(\hat{\mathbf{a}}_t \otimes \hat{\mathbf{a}}_t') = \mathbf{c}_0 + \mathbf{C}_1 \mathbf{w}_t + \mathbf{u}_t, \tag{5.8}$$

where $\text{diag}(\hat{\mathbf{a}}_t \otimes \hat{\mathbf{a}}_t')$ is a $(k \times 1)$ vector of diagonal elements, $\mathbf{c}_0$ is a $(k \times 1)$ vector of constants, $\mathbf{C}_1$ is an appropriate $(k \times n)$ matrix of coefficients, and $\mathbf{w}_t$ is an $(n \times 1)$ vector of principal components, such that $k \leq n \leq s$.[9] The null hypothesis of homoscedasticity of the vector $\mathbf{a}_t$, is given by: $H_0 : \mathbf{C}_1 = \mathbf{0}$ versus $H_1 : \mathbf{C}_1 \neq \mathbf{0}$. The

[9]Note that the upper bound of the number of components might be restricted to $s = [T/2]$, where $[\cdot]$ denotes an integer part.

appropriate LR-based test statistic is given by

$$MARCH = (T - \tau)(\log(|\hat{\mathbf{\Sigma}}_r|) - \log(|\hat{\mathbf{\Sigma}}_u|)) \xrightarrow{d} \chi^2(nk). \tag{5.9}$$

Although principal component analysis can reduce the dimensionality problem, in practice, there is still an ultimate question of how many components to retain. Details about a principal component analysis and various stopping rules for determining the number of principal components are discussed in Section 5.2.4.

5.2.4 Principal Component Analysis

A principal component analysis (PCA) is concerned with explaining the variance-covariance or correlation structure of a set of variables by a few linear combinations of original variables. Formally, the principal components are defined as follows

$$w_{jt} = \mathbf{e}_j'\mathbf{v}_t, \quad \text{for} \quad j = 1, \ldots, s, \quad t = 1, \ldots, T, \tag{5.10}$$

where w_{jt} is the jth-principal component, $\mathbf{e}_j$ is a particular eigenvector associated with the eigenvalue λ_j estimated from the variance-covariance or correlation matrix. It is important to point out that there is no one-to-one mapping between the roots calculated from the variance-covariance matrix and correlation matrix. A problem is that, unlike the correlation matrix, the variance-covariance matrix is not scale invariant and, hence, neither the calculated roots. Therefore, comfortable or not, the use of the correlation matrix is often recommended, especially for heterogenous data sets and/or indicators originally measured in different units, see Jackson (1991, 64–65) for details. For this reason, the correlation matrix is used in this chapter unless otherwise stated. For instance, a vector of additional variables $\mathbf{v}_t$ takes the following form for the MTSAY test based on a VAR(P) filter $\mathbf{v}_t = \text{vech}(\mathbf{z}_t \otimes \mathbf{z}_t')$, where $\mathbf{z}_t = (\mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}')'$ is a vector of predetermined variables. Note that the vector $\mathbf{v}_t$ for the MARCH test is defined in a similar way, see the previous section.

A characteristic feature of PCA is that components are uncorrelated linear combinations of original variables due to orthogonality of the estimated eigenvectors (i.e. $\mathbf{e}_i'\mathbf{e}_j = 0$ for all $i \neq j$). The advantage of this approach is that principal components basically eliminate multicollinearity from a testing procedure.[10] Another advantage of this approach is that, at least for testing purposes, no interpretation of the calculated principal components is required, which significantly simplifies the use of PCA. Another interesting property is that $\mathrm{var}(w_j) = \mathbf{e}_j'\mathrm{var}(\mathbf{v}_t)\mathbf{e}_j = \lambda_j$, for $j = 1, \ldots, s$, which immediately implies that

$$\sum_{j=1}^{s} \mathrm{var}(w_j) = \sum_{j=1}^{s} \mathrm{var}(v_j) = \sum_{j=1}^{s} \lambda_j. \tag{5.11}$$

It is clear from (5.11) that s components are required to reproduce the total variance of the original data set. In practice, however, most of the variance can be accounted for by just a small number of the first components, say $n < s$.

Although PCA can reduce a dimensionality problem, there is still an ultimate question of how many components to retain in practice. Unfortunately, there is no definitive answer to this question. Stopping rules can be theoretically split into four basic categories: (i) purely statistical rules (e.g. a Bartlett test); (ii) graphical rules (e.g. a scree plot); (iii) rule-of-thumb stopping rules; and finally (iv) simulation/bootstrap-based rules. The interested reader is referred to Peres-Neto et al. (2005) for details.

We omit all statistical rules since they might be problematic in the context of a time series analysis. For example, let us consider the Bartlett test of the equivalence of the last $s - n$ roots calculated from the variance-covariance matrix. There are two shortcomings of this test. The main disadvantage is that the limiting distribution of the Bartlett test is based on the assumption of multivariate normality of original observations in the vector $\mathbf{v}_t$, see Jolliffe (2005, p. 53–54), which is rather difficult to justify in the context of a time series analysis and/or non-linearity testing. Another disadvantage is that the limiting distribution of the test is no longer χ^2, provided

[10] Note that it might technically happen that some of the calculated eigenvalues are equal, which means that the choice of eigenvectors, and, subsequently, principal components, is not unique. A standard recommendation, also implemented here, is to use any eigenvectors orthogonal each other. This solution ensures that calculated principal components are still uncorrelated even if not unique, see Johnson and Wichern (2007, Ch. 8) for details.

that the roots are calculated from the correlation matrix, see Jackson (1991, p. 99–101). As a result, the Bartlett test overestimates the actual number of components in practice. Of course, there are some other formal statistical rules which can be used, many of them, unfortunately, suffer from a similar problem like the Bartlett test, see Jolliffe (2005, p. 118–126) for a discussion.

We also skip useful, yet relatively subjective, graphical methods such as a scree plot. The scree plot is a figure depicting sample eigenvalues plotted in descending order against the order number. Provided that just a few first components dominate in magnitude and the rest of eigenvalues is relatively small (and almost equal), then the scree plot does exhibit a break (the so called "elbow" or "broken stick") corresponding to the division of sample eigenvalues into two groups. The order number of eigenvalues around which the break occurs is usually recommended to use as the number of the first principal components to retain. There are two shortcomings of this approach. First, as shown by Izenman (2008, p. 206), the usefulness of the scree plot depends critically on the relationship between the sample size T and the number of variables k. The scree plots seem to be informative only if the sample size T is significantly larger than k, which is rather difficult to guarantee in practice. Second, the scree plot, like any other graphical method, is not a convenient technique for Monte Carlo analysis.

Among those rules successfully applied in the literature, the following three are implemented here:

1. The information criterion rule: The number of principal components can be determined using an automatic selection procedure based on minimizing an appropriate information criterion. Blake and Kapetanios (2003) show, using Monte Carlo experiments, that the BIC approach produces superior results as compared to other methods.

2. The variance rule: Another popular way of selecting the number of principal components is to use the first n components attributing $100\gamma\,\%$ of total vari-

ance of the original set of variables. The usually recommended proportion of total variance recommended in multivariate analysis is $\gamma = 0.9$.

3. The kaiser (root) rule: This rule is based on the fact that the average root calculated from the correlation matrix is equal one. For this reason, the rule suggest to retain all the first eigenvalues larger than 1.

Note that the testing procedure cannot be carried out if no principal component is chosen by the automatic selection procedure. We therefore do not consider this case and start with a minimum of k principal components for the multivariate tests.

Nevertheless, the use of PCA in time series is not without problems. First, PCA is actually built on the independence assumption of original variables (i.e. variables in the vector $\mathbf{v}_t$) over time. However, it can be shown that if PCA is used entirely just for descriptive purposes, not inferential, then weak dependence and/or other non-IID features in the original vector $\mathbf{v}_t$ do not seriously affect the main objective, see Jolliffe (2005, Ch.12) and Jackson (1991, Ch. 4) for details. Second, PCA critically depends on the properties of the variance-covariance or correlation structure of the original observations. This assumption might be problematic as well. It is nowadays well known that non-linear features might be amplified by the presence of outliers and/or structural breaks, see Koop and Potter (2001) or Carrasco (2002) for a discussion. In such a case, the Pearson correlation matrix might not be the best choice and some robust way of calculating the correlation matrix seems to be more appropriate. Although there exist many robust alternatives, see Croux and Haesbroeck (2000), Ma and Genton (2001), or Hubert et al. (2005), among others, they are not convenient for non-linearity testing purposes. The main problem is that robust techniques trim or downweight the aberrant observations, which might be, however, associated with non-linear nature of a given stochastic process. For this reason, the Spearman correlation matrix based on ranks appears a good choice here.

Finally, the following six alternative ways to determine the number of principal components n are used for both multivariate TSAY and ARCH tests in this chapter:

(i) “p-bic” stands for the test with automatically selected number of principal components using the BIC approach and the Pearson correlation matrix; (ii) “p-0.9” denotes the test, where the number of principal components is determined in such a way to cover at least 90 % of the variability of original variables; (iii) “p-k” is the test, where the number of principal components is set using the Kaiser rule with the cutoff 1.0 for eigenvalues calculated from the Pearson correlation matrix; (iv) “s-bic” stands for the test with automatically selected number of principal components using the BIC approach and the Spearman correlation matrix; (v) “s-0.9” denotes the test, where the number of principal components is determined in such a way to cover at least 90 % of the variability of original variables; (vi) “s-k” is the test, where the number of principal components is set using the Kaiser rule with the cutoff 1.0 for eigenvalues calculated from the Spearman correlation matrix.

5.3 Monte Carlo Setup

The statistical properties of both univariate and multivariate (original and principal component-based) non-linearity tests are examined using three sets of multivariate stationary time series models: (i) M-models: three linear VAR models; (ii) N-models: six non-linear conditional mean models; and finally (iii) H-models: six non-linear conditional heteroscedastic models. All data generating processes (DGPs) considered in this chapter are summarized in Tables 5.3 – 5.5. Since a large number of multivariate models is considered in this chapter, some comments on the selected DGPs are in order. M-models are used to assess the size properties of the univariate and multivariate TSAY and ARCH tests: model M1 represents a diagonal VAR(2) model with relatively high persistence, whereas models M2 and M3 are simple VAR(1) models with different persistence. H-models represent various conditional heteroscedastic and stochastic volatility models: H1 and H2 denote different versions of a diagonal VEC-GARCH model with different source of volatility; H3, H4, and H5 represent different versions of a BEKK-GARCH model also with different source of volatility; finally, H6 is a simple multivariate stochastic volatility model. N-models represent a set of non-linear conditional mean models: N1 – N4 denote different specifications of the first-order Taylor approximation of a non-linear VARMA model with a rather distinct source of non-linearity; N5 is a vector TAR model, and N6 is a vector MSAR model. Special attention is paid to the following pairs of models: N1 and N2, N3 and N4, H1 and H2, H3 and H4, which all have the same functional form, but non-linearity is governed in a different way. The above mentioned pairs of models might be considered as a robustness check of the non-linearity tests. Note that all the selected DGPs are time series models borrowed from the literature, see Ling and Li (1997), Duchesne and Lalancette (2003), and Harvill and Ray (1999), among others. Although the list of multivariate time series models is definitely not exhaustive, we are strongly convinced that all the main classes of non-linear models are included. Note also that parameters of all DGPs are set in such a way that the generated series are stationary and imply positive-definite variance-covariance matrices (for H models). The moment condition of the non-linearity tests is by no means easy to check for a given set of multivariate non-linear models. For this reason, the max-sum ratio procedure to check the existence of moments is implemented instead, see Embrechts et al. (2011, p. 309).

The performance of the selected non-linearity tests is assessed using three different sample sizes $T \in \{200, 500, 1000\}$. Originally, $T+100$ observations is simulated in each experiment, but the first 100 of them are discarded in order to eliminate the effect of initial observations. The number of replications of all experiments is set to $R = 1000$. Model innovations $\mathbf{a}_t$ are drawn from a multivariate Gaussian distribution $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with zero means and the variance-covariance matrix $\boldsymbol{\Sigma}$. Two different configurations of $\boldsymbol{\Sigma}$ are considered in this chapter

$$(i) \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix}, \quad (ii) \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}.$$

Non-diagonal $\boldsymbol{\Sigma}_2$ allows for (positive) correlation between model innovations, whereas diagonal $\boldsymbol{\Sigma}_1$ indicates the independence of innovations. Using this specification, we can check the robustness of the non-linearity tests against the correlation structure of model innovations.

In each experiment, the generated series are filtered by either an AR model or a VAR model, depending on a particular test statistic. The lag order p of AR models is selected by the Bayesian information criterion (BIC) developed by Schwarz (1978). Following the arguments in Ng and Perron (2005), a modified version of the criterion is used. They show, based on extensive Monte Carlo experiments, that the best method to give the correct lag order is that with the fixed efficient sample size. Therefore, our criterion is defined as follows

$$BIC_l^u = \log(\hat{\sigma}_l^2) + \frac{l \log(N)}{N},$$

where $\hat{\sigma}^2$ is the estimated variance of residuals, $l \in \{1, \ldots, L\}$, and $N = T - L$ is the efficient sample size, where T is the actual sample size and L is the maximum lag order constrained by $L = [8(T/100)^{0.25}]$. Finally, the lag order p is estimated as $\hat{p} = \min_{l \in \{1,\ldots,L\}}(BIC_l^u)$. The lag order of a VAR model is determined by a multivariate version of the Bayesian information criterion (BIC). The criterion is given by

$$BIC_l^m = \log |\hat{\mathbf{\Sigma}}_l| + \frac{lk^2 \log(N)}{N},$$

where $|\cdot|$ denotes a determinant, $\hat{\mathbf{\Sigma}}$ is the estimated variance-covariance matrix of residuals, T is the actual sample size and L is the maximum lag order constrained by $L = [8(T/100)^{0.25}]$. Finally, the lag order p for VAR(P) models is estimated as $\hat{P} = \min_{l \in \{1,\ldots,L\}}(BIC_l^m)$. Note that the same approach is also used for determining the lag order for the univariate and multivariate ARCH tests.

Table 5.3 List of multivariate time series models: part 1

M1:

$$\mathbf{x}_t = \begin{pmatrix} 0.4 & 0.0 \\ 0.0 & 0.4 \end{pmatrix} \mathbf{x}_{t-1} + \begin{pmatrix} 0.3 & 0.0 \\ 0.0 & 0.3 \end{pmatrix} \mathbf{x}_{t-2} + \mathbf{a}_t.$$

M2:

$$\mathbf{x}_t = \begin{pmatrix} 0.4 & 0.3 \\ 0.3 & 0.4 \end{pmatrix} \mathbf{x}_{t-1} + \mathbf{a}_t.$$

M3:

$$\mathbf{x}_t = \begin{pmatrix} 0.4 & -0.3 \\ -0.3 & 0.4 \end{pmatrix} \mathbf{x}_{t-1} + \mathbf{a}_t.$$

Table 5.4 List of multivariate time series models: part 2

N1:

$$\mathbf{x}_t = \begin{pmatrix} 0.4 & -0.3 \\ -0.3 & 0.4 \end{pmatrix} \mathbf{x}_{t-1} + \begin{pmatrix} 0.1 & 0.0 \\ 0.0 & 0.1 \end{pmatrix} \begin{pmatrix} X_{1t-1}\epsilon_{1t-1} \\ X_{2t-1}\epsilon_{2t-1} \end{pmatrix} + \mathbf{a}_t.$$

N2:

$$\mathbf{x}_t = \begin{pmatrix} 0.4 & -0.3 \\ -0.3 & 0.4 \end{pmatrix} \mathbf{x}_{t-1} + \begin{pmatrix} 0.0 & 0.1 \\ 0.1 & 0.0 \end{pmatrix} \begin{pmatrix} X_{1t-1}\epsilon_{1t-1} \\ X_{2t-1}\epsilon_{2t-1} \end{pmatrix} + \mathbf{a}_t.$$

N3:

$$\mathbf{x}_t = \begin{pmatrix} 0.4 & -0.3 \\ -0.3 & 0.4 \end{pmatrix} \mathbf{x}_{t-1} + \begin{pmatrix} 0.0 & 0.1 \\ 0.1 & 0.0 \end{pmatrix} \begin{pmatrix} X_{1t-1}^2 \\ X_{2t-1}^2 \end{pmatrix} + \mathbf{a}_t.$$

N4:

$$\mathbf{x}_t = \begin{pmatrix} 0.4 & -0.3 \\ -0.3 & 0.4 \end{pmatrix} \mathbf{x}_{t-1} + \begin{pmatrix} 0.1 & 0.0 \\ 0.0 & 0.1 \end{pmatrix} \begin{pmatrix} X_{1t-1}X_{2t-1} \\ X_{2t-1}X_{1t-1} \end{pmatrix} + \mathbf{a}_t.$$

N5:

$$\mathbf{x}_t = \begin{pmatrix} 0.7 & 0.0 \\ 0.3 & 0.7 \end{pmatrix} \mathbf{x}_{t-1} I(X_{1t-1} \le 0) + \begin{pmatrix} -0.7 & 0.0 \\ -0.3 & -0.7 \end{pmatrix} \mathbf{x}_{t-1} I(X_{1t-1} > 0) + \mathbf{a}_t.$$

N6:

$$\mathbf{x}_t = \begin{pmatrix} 0.7 & 0.0 \\ 0.3 & 0.7 \end{pmatrix} \mathbf{x}_{t-1} I(S_t = 1) + \begin{pmatrix} -0.7 & 0.0 \\ -0.3 & -0.7 \end{pmatrix} \mathbf{x}_{t-1} I(S_t = 2) + \mathbf{a}_t,$$

where $p_{11} = \mathbb{P}(S_t = 1 | S_{t-1} = 1) = 0.9$ and $p_{22} = \mathbb{P}(S_t = 2 | S_{t-1} = 2) = 0.7$.

Table 5.5 List of multivariate time series models: part 3

$$\mathbf{x}_t = \begin{pmatrix} 0.4 & -0.3 \\ -0.3 & 0.4 \end{pmatrix} \mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t,$$
$$\boldsymbol{\epsilon}_t = \sqrt{\mathbf{H}_t}\mathbf{a}_t,$$

where $\mathbf{a}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma})$.

H1:

$$\mathrm{diag}(\mathbf{H}_t) = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 0.2 & 0.0 \\ 0.0 & 0.2 \end{pmatrix} \mathrm{diag}(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}) + \begin{pmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{pmatrix} \mathrm{diag}(\mathbf{H}_{t-1}).$$

H2:

$$\mathrm{diag}(\mathbf{H}_t) = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 0.0 & 0.2 \\ 0.2 & 0.0 \end{pmatrix} \mathrm{diag}(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}) + \begin{pmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{pmatrix} \mathrm{diag}(\mathbf{H}_{t-1}).$$

H3:

$$\mathbf{H}_t = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{pmatrix} + \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix} \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1} \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.0 \\ 0.0 & 0.4 \end{pmatrix} \mathbf{H}_{t-1} \begin{pmatrix} 0.4 & 0.0 \\ 0.0 & 0.4 \end{pmatrix}.$$

H4:

$$\mathbf{H}_t = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{pmatrix} + \begin{pmatrix} 0.1 & 0.2 \\ 0.2 & 0.1 \end{pmatrix} \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1} \begin{pmatrix} 0.1 & 0.2 \\ 0.2 & 0.1 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.0 \\ 0.0 & 0.4 \end{pmatrix} \mathbf{H}_{t-1} \begin{pmatrix} 0.4 & 0.0 \\ 0.0 & 0.4 \end{pmatrix}.$$

H5:

$$\mathbf{H}_t = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.1 \end{pmatrix} + \begin{pmatrix} 0.2 & 0.2 \\ 0.2 & 0.2 \end{pmatrix} \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1} \begin{pmatrix} 0.2 & 0.2 \\ 0.2 & 0.2 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.0 \\ 0.0 & 0.4 \end{pmatrix} \mathbf{H}_{t-1} \begin{pmatrix} 0.4 & 0.0 \\ 0.0 & 0.4 \end{pmatrix}.$$

$$\mathbf{x}_t = \begin{pmatrix} 0.4 & -0.3 \\ -0.3 & 0.4 \end{pmatrix} \mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t,$$
$$\boldsymbol{\epsilon}_t = \mathbf{u}_t \odot \exp\{0.5\mathrm{diag}(\mathbf{H}_t)\},$$

where $\mathbf{a}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma})$ and $\mathbf{u}_t \sim NID(\mathbf{0}, \mathbf{I})$.

H6:

$$\mathrm{diag}(\mathbf{H}_t) = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.2 \\ 0.2 & 0.4 \end{pmatrix} \mathrm{diag}(\mathbf{H}_{t-1}) + \mathbf{a}_t.$$

5.4 Monte Carlo Results

5.4.1 Size Properties (M-models)

For each data generating process, each configuration of the variance-covariance matrix $\mathbf{\Sigma}$, and each sample size T, the average rejection frequency is reported for both univariate and multivariate tests as follows

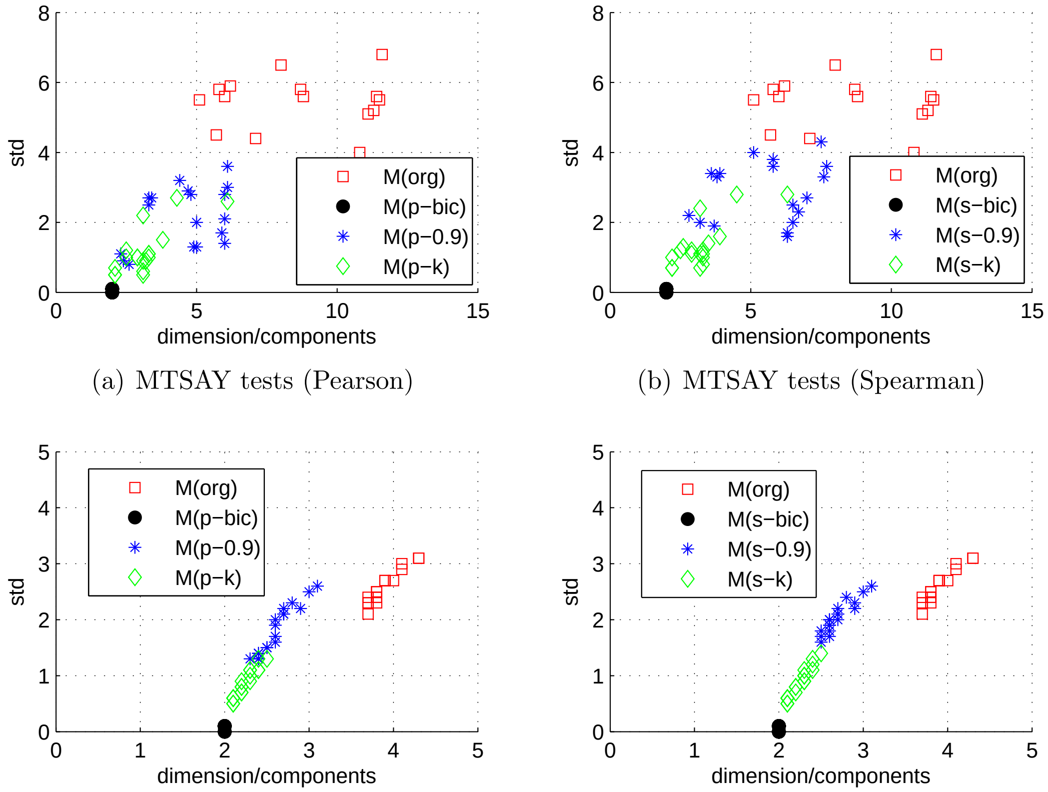
$$avg = \frac{1}{R}\sum_{r=1}^{R} I(\hat{\alpha}_r \leq \alpha),$$

where $R = 1000$ denotes the number of repetitions of each experiment, $I(\cdot)$ is a standard indication function, α is the statistical significance of the test set to 0.05, and $\hat{\alpha}$ is the estimated p-value of a particular test under consideration. The average dimensions of the vector $\mathbf{v}_t$ (i.e. the number of additional variables) and the vector $\mathbf{w}_t$ (i.e. the number of principal components), including the standard deviation of these quantities calculated over all Monte Carlo repetitions, are reported for all the multivariate non-linearity tests as well.

The size results for the selected non-linearity tests are presented in Tables 5.8 – 5.9. The results suggest that the size of the vast majority of the test statistics lies in the range between 0.03 and 0.07, regardless of the data generating process, the sample size T, or the variance-covariance matrix $\mathbf{\Sigma}$. So, it can be concluded that all the univariate and multivariate non-linearity tests are reasonably well sized.

Special attention is paid to evaluating the multivariate tests with respect to the number of principal components selected by different stopping rules. The average number of principal components and the standard deviation of the number of components are reported in Tables 5.10 – 5.13. For higher clarity, the results are depicted in Figure 5.1 as well. The figure depicts the relationship between the average number of principal components (x-axis) and the standard deviation of the number of principal components (y-axis) of the multivariate tests calculated for each DGP configuration. The results suggest the following: (i) Significant differences in the number of principal components and its variability are observed for individual

Figure 5.1 Statistical properties of the stopping rules of the multivariate tests: M-models



(a) MTSAY tests (Pearson)

(b) MTSAY tests (Spearman)

(c) MARCH tests (Pearson)

(d) MARCH tests (Spearman)

Note: M(org) denotes the original multivariate TSAY test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

stopping rules. The BIC stopping rule is apparently the most robust and efficient rule determining almost a constant number of principal components (i.e. $n \approx 2$), regardless of the data generating process, the sample size T, and the configuration of the variance-covariance matrix $\mathbf{\Sigma}$. The Kaiser stopping rule performs significantly better than the variance rule, but worse than the BIC rule; (ii) No noticeable differences are observed between the results of the multivariate tests based on the Pearson and Spearman correlation matrices; (iii) An interesting difference in the statistical properties of the stopping rules is observed between the MTSAY and MARCH tests. In particular, much higher heterogeneity in the average number of components is observed for the MTSAY tests as compared to the MARCH tests. The reason for that lies in the fact that parameters of the MTSAY tests are directly linked to parameters of a filter under the null hypothesis (i.e. the lag order P of a VAR model in our case), whereas parameters of the MARCH tests lack this link. This interesting property might be also used for further improvements of the multivariate non-linearity tests.[11]

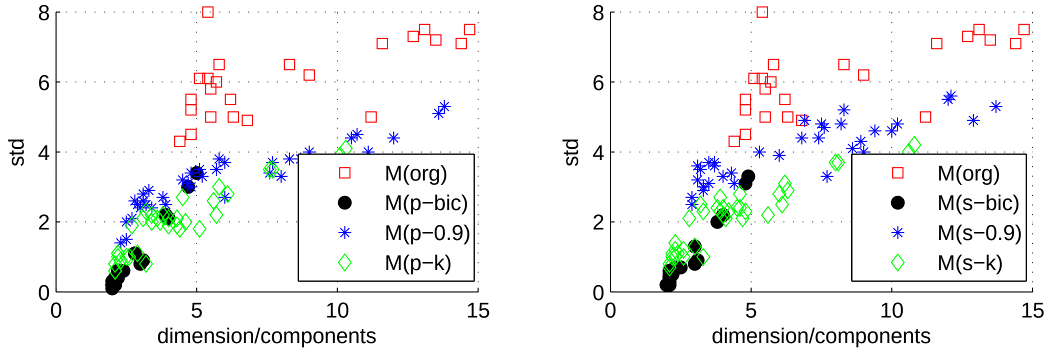
5.4.2 Power Properties (N-models)

For each data generating process, each configuration of the variance-covariance matrix $\mathbf{\Sigma}$, and each sample size T, the average rejection frequency is calculated as in the previous section. The average dimension of the vector $\mathbf{v}_t$ (i.e. the number of additional variables) and the vector $\mathbf{w}_t$ (i.e. the number of principal components), including the standard deviation of these quantities calculated over all Monte Carlo repetitions, are reported for all the multivariate non-linearity tests as well.

The power results of the selected non-linearity tests for N-models are presented in Tables 5.14 – 5.15. Since the ARCH tests have relatively low power against non-linear conditional mean models, the main focus is on the TSAY tests in this section. The ARCH tests results are reported for completeness only. The results show significant differences in the rejection frequency of the multivariate and univariate TSAY

[11] The author, as a part of his research agenda at the National Bank of Slovakia, currently works on developing new multivariate neural network (MNN) tests.

Figure 5.2 Statistical properties of the stopping rules of the multivariate tests: N-models



(a) MTSAY tests (Pearson) (b) MTSAY tests (Spearman)

Note: M(org) denotes the original multivariate TSAY test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

tests, unless the sample size is sufficiently large (i.e. $T = 1000$). In particular, the rejection frequency of the univariate TSAY test, applied to individual series, is significantly lower, especially for N2, N3 and N4 model configurations. Put differently, the univariate TSAY test completely fails, provided that non-linearity in the conditional mean is generated by cross-bilinear terms (N2 case), cross-quadratic terms (N3 case), or cross-product terms (N4 case).

Special attention is paid to evaluating the multivariate tests with respect to the number of components selected by different rules. The average number of principal components and the standard deviation of the number of components are reported in Tables 5.16 – 5.19. For higher clarity, the results are depicted in Figure 5.2 as well. The figure depicts the relationship between the average number of principal components (x-axis) and the standard deviation of the number of principal components (y-axis) of the multivariate tests calculated for each DGP configuration. The results suggest the following: (i) Significant differences in the number of principal components and its variability are observed for individual stopping rules. The BIC stopping rule is apparently the most robust and efficient rule determining almost a constant number of principal components, regardless of the DGP, the sample size T, and the variance-covariance matrix $\mathbf{\Sigma}$. The Kaiser stopping rule performs significantly better than the variance rule, but slightly worse than the BIC rule; (ii) No differences are observed between the results of the multivariate tests based on the Pearson and Spearman correlation matrices.

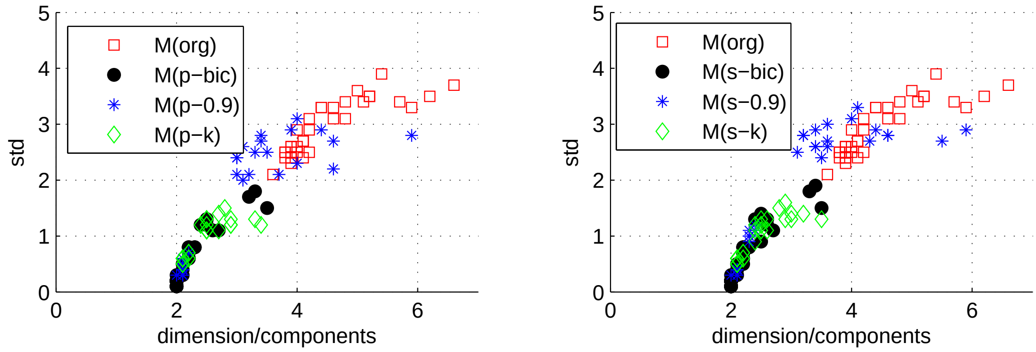
5.4.3 Power Properties (H-models)

For each data generating process, each configuration of the variance-covariance matrix $\mathbf{\Sigma}$, and each sample size T, the average rejection frequency is calculated as in the previous sections. The average dimension of the vector $\mathbf{v}_t$ (i.e. the number of additional variables) and the vector $\mathbf{w}_t$ (i.e. the number of principal components), including their standard deviations, are reported for all the multivariate non-linearity tests as well.

The power results of the selected non-linearity tests for the H-models are presented in Tables 5.20 – 5.21. Since the TSAY tests have very low power against conditional heteroscedastic models, the main focus is on the ARCH tests in this section. The TSAY tests results are reported for completeness only. The results suggest the following: (i) Significant differences are observed in the rejection frequency of the multivariate and univariate ARCH tests, whereas no differences exist among the multivariate ARCH tests, regardless the DGPs, the sample size T, and the configuration of the variance-covariance matrix $\mathbf{\Sigma}$. In particular, the rejection frequency of the univariate ARCH test, applied to individual series, is significantly lower in some cases, especially for H2 and H5 configurations. Put differently, the univariate ARCH test suffers from a serious power distortion, provided that the conditional volatility is generated, for instance, by cross-heteroscedasticity terms rather than individual autoregressive terms.

Special attention is paid to evaluating the multivariate tests with respect to the number of components selected by different rules. The average number of principal components and the standard deviation of the average number of components are re-

Figure 5.3 Power properties of the multivariate tests: H-models



(a) MARCH tests (Pearson) (b) MARCH tests (Spearman)

Note: M(org) denotes the original multivariate TSAY test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

ported in Tables 5.22 – 5.25. For higher clarity, the results are depicted in Figure 5.3 as well. The figure depicts the relationship between the average number of principal components (x-axis) and the standard deviation of the number of principal components (y-axis) of the multivariate tests calculated for each DGP configuration. The results suggest the following: (i) In contrast to MTSAY tests results, the average number of principal components of the MARCH tests is much more concentrated, regardless of the DGPs, the sample size T, and the configuration of the variance-covariance matrix $\mathbf{\Sigma}$. The reason for that is discussed in Section 4.1. There seems not to be a clear cutoff between the BIC approach and the Kaiser rule in selecting the number of principal components. Both rules produce almost identical results. So, the Kaiser rule might be preferred in this case. As in the previous cases, the variance rule is very inefficient and produces similar results to the original MARCH test; (ii) No differences are observed between the results of the multivariate tests based on the Pearson and Spearman correlation matrices.

5.5 Empirical Examples

Having analyzed the behaviour of the proposed tests using extensive Monte Carlo experiments, we can turn our attention to empirical applications of these principal component-based multivariate non-linearity tests. Two examples are provided in this section.

5.5.1 Macroeconomic Example

Following Cho and Moreno (2003), a simple rational expectations model with New Kayenesian features (e.g. the habit formation, the price indexation, etc) is considered in this section. The model describes the behaviour of three agents in the economy (i.e. households, firms, and government). It is assumed that the economy is populated by infinitely lived households who consume and supply labour to firms. Households are assumed to maximize an intertemporal (constant relative risk aversion) utility function subject to a budget constraint. Firms, completely owned by households, produce output using a simple production function. The log-linearized equilibrium (first-order) conditions can be written as follows

$$\begin{aligned}
\pi_t &= \mu_1 \mathbb{E}_t(\pi_{t+1}) + (1-\mu_1)\pi_{t-1} + \kappa y_t + x_t^{\pi}, \\
y_t &= \mu_2 \mathbb{E}_t(y_{t+1}) + (1-\mu_2) y_{t-1} - \omega^{-1}\left(r_t - \mathbb{E}_t(\pi_{t+1})\right) + x_t^{y}, \\
r_t &= \phi \pi_t + \gamma y_t + x_t^{r},
\end{aligned}$$

where π_t denotes the inflation rate, y_t denotes the output, and r_t is the short-term interest rate.[12] All exogenous processes are assumed to follow a simple AR(1) process. The above model can be written in the companion matrix form as follows

$$\begin{aligned}
\mathbf{G}\mathbf{y}_t &= \mathbf{F}\mathbb{E}_t(\mathbf{y}_{t+1}) + \mathbf{H}\mathbf{y}_{t-1} + \mathbf{L}\mathbf{x}_t, \\
\mathbf{x}_t &= \mathbf{N}\mathbf{x}_{t-1} + \mathbf{a}_t,
\end{aligned}$$

where $\mathbf{y}_t = (\pi_t, y_t, r_t)'$ is a vector of dependent variables and $\mathbf{x}_t = (x_t^{\pi}, x_t^{y}, x_t^{r})'$ is a vector of (unobserved) exogenous variables. Uhlig (1995) proposed a simple solution of the model based on a method of undetermined coefficients. It can be shown that

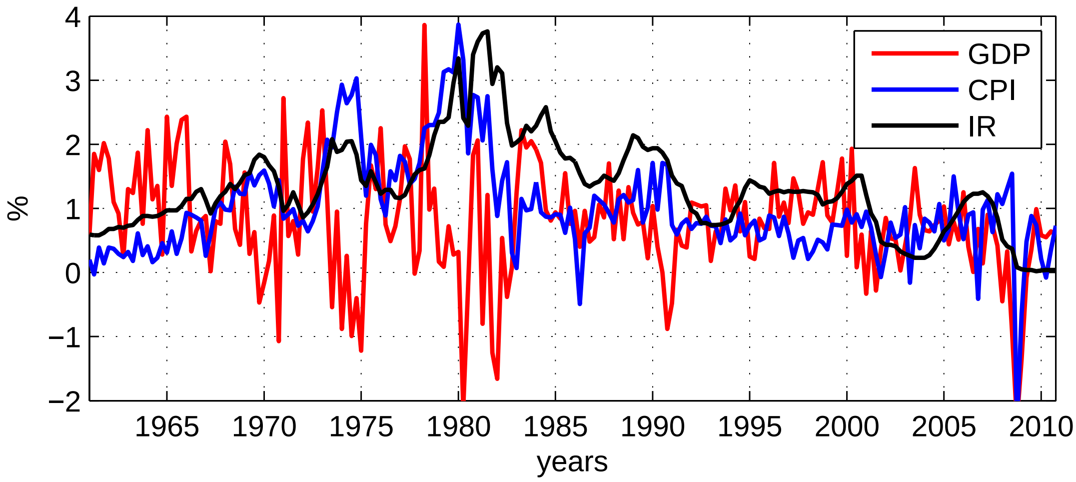
[12] All variables represent deviations from steady state values.

the unobserved (latent) shock variables may be substituted out of the system. The resulting representation is then a simple 3-variate VAR(2) model given by

$$\mathbf{y}_t = \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \mathbf{\Theta} \mathbf{a}_t. \tag{5.14}$$

Three quarterly US economic indicators are considered for $\mathbf{y}$: the output variable y is approximated by the growth rate of real US GDP (GDP); the inflation rate π is approximated by the US CPI inflation rate (CPI), and the policy rate r by the 3M treasure bill rate (IR). The economic time series span the period 1961Q1 – 2010Q4.[13] The series are depicted in Figure 5.4. The appropriate filter for the model in (5.14) seems to be a VAR(2) model. However, it is important to emphasize that the order of the filter is affected by an ad-hoc decision that shock variables follow an AR(1) process. This assumption may not be correct in general. For this reason, an automatic lag order selection procedure to determine the lag order of a VAR filter is implemented as well. The procedure indicates $P = 3$, rather than $P = 2$, for the MTSAY test and $Q = 2$ for the MARCH test.

Figure 5.4 US macroeconomic time series



The p-values of the univariate and multivariate TSAY and ARCH tests are presented in Table 5.6. The results suggest the following: (i) The null hypothesis about

[13] The last observations of the GDP data are not considered deliberately in order to avoid the impact of statistical revisions on the results.

linearity is rejected for both CPI and IR by at least one of the univariate test. However, this is not the case for GDP, the null hypothesis failing to be rejected by none of the univariate non-linearity test at the significance level 0.05; (ii) The null hypothesis is rejected by all MTSAY tests even at the significance level 0.01, regardless of the correlation matrix and the stoping rule. However, the MARCH test results differ depending on the correlation matrix used to calculate principal components. In particular, the null is rejected by all MARCH tests with principal components calculated from the Pearson correlation matrix at the significance level 0.05, whereas in only one case when principal components are calculated from the Spearman rank correlation matrix. A contradiction between Spearman-based and Pearson-based multivariate tests may indicate the presence of aberrant observations, which can inflate the power of Pearson-based multivariate non-linearity tests; (iii) The automatically selected number of principal components varies from 3 to 9, depending on the test and the stopping rule. Either way, even the largest number of principal components (i.e. $n = 9$) is still significantly lower as compared to a dimension of the original MTSAY test (i.e. $s = 45$).

The above results are based on the whole sample consisting of $T = 200$ observations. It might be also interesting to assess the robustness of the tests against the sample size. For this purpose, a rolling-window approach is applied to check the stability of the univariate and multivariate TSAY and ARCH tests. The method is based on splitting the original sample into 101 overlapping sub-samples. Each sub-sample consists of only 100 consecutive observations. It means that the first sub-sample (window) span the period 1961Q1 – 1985Q4, the second one 1961Q2 – 1985Q2, etc. The lag order of a VAR filter is determined using the BIC for each rolling window. The results are depicted in a graphical form in Figures 5.6 – 5.7. The results suggest the following: (i) The power of univariate tests is extremely period (window) dependent, whereas the behaviour of the multivariate tests is significantly more robust; (ii) The number of automatically selected number of components is almost constant over time, whereas a number of additional variables in the original MTSAY and MARCH varies significantly.

Table 5.6 P-values of the non-linearity tests

	univariate tests			**multivariate tests**	
filter	AR(2)	AR(3)	AR(6)	VAR(3)	
tests/variables	GDP	CPI	IR	$\mathbf{y}$ =(CPI,GDP,IR)	n
TSAY	0.760	0.028	0.000	–	–
MTSAY(org)	–	–	–	0.000	45
MTSAY(p-bic)	–	–	–	0.000	5
MTSAY(p-0.9)	–	–	–	0.000	6
MTSAY(p-k)	–	–	–	0.000	6
MTSAY(s-bic)	–	–	–	0.000	5
MTSAY(s-0.9)	–	–	–	0.000	5
MTSAY(s-k)	–	–	–	0.000	6
ARCH	0.318	0.908	0.000	–	–
MARCH(org)	–	–	–	0.000	12
MARCH(p-bic)	–	–	–	0.005	3
MARCH(p-0.9)	–	–	–	0.000	6
MARCH(p-k)	–	–	–	0.017	4
MARCH(s-bic)	–	–	–	0.296	3
MARCH(s-0.9)	–	–	–	0.000	9
MARCH(s-k)	–	–	–	0.125	5

[a] GDP denotes the growth rate of real US GDP, CPI represents the US CPI inflation rate, and IR stands for the US 3M treasury bill rate, and n denotes the number of principal components selected by the stopping rules and/or the number of additional variables required by the original multivariate tests.

[b] M(org) denotes the original multivariate TSAY test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule. n denotes a number of additional variables and/or principal components.

Finally, it can be concluded that using linear multivariate economic models (e.g. VAR and/or DSGE models) such as in Smets and Wouters (2002), Smets and Wouters (2003), Smets and Wouters (2007), Del Negro et al. (2007), Adolfson et al. (2007), Adolfson et al. (2008), or Adolfson et al. (2008) is in sharp contrast with our empirical findings, and, thus, highly questionable.

5.5.2 Financial Example

Following Diebold and Li (2006) and Diebold et al. (2006), a dynamic Nelson-Siegel model is considered as a representative model of the US yield curve. The model can be formally written as follows

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-\exp\{-\lambda_t\tau\}}{\lambda_t\tau}\right) + \beta_{3t}\left(\frac{1-\exp\{-\lambda_t\tau\}}{\lambda_t\tau} - \exp\{-\lambda_t\tau\}\right) + u_t(\tau), \tag{5.15}$$

where $y_t(\tau)$ denotes the yield at time t with maturity τ, λ_t is the decay parameter, $u_t(\tau)$ describes an error term, and $\beta_{1t}, \beta_{2t}, \beta_{3t}$ are factor loadings representing a level, slope, and curvature effect of the yield curve. Keeping the decay parameter fixed (i.e. $\lambda_t = \lambda$), the unknown parameter vector $\boldsymbol{\beta}_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})'$ can be easily estimated by the OLS/WLS method.[14] It is important to emphasize that although a no-arbitrage condition does not hold in the original Nelson-Siegel model, it posses no problem from the non-linearity testing point of view.[15] Under the null hypothesis of linearity, the factor loadings are assumed to follow some finite order VAR model given by

$$\boldsymbol{\beta}_t = \boldsymbol{\Phi}_0 + \sum_{i=1}^{P} \boldsymbol{\Phi}_i \boldsymbol{\beta}_{t-i} + \mathbf{a}_t, \tag{5.16}$$

[14] The WLS method seems to be preferred compared to OLS due to existing residual heteroscedasticity.

[15] Christensen et al. (2011) derived an arbitrage-free version of the dynamic Neslon-Siegel model. The modification consists in introducing a correction term which varies only across maturities but is constant over time. This means that the correction term itself has no effect on the dynamics of the estimated loading factors, and, thus, on testing for non-linearity. Therefore, the issue related to an arbitrage condition is omitted in our case.

where $\mathbf{\Phi}_0$ is a (3×1) constant vector and $\mathbf{\Phi}_i$, for $i \in \{1, \ldots, P\}$, are (3×3) parameter matrices.

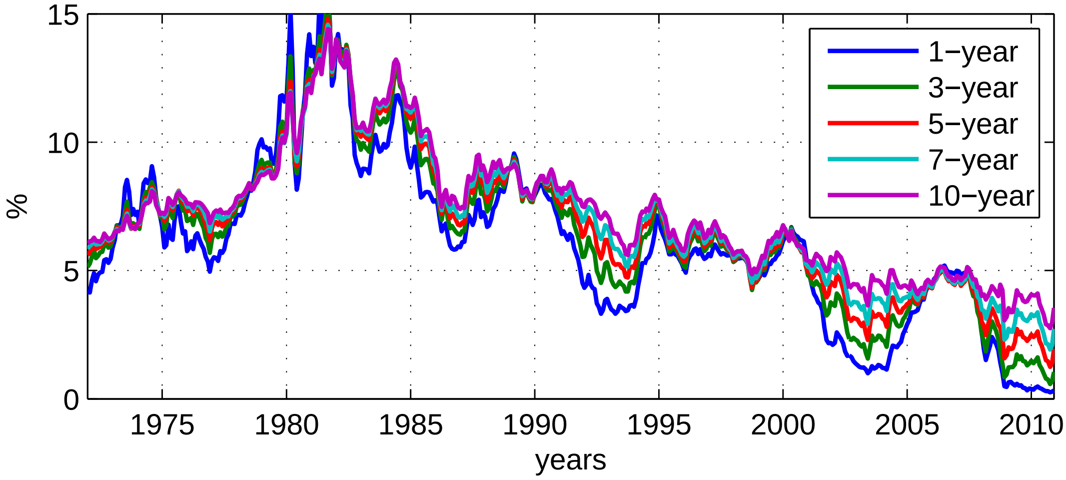
The monthly US interest rate data covering the maturities from 1 to 15 years are considered. The interested reader is referred to Gürkaynak et al. (2007) for methodological notes about the data set. The data span the period 1972M1 – 2010M12. The selected interest rates and the estimated factor loadings are depicted in Figure 5.5. An automatic lag order selection procedure to determine the lag order of a VAR filter is implemented as in the previous example. The procedure indicates the lag order $P = 4$ and $Q = 4$.

The p-values of the univariate and multivariate TSAY and ARCH tests are presented in Table 5.7. These suggest the following: (i) The null hypothesis of linearity of a vector of factor loadings is rejected by all univariate and multivariate TSAY and ARCH tests at the significance level 0.05; (ii) As in the previous example, the principal component based tests lead to an extraordinary dimensionality reduction without any inferential differences. In particular, the automatically selected number of principal components varies from 3 to 18, depending on the test and the stopping rule. Either way, even the largest number of principal components (i.e. $n = 18$) is still significantly lower as compared to a dimension of the original MTSAY test (i.e. $s = 78$).

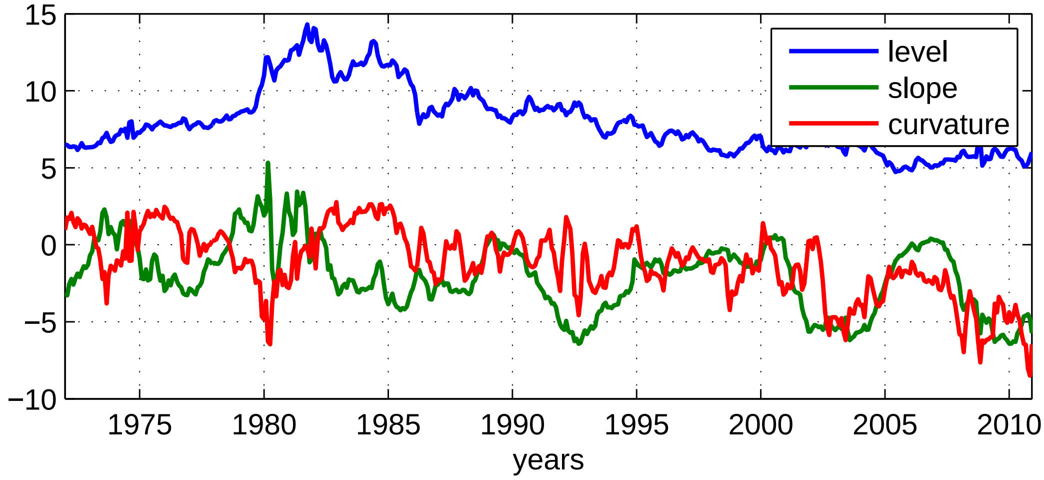
Finally, it can be concluded that using linear multivariate dynamic Nelson-Siegel models such as in Diebold and Li (2006) and Diebold et al. (2006) is in sharp contrast with our empirical findings, and, thus, highly questionable.

The above results are based on the whole sample consisting of $T = 492$ monthly observations. It might be also interesting to assess the robustness of the univariate and multivariate tests against the sample size. For this purpose, a rolling-window approach is applied. The method is based on splitting the original sample into 168 overlapping sub-samples. Each sub-sample consists of only 300 consecutive

Figure 5.5 US financial time series



(a) Interest rates with different maturities



(b) Estimated loadings of the yiled curve

Table 5.7 P-values of the non-linearity tests

	univariate tests			multivariate tests	
filter	AR(3)	AR(3)	AR(6)	VAR(4)	
tests/variables	level	shift	curvature	$\boldsymbol{\beta}$ =(level,shift,curvature)	n
TSAY	0.043	0.000	0.002	–	–
MTSAY(org)	–	–	–	0.000	78
MTSAY(p-bic)	–	–	–	0.023	3
MTSAY(p-0.9)	–	–	–	0.023	3
MTSAY(p-k)	–	–	–	0.000	7
MTSAY(s-bic)	–	–	–	0.032	3
MTSAY(s-0.9)	–	–	–	0.029	4
MTSAY(s-k)	–	–	–	0.011	6
ARCH	0.000	0.000	0.000	–	–
MARCH(org)	–	–	–	0.000	24
MARCH(p-bic)	–	–	–	0.000	11
MARCH(p-0.9)	–	–	–	0.000	14
MARCH(p-k)	–	–	–	0.000	9
MARCH(s-bic)	–	–	–	0.000	6
MARCH(s-0.9)	–	–	–	0.000	18
MARCH(s-k)	–	–	–	0.000	10

[a] M(org) denotes the original multivariate TSAY test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule. n denotes a number of additional variables and/or principal components.

monthly observations. It means that the first sub-sample (window) span the period 1972M2 – 1996M12, the second one 1972M2 – 1997M1, etc. The lag order of a VAR filter is determined using the BIC for each rolling window. The results are depicted in a graphical form in Figures 5.8 – 5.9. The results suggest the following: (i) The univariate and multivariate ARCH tests do produce very similar results, whereas the TSAY tests differ considerably. The multivariate test, however, do produce more robust results as compared to the univariate counterparts; (ii) As in the previous example, the number of automatically selected number of components is almost constant over time, whereas a number of additional variables in the original MTSAY and MARCH varies significantly.

5.6 Conclusion

It has been demonstrated that the issue of dimensionality and possible multicolinearity in the multivariate non-linearity tests can be easily bypassed using a principal component analysis. All the modified principal component-based non-linearity tests do exhibit very good size and power properties. In particular, the principal component-based tests do offer a remarkable dimensionality reduction (in average about 70 %) without any systematic power distortion. Nevertheless, special care should be exercised to stopping rules determining the number of components. The Monte Carlo results suggest that the BIC rule performs best as compared to the variance rule and the Kaiser rule, although the results are test dependent. Our Monte Carlo results also clearly confirm that the univariate non-linearity tests might not be adequate tools for testing for non-linearity in the case of multivariate time series. The univariate tests can suffer from a serious power distortion, and, thus, can lead to misleading inference.

Finally, it can be concluded that using linear dynamic time series models (e.g. VAR-type models) for modelling both macroeconomic and financial variables is in sharp contrast with our empirical findings. Put differently, omitting non-linearity in both macroeconomic and financial indicators may cause serious problems for a central bank when conducting monetary policy in practice.

5.7 Appendix A: Tables

Table 5.8 Size properties of the TSAY tests: M-models

		T=200			**T=500**			**T=1000**		
matrix	test	M1	M2	M3	M1	M2	M3	M1	M2	M3
$\mathbf{\Sigma}_1$	M(org)	0.059	0.040	0.051	0.034	0.055	0.044	0.059	0.046	0.044
	M(p-bic)	0.057	0.037	0.050	0.045	0.045	0.047	0.048	0.052	0.039
	M(p-0.9)	0.042	0.042	0.047	0.041	0.041	0.040	0.042	0.044	0.039
	M(p-k)	0.053	0.042	0.046	0.047	0.047	0.051	0.050	0.052	0.034
	M(s-bic)	0.048	0.045	0.056	0.051	0.048	0.043	0.037	0.051	0.032
	M(s-0.9)	0.049	0.043	0.047	0.042	0.052	0.038	0.043	0.039	0.044
	M(s-k)	0.048	0.045	0.051	0.051	0.043	0.049	0.052	0.056	0.037
	X_1	0.041	0.050	0.041	0.038	0.064	0.053	0.046	0.051	0.041
	X_2	0.037	0.046	0.070	0.050	0.048	0.060	0.056	0.061	0.047
$\mathbf{\Sigma}_2$	M(org)	0.058	0.041	0.052	0.057	0.041	0.046	0.053	0.056	0.050
	M(p-bic)	0.053	0.053	0.057	0.062	0.042	0.044	0.054	0.044	0.047
	M(p-0.9)	0.042	0.049	0.053	0.061	0.044	0.043	0.050	0.045	0.045
	M(p-k)	0.034	0.048	0.052	0.064	0.042	0.042	0.052	0.042	0.040
	M(s-bic)	0.051	0.051	0.051	0.055	0.044	0.044	0.051	0.049	0.051
	M(s-0.9)	0.042	0.042	0.057	0.052	0.048	0.039	0.048	0.044	0.039
	M(s-k)	0.039	0.049	0.049	0.061	0.045	0.037	0.049	0.046	0.037
	X_1	0.034	0.049	0.058	0.055	0.058	0.066	0.039	0.054	0.054
	X_2	0.021	0.050	0.053	0.045	0.054	0.054	0.052	0.057	0.058

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.9 Size properties of the ARCH tests: M-models

		T=200			T=500			T=1000		
matrix	test	M1	M2	M3	M1	M2	M3	M1	M2	M3
$\boldsymbol{\Sigma}_1$	M(org)	0.063	0.067	0.062	0.050	0.059	0.055	0.068	0.053	0.050
	M(p-bic)	0.066	0.070	0.067	0.047	0.051	0.058	0.048	0.059	0.061
	M(p-0.9)	0.062	0.072	0.063	0.050	0.057	0.057	0.057	0.052	0.060
	M(p-k)	0.064	0.074	0.063	0.052	0.053	0.057	0.050	0.056	0.061
	M(s-bic)	0.066	0.072	0.072	0.043	0.047	0.064	0.051	0.054	0.063
	M(s-0.9)	0.062	0.070	0.065	0.047	0.051	0.061	0.054	0.056	0.065
	M(s-k)	0.056	0.070	0.061	0.045	0.046	0.063	0.051	0.057	0.065
	X_1	0.049	0.055	0.069	0.054	0.049	0.048	0.052	0.044	0.059
	X_2	0.047	0.050	0.064	0.056	0.061	0.058	0.055	0.057	0.066
$\boldsymbol{\Sigma}_2$	M(org)	0.050	0.068	0.061	0.056	0.050	0.058	0.065	0.046	0.045
	M(p-bic)	0.050	0.063	0.068	0.060	0.055	0.054	0.060	0.047	0.056
	M(p-0.9)	0.052	0.066	0.057	0.055	0.053	0.051	0.059	0.048	0.056
	M(p-k)	0.049	0.060	0.057	0.055	0.049	0.049	0.056	0.046	0.051
	M(s-bic)	0.052	0.066	0.068	0.059	0.053	0.052	0.062	0.050	0.055
	M(s-0.9)	0.049	0.062	0.060	0.058	0.054	0.052	0.061	0.049	0.054
	M(s-k)	0.049	0.059	0.056	0.057	0.047	0.048	0.058	0.049	0.051
	X_1	0.051	0.050	0.046	0.052	0.057	0.044	0.063	0.056	0.065
	X_2	0.058	0.042	0.051	0.052	0.054	0.049	0.052	0.047	0.061

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.10 Number of additional variables/components of the MTSAY tests: M-models

		T=200			T=500			T=1000		
matrix	test	M1	M2	M3	M1	M2	M3	M1	M2	M3
$\mathbf{\Sigma}_1$	M(org)	11.3	6.2	6.0	10.8	8.7	8.8	10.6	11.4	11.1
	M(p-bic)	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	M(p-0.9)	6.0	3.4	3.3	5.9	4.7	4.8	6.0	6.1	6.0
	M(p-k)	3.8	2.5	2.5	3.3	2.9	2.9	3.1	3.3	3.2
	M(s-bic)	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	M(s-0.9)	7.0	3.9	3.8	6.7	5.8	5.8	6.5	7.7	7.6
	M(s-k)	3.9	2.6	2.5	3.3	2.9	2.9	3.2	3.3	3.3
$\mathbf{\Sigma}_2$	M(org)	11.5	5.1	5.8	10.7	5.7	8.0	10.7	7.1	11.6
	M(p-bic)	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	M(p-0.9)	5.0	2.3	3.3	4.9	2.4	4.4	5.0	2.6	6.1
	M(p-k)	3.3	2.1	3.1	3.1	2.1	4.3	3.1	2.1	6.1
	M(s-bic)	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	M(s-0.9)	6.5	2.8	3.6	6.3	3.2	5.1	6.3	3.7	7.5
	M(s-k)	3.5	2.2	3.2	3.3	2.2	4.5	3.2	2.2	6.3

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.11 Number of additional variables/components of the MARCH tests: M-models

		T=200			T=500			T=1000		
matrix	test	M1	M2	M3	M1	M2	M3	M1	M2	M3
$\mathbf{\Sigma}_1$	M(org)	4.3	4.1	4.0	3.8	3.9	3.7	3.8	3.8	3.7
	M(p-bic)	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	M(p-0.9)	3.1	3.0	2.9	2.7	2.8	2.6	2.7	2.7	2.6
	M(p-k)	2.5	2.4	2.4	2.3	2.4	2.3	2.3	2.3	2.2
	M(s-bic)	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	M(s-0.9)	3.1	3.0	2.9	2.7	2.8	2.6	2.7	2.7	2.6
	M(s-k)	2.5	2.4	2.4	2.3	2.4	2.3	2.3	2.3	2.3
$\mathbf{\Sigma}_2$	M(org)	4.1	4.3	3.9	3.8	3.8	3.9	3.8	3.7	3.7
	M(p-bic)	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	M(p-0.9)	2.6	2.6	2.5	2.4	2.4	2.5	2.4	2.3	2.4
	M(p-k)	2.2	2.2	2.2	2.1	2.1	2.2	2.1	2.1	2.1
	M(s-bic)	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	M(s-0.9)	2.7	2.9	2.6	2.5	2.6	2.6	2.6	2.5	2.5
	M(s-k)	2.2	2.3	2.2	2.1	2.1	2.2	2.1	2.1	2.1

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.12 Variability of the number of additional variables/components of the TSAY tests: M-models

		T=200			**T=500**			**T=1000**		
matrix	test	M1	M2	M3	M1	M2	M3	M1	M2	M3
$\boldsymbol{\Sigma}_1$	M(org)	5.2	5.9	5.6	4.0	5.8	5.6	3.3	5.6	5.1
	M(p-bic)	0.1	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0
	M(p-0.9)	2.1	2.7	2.5	1.7	2.9	2.8	1.4	3.0	2.8
	M(p-k)	1.5	1.2	1.0	1.1	1.0	1.0	0.9	1.0	0.9
	M(s-bic)	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0
	M(s-0.9)	2.7	3.4	3.3	2.3	3.8	3.6	2.0	3.6	3.3
	M(s-k)	1.6	1.3	1.2	1.2	1.2	1.1	1.1	1.1	1.0
$\boldsymbol{\Sigma}_2$	M(org)	5.5	5.5	5.8	3.3	4.5	6.5	3.2	4.4	6.8
	M(p-bic)	0.1	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0
	M(p-0.9)	2.0	1.1	2.7	1.3	0.9	3.2	1.3	0.8	3.6
	M(p-k)	1.1	0.7	2.2	0.6	0.5	2.7	0.5	0.5	2.6
	M(s-bic)	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.0
	M(s-0.9)	2.5	2.2	3.4	1.7	2.0	4.0	1.6	1.9	4.3
	M(s-k)	1.4	1.0	2.4	0.8	0.7	2.8	0.7	0.7	2.8

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.13 Variability of the number of additional variables/components of the MARCH tests: M-models

matrix	test	T=200			T=500			T=1000		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
$\mathbf{\Sigma}_1$	M(org)	3.1	3.0	2.7	2.4	2.7	2.3	2.4	2.5	2.1
	M(p-bic)	0.1	0.1	0.1	0.1	0.0	0.1	0.0	0.1	0.1
	M(p-0.9)	2.6	2.5	2.2	2.1	2.3	2.0	2.1	2.2	1.9
	M(p-k)	1.3	1.3	1.1	1.0	1.1	0.9	1.1	1.1	0.9
	M(s-bic)	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.1	0.0
	M(s-0.9)	2.6	2.5	2.3	2.1	2.4	2.0	2.1	2.2	1.9
	M(s-k)	1.4	1.3	1.1	1.0	1.2	1.0	1.0	1.1	0.9
$\mathbf{\Sigma}_2$	M(org)	2.9	3.1	2.7	2.4	2.5	2.7	2.3	2.3	2.4
	M(p-bic)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0
	M(p-0.9)	1.6	1.7	1.5	1.3	1.4	1.5	1.3	1.3	1.4
	M(p-k)	0.8	0.8	0.7	0.6	0.6	0.7	0.5	0.5	0.6
	M(s-bic)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0
	M(s-0.9)	2.0	2.2	1.9	1.7	1.8	2.0	1.7	1.6	1.8
	M(s-k)	0.8	0.9	0.8	0.6	0.6	0.7	0.5	0.5	0.6

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.14 Power properties of the TSAY tests: N-models

matrix	test	T=200						T=500						T=1000					
		N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6
$\mathbf{\Sigma}_1$	M(org)	0.34	0.36	0.81	0.47	1.00	0.49	0.76	0.77	0.99	0.87	1.00	0.53	0.98	0.97	1.00	0.99	1.00	0.58
	M(p-bic)	0.15	0.20	0.81	0.47	0.99	0.33	0.28	0.33	0.98	0.84	1.00	0.34	0.42	0.40	1.00	0.97	1.00	0.35
	M(p-0.9)	0.29	0.30	0.81	0.46	1.00	0.36	0.66	0.69	0.99	0.88	1.00	0.41	0.96	0.94	1.00	0.99	1.00	0.45
	M(p-k)	0.23	0.23	0.81	0.44	0.99	0.35	0.49	0.53	0.99	0.80	1.00	0.38	0.80	0.77	1.00	0.97	1.00	0.41
	M(s-bic)	0.16	0.17	0.81	0.47	0.99	0.30	0.26	0.30	0.98	0.84	1.00	0.32	0.38	0.38	1.00	0.98	1.00	0.34
	M(s-0.9)	0.34	0.34	0.82	0.46	1.00	0.40	0.74	0.74	0.99	0.87	1.00	0.44	0.98	0.97	1.00	0.99	1.00	0.49
	M(s-k)	0.23	0.23	0.81	0.44	1.00	0.35	0.47	0.53	0.98	0.81	1.00	0.38	0.78	0.77	1.00	0.98	1.00	0.41
	X_1	0.34	0.06	0.07	0.10	0.99	0.21	0.72	0.06	0.06	0.17	1.00	0.25	0.94	0.09	0.07	0.32	1.00	0.27
	X_2	0.26	0.06	0.05	0.10	0.32	0.38	0.63	0.11	0.05	0.19	0.53	0.43	0.92	0.14	0.06	0.31	0.74	0.48
$\mathbf{\Sigma}_2$	M(org)	0.39	0.40	0.81	0.25	1.00	0.49	0.83	0.81	0.99	0.54	1.00	0.51	0.99	0.99	1.00	0.80	1.00	0.59
	M(p-bic)	0.24	0.22	0.79	0.23	1.00	0.37	0.38	0.38	0.96	0.43	1.00	0.33	0.33	0.37	1.00	0.56	1.00	0.36
	M(p-0.9)	0.34	0.36	0.81	0.23	1.00	0.39	0.73	0.73	0.99	0.51	1.00	0.40	0.97	0.96	1.00	0.82	1.00	0.46
	M(p-k)	0.33	0.32	0.81	0.23	1.00	0.36	0.73	0.73	0.99	0.51	1.00	0.36	0.97	0.96	1.00	0.83	1.00	0.42
	M(s-bic)	0.21	0.20	0.78	0.24	1.00	0.34	0.32	0.31	0.93	0.46	1.00	0.32	0.24	0.25	0.98	0.65	1.00	0.36
	M(s-0.9)	0.35	0.35	0.81	0.23	1.00	0.41	0.74	0.75	0.99	0.52	1.00	0.43	0.98	0.97	1.00	0.84	1.00	0.48
	M(s-k)	0.32	0.32	0.81	0.22	1.00	0.35	0.73	0.71	0.99	0.51	1.00	0.36	0.97	0.96	1.00	0.84	1.00	0.42
	X_1	0.33	0.05	0.06	0.09	0.99	0.19	0.69	0.06	0.07	0.13	1.00	0.24	0.91	0.15	0.12	0.29	1.00	0.26
	X_2	0.29	0.07	0.06	0.10	0.98	0.41	0.62	0.11	0.08	0.14	1.00	0.43	0.91	0.21	0.12	0.29	1.00	0.51

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.15 Power properties of the ARCH tests: N-models

		T=200						**T=500**						**T=1000**					
matrix	test	N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6
$\mathbf{\Sigma}_1$	M(org)	0.11	0.12	0.18	0.09	0.88	0.94	0.16	0.19	0.32	0.12	1.00	1.00	0.29	0.32	0.49	0.13	1.00	1.00
	M(p-bic)	0.12	0.12	0.16	0.09	0.85	0.94	0.16	0.19	0.30	0.11	0.99	1.00	0.26	0.31	0.45	0.11	1.00	1.00
	M(p-0.9)	0.11	0.11	0.17	0.08	0.86	0.94	0.16	0.19	0.29	0.11	1.00	1.00	0.27	0.31	0.45	0.12	1.00	1.00
	M(p-k)	0.10	0.11	0.16	0.08	0.84	0.94	0.15	0.19	0.29	0.10	0.99	1.00	0.26	0.31	0.44	0.11	1.00	1.00
	M(s-bic)	0.12	0.11	0.17	0.09	0.84	0.94	0.15	0.20	0.28	0.10	0.99	1.00	0.26	0.30	0.42	0.12	1.00	1.00
	M(s-0.9)	0.11	0.10	0.16	0.08	0.85	0.95	0.15	0.20	0.27	0.10	0.99	1.00	0.26	0.30	0.43	0.12	1.00	1.00
	M(s-k)	0.11	0.10	0.15	0.08	0.83	0.94	0.15	0.20	0.27	0.09	0.99	1.00	0.25	0.30	0.41	0.11	1.00	1.00
	X_1	0.10	0.06	0.05	0.07	0.44	0.86	0.15	0.07	0.07	0.09	0.82	1.00	0.23	0.06	0.07	0.11	0.98	1.00
	X_2	0.10	0.05	0.06	0.06	0.71	0.94	0.14	0.06	0.07	0.08	0.96	1.00	0.23	0.06	0.08	0.10	1.00	1.00
$\mathbf{\Sigma}_2$	M(org)	0.11	0.13	0.14	0.09	0.85	0.94	0.17	0.17	0.22	0.10	1.00	1.00	0.27	0.29	0.41	0.14	1.00	1.00
	M(p-bic)	0.12	0.13	0.15	0.10	0.81	0.96	0.17	0.18	0.24	0.10	0.99	1.00	0.29	0.30	0.42	0.16	1.00	1.00
	M(p-0.9)	0.11	0.13	0.15	0.09	0.81	0.96	0.18	0.19	0.24	0.10	1.00	1.00	0.30	0.31	0.43	0.15	1.00	1.00
	M(p-k)	0.11	0.12	0.15	0.09	0.79	0.95	0.17	0.18	0.23	0.10	1.00	1.00	0.29	0.30	0.42	0.16	1.00	1.00
	M(s-bic)	0.12	0.14	0.15	0.10	0.82	0.96	0.17	0.18	0.24	0.10	0.99	1.00	0.29	0.31	0.42	0.16	1.00	1.00
	M(s-0.9)	0.11	0.13	0.15	0.09	0.81	0.96	0.18	0.19	0.24	0.10	1.00	1.00	0.29	0.31	0.43	0.15	1.00	1.00
	M(s-k)	0.11	0.12	0.15	0.09	0.80	0.95	0.17	0.18	0.23	0.10	1.00	1.00	0.29	0.31	0.42	0.16	1.00	1.00
	X_1	0.11	0.06	0.05	0.09	0.43	0.88	0.13	0.07	0.07	0.09	0.82	1.00	0.22	0.09	0.08	0.10	0.98	1.00
	X_2	0.08	0.06	0.06	0.06	0.63	0.96	0.15	0.07	0.05	0.09	0.93	1.00	0.23	0.08	0.07	0.10	1.00	1.00

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.16 Number of additional variables/principal components of the MTSAY tests: N-models

		T=200						**T=500**						**T=1000**					
matrix	test	N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6
$\mathbf{\Sigma}_1$	M(org)	13.1	12.7	4.4	6.2	5.1	13.9	14.4	14.7	4.8	9.0	5.4	14.6	18.5	19.5	6.3	11.2	5.5	15.5
	M(p-bic)	2.1	2.0	2.1	2.0	3.1	2.1	2.1	2.1	2.2	2.1	4.0	2.1	2.2	2.2	2.4	2.2	4.7	2.0
	M(p-0.9)	7.7	7.6	2.7	3.4	2.9	4.8	8.7	9.0	3.0	4.8	3.1	5.2	11.1	12.0	3.8	6.0	3.1	5.7
	M(p-k)	4.3	4.3	2.3	2.5	2.2	3.4	4.4	4.6	2.3	2.9	2.2	3.7	5.1	5.7	2.6	3.2	2.1	4.0
	M(s-bic)	2.1	2.1	2.1	2.0	3.1	2.1	2.1	2.1	2.2	2.1	4.0	2.0	2.2	2.2	2.5	2.2	4.8	2.0
	M(s-0.9)	8.9	8.6	2.9	4.0	3.2	6.9	10.0	10.2	3.3	6.0	3.5	7.6	12.9	13.7	4.3	7.7	3.7	8.2
	M(s-k)	4.6	4.5	2.3	2.6	2.3	3.9	4.7	4.8	2.4	3.0	2.2	4.0	5.6	6.1	2.6	3.3	2.2	4.1
$\mathbf{\Sigma}_2$	M(org)	13.5	14.2	4.8	5.8	5.4	13.7	18.0	18.2	5.5	8.3	4.8	14.4	24.6	23.9	6.8	11.6	5.7	15.9
	M(p-bic)	2.1	2.1	2.1	2.0	3.0	2.1	2.1	2.1	2.2	2.0	3.9	2.1	2.2	2.2	2.8	2.1	5.0	2.1
	M(p-0.9)	8.0	8.3	2.8	3.3	2.5	4.7	10.5	10.7	3.2	4.5	2.3	5.1	13.8	13.6	3.9	6.0	2.5	5.8
	M(p-k)	5.6	5.8	2.7	3.2	2.2	3.4	7.6	7.7	3.1	4.5	2.1	3.6	10.3	10.1	3.8	6.1	2.1	4.0
	M(s-bic)	2.1	2.1	2.1	2.0	3.0	2.1	2.1	2.1	2.3	2.0	3.8	2.1	2.1	2.1	3.0	2.1	4.9	2.1
	M(s-0.9)	9.0	9.4	3.1	3.7	3.1	6.8	12.0	12.1	3.5	5.3	2.9	7.5	16.2	15.7	4.4	7.4	3.3	8.3
	M(s-k)	6.0	6.2	2.8	3.2	2.3	3.8	8.0	8.1	3.3	4.6	2.1	3.9	10.8	10.6	4.0	6.3	2.1	4.2

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.17 Number of additional variables/principal components of the MARCH tests: N-models

		T=200						**T=500**						**T=1000**					
matrix	test	N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6
$\mathbf{\Sigma}_1$	M(org)	4.1	4.3	4.0	4.1	4.3	4.9	3.8	3.9	3.8	3.8	4.3	5.0	3.7	3.7	3.6	3.8	4.6	5.7
	M(p-bic)	2.0	2.0	2.0	2.0	2.2	2.3	2.0	2.0	2.0	2.0	2.6	2.7	2.0	2.0	2.0	2.0	3.4	3.5
	M(p-0.9)	3.0	3.1	2.9	3.0	3.0	2.6	2.7	2.8	2.7	2.7	3.0	2.6	2.6	2.6	2.6	2.7	3.4	2.8
	M(p-k)	2.4	2.5	2.4	2.4	2.4	2.3	2.3	2.3	2.3	2.3	2.4	2.3	2.3	2.3	2.2	2.3	2.5	2.3
	M(s-bic)	2.0	2.0	2.0	2.0	2.2	2.3	2.0	2.0	2.0	2.0	2.6	2.7	2.0	2.0	2.0	2.0	3.5	3.5
	M(s-0.9)	3.0	3.1	2.9	3.0	3.1	3.2	2.7	2.8	2.7	2.7	3.2	3.3	2.6	2.6	2.6	2.7	3.5	3.8
	M(s-k)	2.4	2.5	2.4	2.5	2.5	2.4	2.3	2.4	2.3	2.3	2.4	2.3	2.3	2.3	2.2	2.3	2.6	2.3
$\mathbf{\Sigma}_2$	M(org)	4.1	4.2	4.1	4.1	4.3	4.8	3.7	3.8	3.8	3.8	4.3	5.3	3.6	3.8	3.7	3.7	4.3	5.8
	M(p-bic)	2.0	2.0	2.0	2.0	2.2	2.3	2.0	2.0	2.0	2.0	2.6	2.7	2.0	2.0	2.0	2.0	3.1	3.6
	M(p-0.9)	2.6	2.6	2.6	2.5	2.7	2.6	2.4	2.4	2.4	2.4	2.7	2.7	2.3	2.4	2.3	2.3	2.7	2.8
	M(p-k)	2.2	2.2	2.2	2.2	2.3	2.3	2.1	2.1	2.1	2.1	2.3	2.4	2.1	2.1	2.1	2.1	2.2	2.4
	M(s-bic)	2.0	2.0	2.0	2.0	2.3	2.3	2.0	2.0	2.0	2.0	2.6	2.8	2.0	2.0	2.0	2.0	3.1	3.6
	M(s-0.9)	2.8	2.8	2.7	2.7	2.9	3.1	2.5	2.6	2.6	2.6	2.9	3.5	2.4	2.6	2.5	2.5	2.9	3.8
	M(s-k)	2.2	2.2	2.2	2.2	2.3	2.3	2.1	2.1	2.2	2.1	2.2	2.4	2.1	2.1	2.1	2.1	2.2	2.4

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.18 Variability of the number of additional variables/principal components of the MTSAY tests: N-models

		T=200						**T=500**						**T=1000**					
matrix	test	N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6
$\mathbf{\Sigma}_1$	M(org)	7.5	7.3	4.3	5.5	6.1	11.3	7.1	7.5	4.5	6.2	6.1	9.8	7.3	8.0	5.0	5.0	5.8	9.4
	M(p-bic)	0.2	0.2	0.3	0.2	0.9	0.4	0.3	0.3	0.4	0.3	2.1	0.3	0.5	0.5	0.6	0.6	3.0	0.3
	M(p-0.9)	3.7	3.4	2.1	2.4	2.5	3.4	3.8	4.0	2.4	3.0	2.8	3.3	4.0	4.4	2.7	2.7	2.6	3.5
	M(p-k)	2.1	1.9	0.9	1.0	1.1	2.2	1.8	2.0	0.9	1.1	1.0	2.0	1.8	2.2	1.0	0.8	0.8	1.9
	M(s-bic)	0.3	0.2	0.3	0.2	0.9	0.4	0.4	0.4	0.5	0.3	2.2	0.2	0.5	0.5	0.7	0.5	3.1	0.2
	M(s-0.9)	4.3	4.1	2.7	3.3	3.5	4.9	4.6	4.8	3.0	3.9	3.7	4.7	4.9	5.3	3.4	3.3	3.7	4.8
	M(s-k)	2.4	2.4	1.0	1.2	1.2	2.7	2.1	2.3	1.1	1.3	1.1	2.2	2.2	2.5	1.0	1.0	1.0	2.1
$\mathbf{\Sigma}_2$	M(org)	7.2	8.4	5.5	6.5	8.0	10.6	9.0	9.3	5.0	6.5	5.2	10.2	10.8	10.2	4.9	7.1	6.0	10.7
	M(p-bic)	0.3	0.2	0.3	0.1	0.8	0.4	0.4	0.4	0.6	0.2	2.2	0.3	0.6	0.6	1.1	0.3	3.4	0.3
	M(p-0.9)	3.3	3.8	2.6	2.9	2.0	3.1	4.4	4.5	2.5	3.2	1.4	3.5	5.3	5.1	2.5	3.7	1.5	3.8
	M(p-k)	2.6	3.0	1.9	2.3	1.1	2.0	3.5	3.5	2.1	2.7	0.6	2.1	4.1	3.9	2.2	2.8	0.6	2.1
	M(s-bic)	0.3	0.3	0.3	0.2	0.8	0.4	0.4	0.4	0.7	0.2	2.0	0.3	0.6	0.5	1.3	0.2	3.3	0.3
	M(s-0.9)	4.0	4.6	3.2	3.6	3.6	4.4	5.5	5.6	3.1	4.0	2.5	4.8	6.6	6.2	3.1	4.4	2.9	5.2
	M(s-k)	2.8	3.1	2.1	2.5	1.4	2.4	3.7	3.7	2.3	2.8	0.7	2.3	4.2	4.0	2.4	2.9	0.8	2.3

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.19 Variability of the number of additional variables/principal components of the MARCH tests: N-models

		T=200						**T=500**						**T=1000**					
matrix	test	N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6
$\boldsymbol{\Sigma}_1$	M(org)	3.0	3.2	2.7	3.0	3.0	3.6	2.5	2.6	2.3	2.4	2.8	3.3	2.2	2.2	2.1	2.3	2.9	3.3
	M(p-bic)	0.1	0.1	0.1	0.1	0.5	0.8	0.1	0.1	0.2	0.1	0.9	1.1	0.1	0.1	0.2	0.1	1.6	1.6
	M(p-0.9)	2.5	2.6	2.3	2.5	2.3	1.4	2.2	2.3	2.1	2.1	2.2	1.3	1.9	2.0	1.9	2.1	2.4	1.4
	M(p-k)	1.2	1.3	1.1	1.2	1.1	0.9	1.1	1.1	1.0	1.0	1.0	0.8	0.9	0.9	0.9	1.0	1.1	0.8
	M(s-bic)	0.1	0.1	0.2	0.2	0.5	0.8	0.1	0.1	0.2	0.1	0.9	1.2	0.1	0.1	0.2	0.1	1.6	1.5
	M(s-0.9)	2.5	2.7	2.3	2.5	2.5	2.3	2.2	2.3	2.1	2.1	2.4	2.2	2.0	2.0	1.9	2.1	2.6	2.2
	M(s-k)	1.3	1.3	1.2	1.3	1.2	0.9	1.1	1.1	1.0	1.0	1.1	0.8	0.9	0.9	0.9	1.0	1.2	0.8
$\boldsymbol{\Sigma}_2$	M(org)	3.1	2.9	2.8	2.8	3.1	3.4	2.3	2.6	2.5	2.5	2.8	3.6	2.2	2.6	2.2	2.3	2.5	3.4
	M(p-bic)	0.1	0.1	0.1	0.1	0.5	0.7	0.1	0.1	0.1	0.1	0.8	1.2	0.1	0.1	0.1	0.1	1.1	1.6
	M(p-0.9)	1.7	1.6	1.6	1.5	1.8	1.4	1.3	1.4	1.4	1.4	1.6	1.4	1.2	1.5	1.2	1.3	1.4	1.4
	M(p-k)	0.8	0.8	0.7	0.7	1.0	0.8	0.5	0.6	0.6	0.6	0.8	0.9	0.5	0.6	0.5	0.5	0.7	0.9
	M(s-bic)	0.1	0.1	0.1	0.1	0.5	0.7	0.1	0.1	0.1	0.1	0.8	1.3	0.1	0.1	0.1	0.1	1.1	1.7
	M(s-0.9)	2.1	2.1	2.0	2.0	2.2	2.2	1.7	1.8	1.8	1.8	2.0	2.4	1.6	1.9	1.6	1.6	1.8	2.3
	M(s-k)	0.9	0.8	0.8	0.8	1.0	0.9	0.6	0.6	0.7	0.6	0.7	0.9	0.5	0.6	0.5	0.5	0.6	0.9

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.20 Power properties of the TSAY tests: H-models

		T=200						T=500						T=1000					
matrix	test	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6
$\mathbf{\Sigma}_1$	M(org)	0.19	0.23	0.09	0.09	0.10	0.32	0.28	0.33	0.08	0.08	0.09	0.42	0.37	0.41	0.07	0.09	0.09	0.52
	M(p-bic)	0.10	0.17	0.09	0.09	0.10	0.21	0.11	0.16	0.09	0.09	0.10	0.20	0.15	0.17	0.08	0.11	0.08	0.19
	M(p-0.9)	0.11	0.17	0.09	0.09	0.10	0.21	0.18	0.25	0.09	0.09	0.11	0.30	0.24	0.27	0.09	0.11	0.09	0.37
	M(p-k)	0.10	0.15	0.10	0.09	0.09	0.20	0.12	0.18	0.09	0.09	0.11	0.22	0.16	0.20	0.09	0.10	0.09	0.24
	M(s-bic)	0.11	0.16	0.09	0.09	0.11	0.19	0.11	0.15	0.09	0.09	0.10	0.20	0.15	0.16	0.08	0.10	0.08	0.19
	M(s-0.9)	0.14	0.20	0.09	0.09	0.10	0.24	0.21	0.26	0.08	0.09	0.10	0.32	0.28	0.31	0.09	0.11	0.10	0.38
	M(s-k)	0.10	0.16	0.09	0.09	0.10	0.17	0.12	0.19	0.09	0.09	0.11	0.20	0.16	0.19	0.09	0.10	0.09	0.24
	X_1	0.11	0.07	0.09	0.09	0.07	0.12	0.18	0.09	0.09	0.11	0.08	0.14	0.19	0.11	0.13	0.11	0.08	0.23
	X_2	0.11	0.09	0.12	0.10	0.09	0.13	0.17	0.10	0.10	0.11	0.10	0.17	0.21	0.10	0.10	0.12	0.11	0.21
$\mathbf{\Sigma}_2$	M(org)	0.22	0.24	0.09	0.11	0.16	0.43	0.34	0.31	0.12	0.13	0.21	0.55	0.44	0.39	0.12	0.13	0.24	0.72
	M(p-bic)	0.17	0.18	0.09	0.11	0.16	0.28	0.19	0.17	0.13	0.13	0.18	0.26	0.20	0.21	0.12	0.15	0.16	0.29
	M(p-0.9)	0.20	0.21	0.09	0.11	0.15	0.32	0.28	0.25	0.13	0.13	0.20	0.39	0.36	0.32	0.12	0.15	0.19	0.50
	M(p-k)	0.19	0.20	0.09	0.12	0.16	0.27	0.25	0.24	0.13	0.13	0.20	0.31	0.34	0.31	0.12	0.15	0.22	0.36
	M(s-bic)	0.17	0.18	0.11	0.11	0.16	0.26	0.19	0.15	0.12	0.13	0.18	0.25	0.20	0.18	0.12	0.15	0.17	0.28
	M(s-0.9)	0.21	0.22	0.09	0.11	0.16	0.32	0.30	0.26	0.13	0.13	0.20	0.42	0.38	0.34	0.13	0.15	0.22	0.58
	M(s-k)	0.20	0.20	0.10	0.12	0.16	0.26	0.28	0.25	0.13	0.13	0.21	0.28	0.36	0.32	0.13	0.15	0.22	0.35
	X_1	0.17	0.07	0.13	0.14	0.07	0.16	0.28	0.10	0.17	0.16	0.09	0.23	0.35	0.12	0.18	0.17	0.09	0.32
	X_2	0.09	0.08	0.15	0.16	0.12	0.15	0.13	0.10	0.19	0.19	0.15	0.21	0.15	0.13	0.19	0.22	0.15	0.27

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.21 Power properties of the ARCH tests: H-models

		T=200						**T=500**						**T=1000**					
matrix	test	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6
$\mathbf{\Sigma}_1$	M(org)	0.74	0.72	0.36	0.34	0.41	0.38	0.99	0.98	0.72	0.69	0.81	0.62	1.00	1.00	0.94	0.96	0.98	0.82
	M(p-bic)	0.72	0.67	0.40	0.40	0.47	0.35	0.98	0.95	0.76	0.74	0.85	0.59	1.00	1.00	0.95	0.98	0.99	0.81
	M(p-0.9)	0.73	0.69	0.40	0.39	0.47	0.34	0.97	0.97	0.76	0.74	0.85	0.58	1.00	1.00	0.95	0.98	0.99	0.81
	M(p-k)	0.72	0.67	0.40	0.40	0.47	0.32	0.97	0.96	0.76	0.74	0.85	0.58	1.00	1.00	0.96	0.98	0.99	0.81
	M(s-bic)	0.72	0.70	0.40	0.40	0.47	0.37	0.98	0.96	0.76	0.74	0.85	0.60	1.00	1.00	0.95	0.98	0.99	0.82
	M(s-0.9)	0.71	0.70	0.39	0.39	0.47	0.36	0.97	0.96	0.76	0.75	0.85	0.60	1.00	1.00	0.96	0.98	0.99	0.82
	M(s-k)	0.70	0.68	0.40	0.40	0.47	0.34	0.97	0.96	0.76	0.75	0.85	0.59	1.00	1.00	0.96	0.98	0.99	0.82
	X_1	0.49	0.16	0.38	0.35	0.11	0.24	0.89	0.28	0.74	0.68	0.21	0.44	0.99	0.42	0.95	0.94	0.32	0.65
	X_2	0.54	0.17	0.38	0.38	0.28	0.25	0.86	0.30	0.76	0.69	0.56	0.44	0.99	0.39	0.94	0.93	0.82	0.63
$\mathbf{\Sigma}_2$	M(org)	0.72	0.72	0.60	0.60	0.87	0.54	0.99	0.98	0.94	0.96	1.00	0.74	1.00	1.00	1.00	1.00	1.00	0.94
	M(p-bic)	0.74	0.72	0.66	0.65	0.91	0.48	0.99	0.98	0.96	0.97	1.00	0.73	1.00	1.00	1.00	1.00	1.00	0.94
	M(p-0.9)	0.76	0.74	0.66	0.65	0.91	0.47	0.99	0.98	0.97	0.97	1.00	0.73	1.00	1.00	1.00	1.00	1.00	0.94
	M(p-k)	0.74	0.73	0.66	0.65	0.91	0.46	0.98	0.98	0.97	0.97	1.00	0.72	1.00	1.00	1.00	1.00	1.00	0.93
	M(s-bic)	0.74	0.73	0.66	0.65	0.91	0.51	0.99	0.98	0.96	0.97	1.00	0.75	1.00	1.00	1.00	1.00	1.00	0.95
	M(s-0.9)	0.76	0.75	0.66	0.65	0.91	0.50	0.99	0.98	0.96	0.97	1.00	0.74	1.00	1.00	1.00	1.00	1.00	0.95
	M(s-k)	0.74	0.73	0.65	0.65	0.91	0.49	0.99	0.98	0.97	0.97	1.00	0.74	1.00	1.00	1.00	1.00	1.00	0.95
	X_1	0.70	0.22	0.66	0.60	0.19	0.33	0.98	0.36	0.95	0.95	0.38	0.52	1.00	0.56	1.00	1.00	0.61	0.74
	X_2	0.32	0.24	0.72	0.72	0.47	0.35	0.64	0.39	0.98	0.97	0.85	0.60	0.90	0.66	1.00	1.00	0.99	0.85

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.22 Number of additional variables/principal components of the MTSAY tests: H-models

		T=200						**T=500**						**T=1000**					
matrix	test	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6
$\boldsymbol{\Sigma}_1$	M(org)	7.6	6.6	4.9	4.6	5.6	8.6	9.8	9.1	4.6	4.8	6.9	10.1	12.1	11.7	5.1	5.3	9.3	12.8
	M(p-bic)	2.0	2.0	2.0	2.0	2.0	2.1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	M(p-0.9)	3.9	3.4	2.5	2.4	2.5	3.9	5.1	4.7	2.4	2.5	2.7	4.9	6.3	6.1	2.6	2.6	3.2	6.4
	M(p-k)	2.8	2.6	2.4	2.4	2.5	3.0	3.1	2.9	2.4	2.4	2.7	3.2	3.4	3.4	2.4	2.5	3.2	3.5
	M(s-bic)	2.0	2.0	2.0	2.0	2.0	2.1	2.0	2.0	2.0	2.0	2.0	2.1	2.0	2.0	2.0	2.0	2.0	2.0
	M(s-0.9)	4.7	4.2	2.8	2.7	2.8	5.0	6.5	6.0	2.7	2.8	3.2	6.3	8.2	8.0	2.9	3.0	4.0	8.3
	M(s-k)	2.9	2.7	2.5	2.4	2.5	3.0	3.2	3.0	2.4	2.5	2.8	3.2	3.5	3.4	2.5	2.6	3.2	3.6
$\boldsymbol{\Sigma}_2$	M(org)	7.1	6.8	5.0	4.4	6.8	8.6	9.7	9.5	4.4	3.8	10.5	10.6	12.8	12.4	4.1	3.9	19.0	13.2
	M(p-bic)	2.0	2.0	2.0	2.0	2.0	2.1	2.0	2.0	2.0	2.0	2.0	2.1	2.0	2.0	2.0	2.0	2.0	2.1
	M(p-0.9)	3.8	3.7	2.3	2.1	2.5	3.9	5.1	5.1	2.2	2.1	2.9	5.1	6.5	6.7	2.1	2.1	4.6	6.4
	M(p-k)	3.3	3.3	2.4	2.3	2.8	3.0	4.3	4.4	2.3	2.1	3.5	3.3	5.3	5.8	2.2	2.2	5.5	3.6
	M(s-bic)	2.0	2.0	2.0	2.0	2.0	2.1	2.0	2.0	2.0	2.0	2.0	2.1	2.0	2.0	2.0	2.0	2.0	2.1
	M(s-0.9)	4.4	4.2	2.5	2.3	2.8	5.1	6.1	6.0	2.3	2.2	3.7	6.7	8.0	8.0	2.2	2.2	5.9	8.5
	M(s-k)	3.6	3.5	2.4	2.3	2.8	3.1	4.9	4.9	2.3	2.1	3.5	3.2	6.4	6.3	2.2	2.2	5.5	3.6

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.23 Number of additional variables/principal components of the MARCH tests: H-models

		T=200						T=500						T=1000					
matrix	test	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6
$\boldsymbol{\Sigma}_1$	M(org)	5.4	4.6	4.2	4.0	3.9	4.4	5.9	5.2	3.9	3.9	3.8	4.2	7.4	5.7	3.9	3.9	3.6	4.6
	M(p-bic)	2.2	2.2	2.0	2.0	2.0	2.2	2.3	2.6	2.0	2.0	2.0	2.2	2.5	3.2	2.0	2.0	2.0	2.3
	M(p-0.9)	4.0	3.3	2.2	2.1	2.1	3.1	4.6	3.9	2.1	2.1	2.1	3.0	5.9	4.4	2.1	2.1	2.0	3.4
	M(p-k)	2.8	2.5	2.2	2.2	2.1	2.5	2.9	2.7	2.1	2.1	2.1	2.4	3.4	2.9	2.1	2.1	2.1	2.5
	M(s-bic)	2.2	2.3	2.0	2.0	2.0	2.2	2.5	2.6	2.0	2.0	2.0	2.2	2.6	3.3	2.0	2.0	2.0	2.2
	M(s-0.9)	4.1	3.4	2.4	2.4	2.1	3.2	4.6	4.0	2.3	2.3	2.1	3.1	5.9	4.4	2.3	2.3	2.0	3.4
	M(s-k)	2.9	2.6	2.2	2.2	2.1	2.5	3.0	2.8	2.1	2.1	2.1	2.4	3.5	3.0	2.1	2.1	2.1	2.5
$\boldsymbol{\Sigma}_2$	M(org)	5.1	5.0	4.1	4.0	3.9	4.4	6.6	5.2	4.1	4.0	3.8	4.8	7.7	6.2	4.2	4.1	3.8	4.8
	M(p-bic)	2.2	2.2	2.0	2.0	2.1	2.2	2.7	2.5	2.0	2.0	2.1	2.4	3.5	3.3	2.0	2.0	2.1	2.5
	M(p-0.9)	3.1	3.0	2.1	2.1	2.1	3.0	4.0	3.2	2.1	2.1	2.1	3.4	4.6	3.7	2.1	2.1	2.1	3.5
	M(p-k)	2.5	2.5	2.2	2.1	2.1	2.5	2.9	2.5	2.2	2.1	2.1	2.5	3.3	2.7	2.1	2.1	2.1	2.5
	M(s-bic)	2.2	2.2	2.0	2.0	2.1	2.2	2.7	2.5	2.0	2.0	2.1	2.4	3.5	3.4	2.0	2.0	2.1	2.5
	M(s-0.9)	3.5	3.4	2.1	2.1	2.1	3.2	4.6	3.6	2.1	2.1	2.1	3.6	5.5	4.3	2.1	2.1	2.1	3.6
	M(s-k)	2.5	2.5	2.2	2.1	2.1	2.4	2.9	2.5	2.2	2.1	2.1	2.5	3.2	2.6	2.1	2.1	2.1	2.4

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.24 Variability of the number of additional variables/principal components of the TSAY tests: H-models

		T=200						T=500						T=1000					
matrix	test	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6
$\boldsymbol{\Sigma}_1$	M(org)	9.5	6.6	6.9	5.7	6.6	11.6	8.9	7.2	5.0	5.4	6.9	10.1	7.2	7.0	5.1	5.2	7.7	10.4
	M(p-bic)	0.2	0.2	0.1	0.1	0.1	0.4	0.1	0.2	0.1	0.0	0.1	0.3	0.1	0.2	0.0	0.0	0.0	0.3
	M(p-0.9)	3.8	2.7	2.0	1.6	1.5	3.7	4.0	3.3	1.5	1.6	1.6	4.0	3.6	3.4	1.5	1.6	1.9	4.6
	M(p-k)	2.1	1.4	1.9	1.6	1.6	2.4	1.7	1.4	1.4	1.6	1.7	2.1	1.4	1.3	1.4	1.5	2.0	2.1
	M(s-bic)	0.3	0.2	0.1	0.1	0.1	0.6	0.1	0.2	0.1	0.0	0.1	0.4	0.1	0.3	0.0	0.0	0.1	0.3
	M(s-0.9)	5.2	3.8	2.9	2.4	2.1	5.7	5.5	4.4	2.2	2.4	2.3	5.8	4.6	4.4	2.3	2.4	2.6	6.0
	M(s-k)	2.5	1.6	2.2	1.8	1.7	2.7	1.9	1.6	1.6	1.7	1.8	2.2	1.5	1.5	1.6	1.7	2.1	2.2
$\boldsymbol{\Sigma}_2$	M(org)	9.1	8.4	7.2	5.3	8.8	11.4	8.9	8.4	6.0	3.6	10.8	10.7	8.6	8.1	4.5	4.3	16.0	10.2
	M(p-bic)	0.2	0.1	0.1	0.1	0.1	0.6	0.3	0.3	0.1	0.1	0.1	0.5	0.4	0.1	0.0	0.1	0.1	0.6
	M(p-0.9)	3.8	3.6	1.3	0.9	1.7	3.7	4.1	3.9	1.1	0.6	2.3	4.4	4.0	4.1	0.8	0.8	3.9	4.2
	M(p-k)	2.6	2.6	1.6	1.1	2.1	2.4	2.8	2.6	1.4	0.8	2.7	2.3	2.5	2.6	1.0	1.0	4.2	2.1
	M(s-bic)	0.2	0.2	0.1	0.1	0.2	0.7	0.4	0.3	0.0	0.1	0.2	0.6	0.3	0.1	0.1	0.1	0.1	0.6
	M(s-0.9)	5.0	4.7	1.9	1.4	2.2	5.7	5.3	5.1	1.6	0.9	2.9	6.2	5.2	5.0	1.2	1.2	4.5	6.0
	M(s-k)	3.2	3.0	1.8	1.3	2.1	2.7	3.4	3.2	1.4	0.8	2.7	2.4	3.2	3.0	1.1	1.0	4.2	2.2

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

[b] The significance level is set to $\alpha = 0.05$.

Table 5.25 Variability of the number of additional variables/principal components of the MARCH tests: H-models

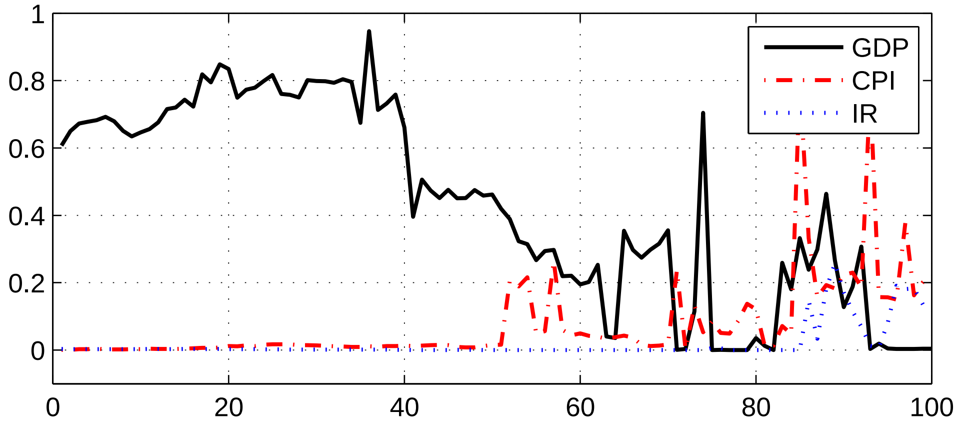
		T=200						T=500						T=1000					
matrix	test	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6
$\mathbf{\Sigma}_1$	M(org)	3.9	3.1	3.1	2.9	2.5	3.3	3.3	3.5	2.4	2.6	2.5	2.9	3.4	3.4	2.4	2.3	2.1	3.3
	M(p-bic)	0.6	0.7	0.2	0.1	0.2	0.8	0.8	1.1	0.1	0.1	0.1	0.6	1.3	1.7	0.2	0.1	0.2	0.8
	M(p-0.9)	3.1	2.5	0.7	0.6	0.4	2.6	2.7	2.9	0.5	0.5	0.4	2.4	2.8	2.9	0.4	0.4	0.3	2.8
	M(p-k)	1.5	1.2	0.7	0.7	0.6	1.3	1.2	1.4	0.6	0.6	0.6	1.2	1.2	1.2	0.5	0.5	0.5	1.3
	M(s-bic)	0.5	0.8	0.1	0.1	0.2	0.8	0.9	1.2	0.1	0.2	0.1	0.7	1.3	1.8	0.2	0.1	0.1	0.8
	M(s-0.9)	3.3	2.6	1.2	1.1	0.4	2.8	2.8	3.1	1.0	1.1	0.4	2.5	2.9	2.9	1.0	0.9	0.3	2.9
	M(s-k)	1.6	1.3	0.7	0.7	0.6	1.3	1.3	1.5	0.6	0.6	0.6	1.1	1.3	1.4	0.5	0.5	0.5	1.2
$\mathbf{\Sigma}_2$	M(org)	3.4	3.6	2.7	2.6	2.6	3.3	3.7	3.5	2.7	2.5	2.5	3.4	3.4	3.5	2.5	2.4	2.4	3.1
	M(p-bic)	0.6	0.8	0.2	0.2	0.3	0.6	1.1	1.3	0.2	0.2	0.4	1.2	1.5	1.8	0.3	0.2	0.5	1.2
	M(p-0.9)	2.0	2.1	0.4	0.4	0.4	2.4	2.3	2.1	0.4	0.3	0.4	2.7	2.2	2.1	0.3	0.3	0.4	2.5
	M(p-k)	1.1	1.1	0.6	0.6	0.6	1.3	1.3	1.1	0.6	0.6	0.6	1.3	1.3	1.1	0.5	0.5	0.6	1.1
	M(s-bic)	0.6	0.7	0.2	0.2	0.3	0.7	1.1	1.3	0.2	0.2	0.4	1.3	1.5	1.9	0.3	0.2	0.5	1.4
	M(s-0.9)	2.4	2.6	0.4	0.4	0.4	2.8	2.8	2.6	0.4	0.4	0.4	3.0	2.7	2.7	0.3	0.3	0.4	2.7
	M(s-k)	1.1	1.2	0.6	0.6	0.6	1.2	1.3	1.1	0.6	0.6	0.6	1.2	1.4	1.1	0.5	0.5	0.6	0.9

[a] M(org) denotes the original multivariate test, M(p-) denotes the multivariate test based on the Pearson correlation matrix, whereas M(s-) on the Spearman correlation matrix. M(-bic) denotes the multivariate test with the number of principal components selected by the BIC approach, whereas M(-0.9) by the variance rule with the cut-off variance 0.9, and M(-k) by the Kaiser rule.

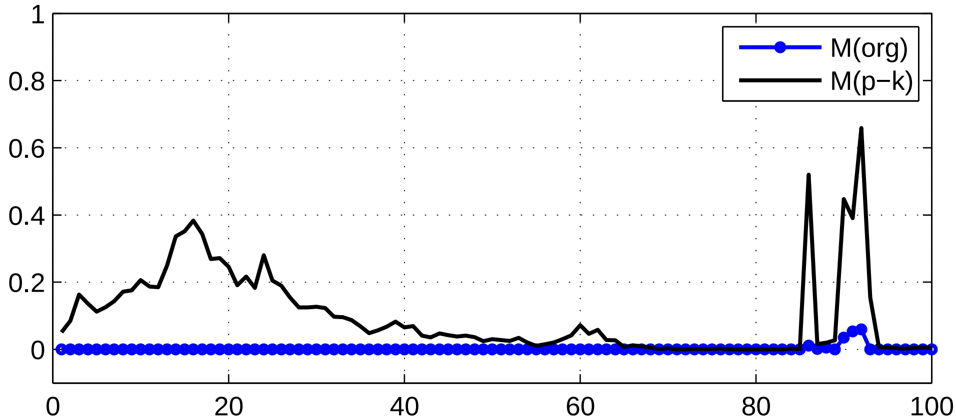
[b] The significance level is set to $\alpha = 0.05$.

5.8 Appendix B: Figures

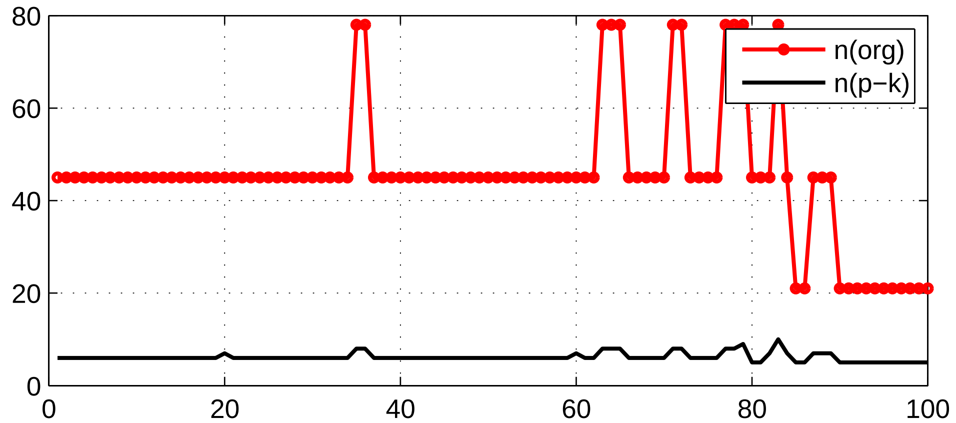
Figure 5.6 Empirical results of the TSAY tests: 25 year rolling windows



(a) p-values of the univariate tests



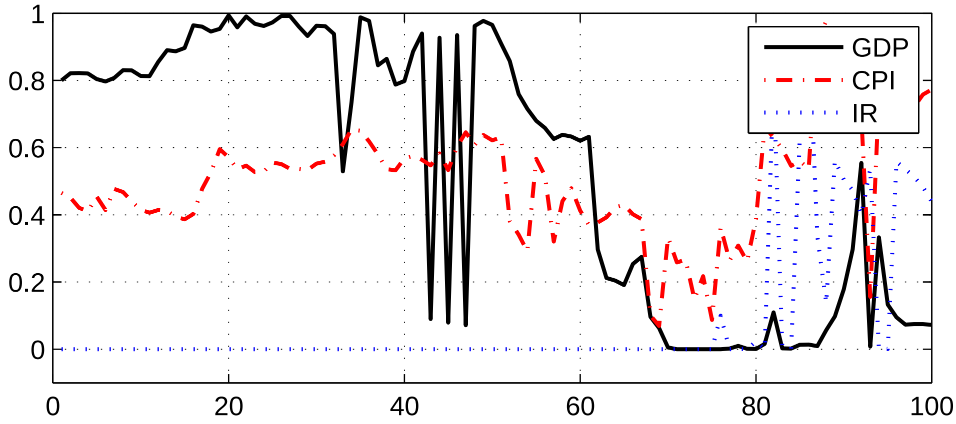
(b) p-values of the multivariate tests



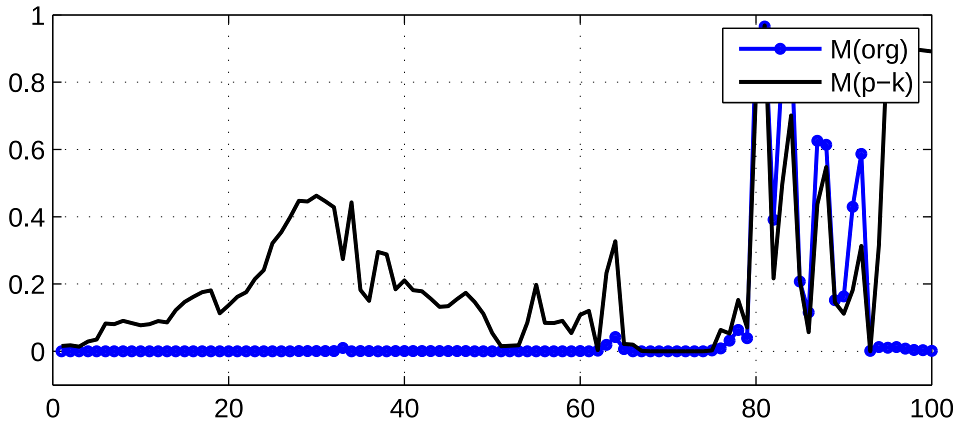
(c) number of additional variables/number of components of the multivariate tests

Note: M(org) denotes the average rejection frequency of the original multivariate TSAY test, M(p-k) stands for the multivariate TSAY test based on the automatically selected number of principal components. n(org) and n(p-k) denote a number of additional variables and principal components used by the multivariate tests.

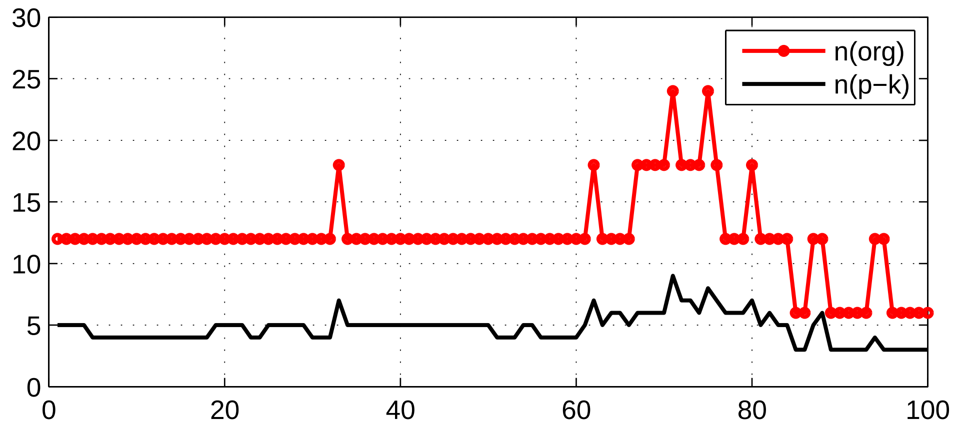
Figure 5.7 Empirical results of the ARCH tests: 25 year rolling windows



(a) p-values of the univariate tests



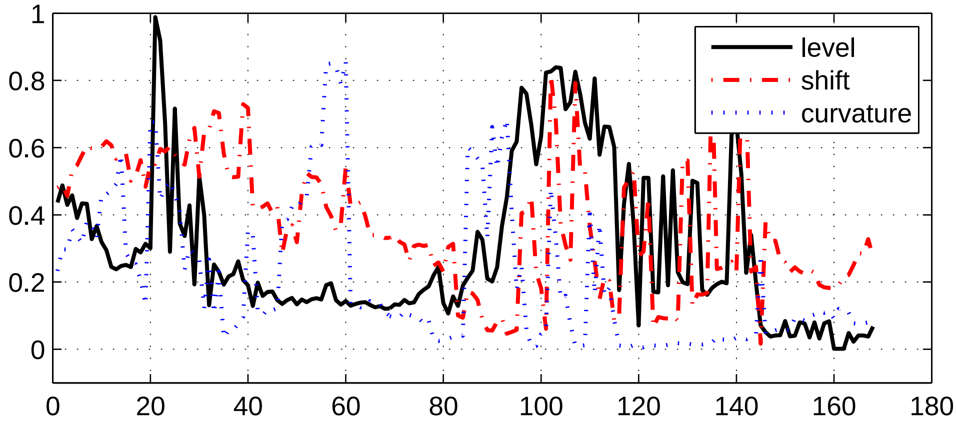
(b) p-values of the multivariate tests



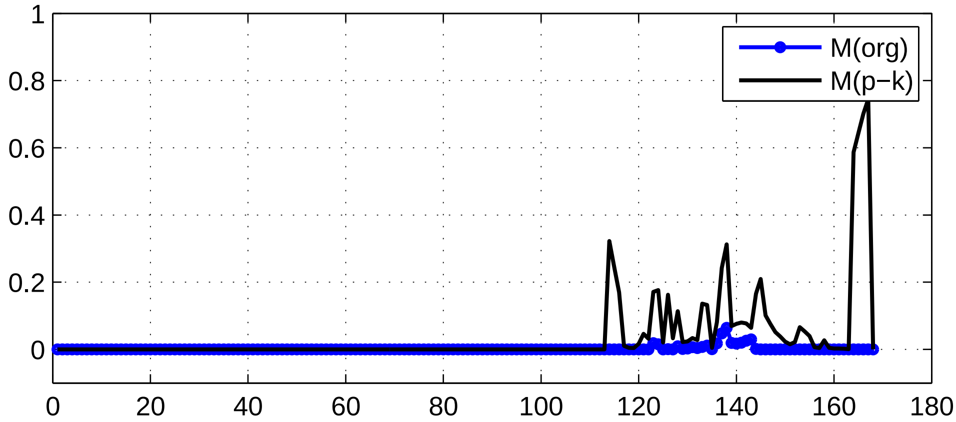
(c) number of additional variables/number of components of the multivariate tests

Note: M(org) denotes the average rejection frequency of the original multivariate ARCH test, M(p-k) stands for the multivariate ARCH test based on the automatically selected number of principal components. n(org) and n(p-k) denote a number of additional variables and principal components used by the multivariate tests.

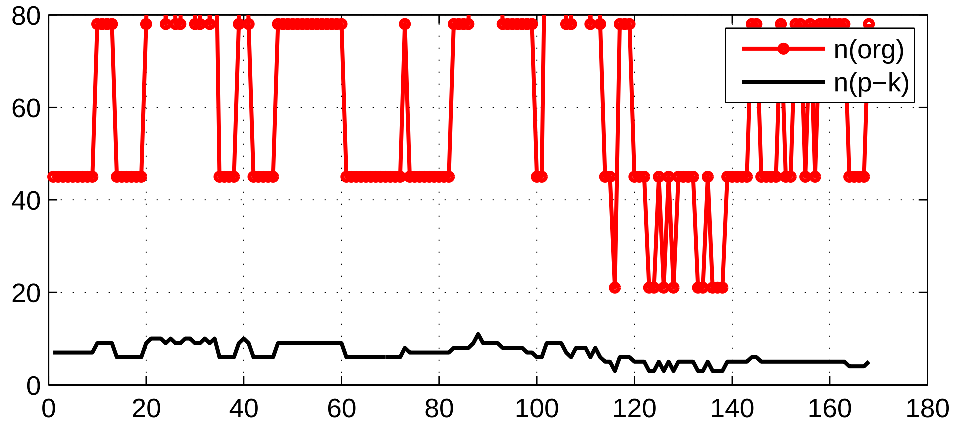
Figure 5.8 Empirical results of the TSAY tests: 25 year rolling windows



(a) p-values of the univariate tests



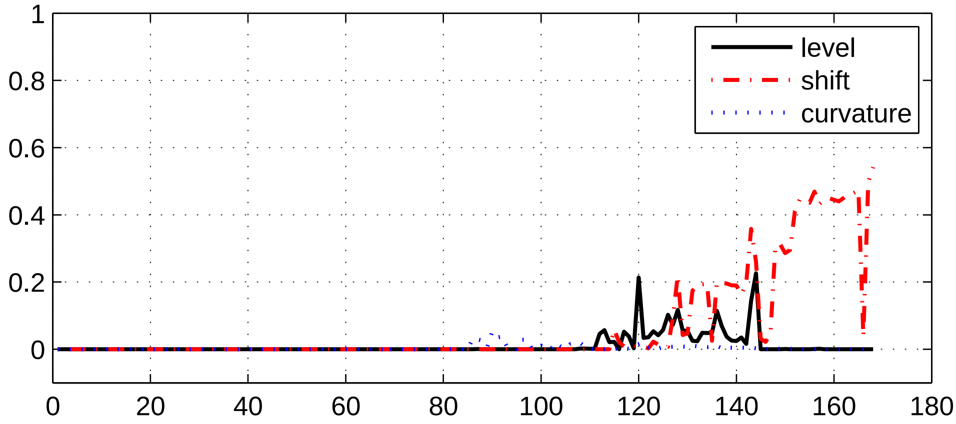
(b) p-values of the multivariate tests



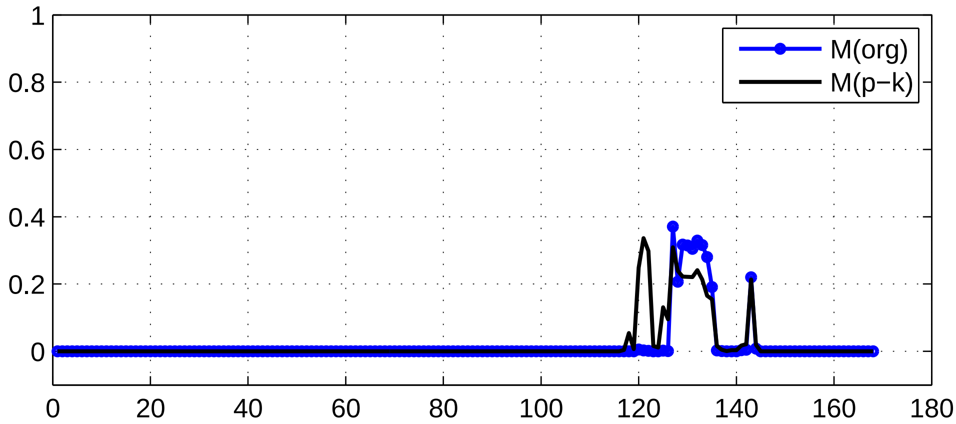
(c) number of additional variables/number of components of the multivariate tests

Note: M(org) denotes the average rejection frequency of the original multivariate TSAY test, M(p-k) stands for the multivariate TSAY test based on the automatically selected number of principal components. n(org) and n(p-k) denote a number of additional variables and principal components used by the multivariate tests.

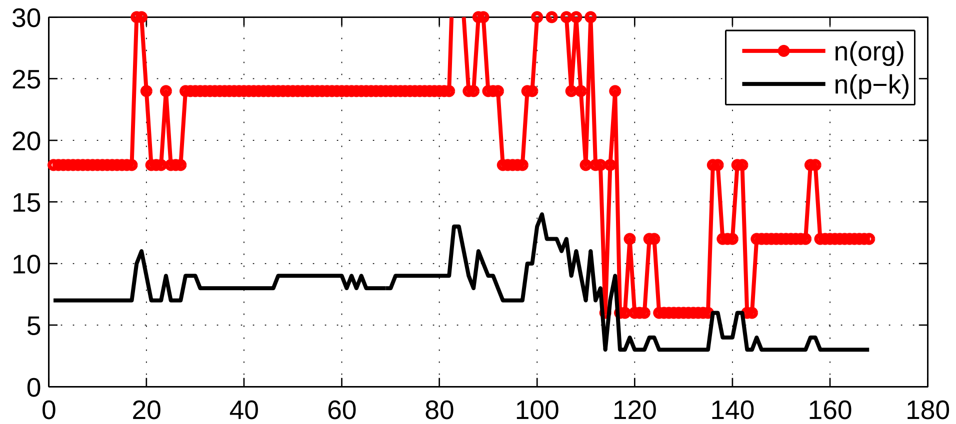
Figure 5.9 Empirical results of the ARCH tests: 25 year rolling windows



(a) p-values of the univariate tests



(b) p-values of the multivariate tests



(c) number of additional variables/number of components of the multivariate tests

Note: M(org) denotes the average rejection frequency of the original multivariate ARCH test, M(p-k) stands for the multivariate ARCH test based on the automatically selected number of principal components. n(org) and n(p-k) denote a number of additional variables and principal components used by the multivariate tests.

Chapter 6

Conclusion and Further Research

6.1 Conclusion

The second chapter examines the statistical properties of the selected univariate non-linearity tests under different conditions such as moment condition failure, asymmetry of innovations, and various parameter configurations of data generating processes. Since analytical results are available only for a very limited number of the test statistics, an extensive Monte Carlo approach is implemented instead. Our goal has been to provide evidence whether or not there exists a superior test, which should be preferred in applied research. The Monte Carlo results reveal several new findings. First, our results suggest that one should interpret the results about non-linearity testing with caution, since the power of the tests is very sensitive on the parameter configuration of data generating processes. In particular, the power of the non-linearity tests is robust (i.e. rejecting or not rejecting linearity) against DGP parameters only in less than 50 % of cases. Second, the power of the tests is statistically significantly inflated under asymmetry of innovations and moment condition failure. Third, based on a multi-criterion evaluation, there is no superior test among those considered in this thesis. Nevertheless, the BDS and NN tests perform very well in general. Two important conclusion emerge from our results for applied time series modelling. First, the statistical properties of the selected non-linear tests based on Gaussian innovations can be considered as conservative as compared to other configurations of model innovations. For this reason, the use of Gaussian innovations in the context of non-linearity testing might be recommended. Second, it is absolutely crucial to examine the whole parameter space of a given stochastic process when evaluating the finite sample performance of the non-linearity tests. Relying just on a few parameter configurations may easily lead to misleading inference.

The third chapter presents theoretical, Monte carlo and empirical results of a new version of the portmanteau Q test based on autocorrelations and crosscorrelations. There have been two good reasons for considering a new version of the Q test. First, the Q test is a very popular test statistic routinely used in the time series literature, the test is easy to calculate and follows a standard limiting distribution. Second, according to the results presented in Chapter 2, there is still a large room for possible improvements of the Q test. The main task of this chapter has been to concentrate on the power improvements of the Q test in order to bypass some of the shortcomings discovered in Chapter 2. Our results suggest the following. First, it is demonstrated that inspecting residual autocorrelations and crosscorrelations can be an useful, yet very simple, tool for testing linearity. The new Q test significantly improves the power against some non-linear time series models (e.g. a threshold autoregressive model or a threshold moving average model). In addition, the new Q test is capable to detect some interesting non-linear processes(e.g. a non-linear moving average model), for which the standard Mcleod and Li Q test completely fails. Second, the power of the new Q tests is even higher as compared to the BDS and NN tests, two recommended tests from Chapter 2. As a result, it can be concluded that there seems to be no point in applying the whole battery of the non-linearity tests and the use of the new Q tests appears to be fully sufficient.

The fourth chapter concentrates on modifying a quantile-based test for testing symmetry of the marginal of weakly dependent stochastic processes. It has been shown that the test is intuitive, easy to calculate, follows standard limiting distribution, and much more importantly, it is robust against weak dependence of observations. Especially the last feature makes the test very attractive for applied research since it reduces the inferential errors coming from the incorrect estimation of the key quantities of the test. Monte Carlo results suggest that the finite sample properties of the QS test significantly outperforms the skewness-based symmetry test and compares favorably with the bootstrap-based Kolmogorov-Smirnov test. It can be concluded that in situations where the user is not proficient in bootstrap, or it is not clear which bootstrap method should be implemented, the QS test may serve as a valuable alternative.

The fifth chapter focuses on proposing two new principal component-based multivariate non-linearity tests. It is shown that testing non-linearity of economic indicators, which are to some extent dependent, using univariate test statistics can lead to misleading inference. However, standard multivariate non-linearity tests suffer from a dimensionality problem. The main goal of this chapter has been to modify two well known multivariate test statistics for their use in relatively small samples usually observed in applied macroeconomics. Our results, based on extensive Monte Carlo experiments, suggest the following. First, it is shown that the use of a principal component analysis can bypass the dimensionality problem very efficiently without any systematic power distortion. According to our results, the BIC rule performs best as compared to other stopping rules considered in this chapter. Second, the power of the multivariate tests is significantly higher as compared to univariate test statistics. Third, the principal component-based tests are insensitive to a choice of the correlation matrix used to calculate principal components. It can be concluded from the results that a principal component analysis can serve as an useful, yet very simple, tool in the context of non-linearity testing.

6.2 Further Research

The submitted Ph.D. thesis is concerned with various issues of testing for non-linearity and marginal asymmetry. Some issues, however, have been left for further research. Among those, an issue of discrimination among non-linear models plays a key role.

In the past two decades, main attention has been paid to developing many different non-linear models and related test statistics. Empirical evidence, partly supported by the results presented in Chapters 3 and 5 of this thesis, clearly indicates that not rejecting linearity is an exception rather than a rule in practice. In such a case, a natural question arises: "*What is the right alternative model after rejecting the null hypothesis of linearity?*" However, discrimination among non-nested models is much more involved, especially when competing models are non-linear. This issue has been almost completely overlooked in the literature despite its importance in the whole process of non-linear time series modelling. An exception is the influential work of Chen et al. (1997) who offer an interesting solution to this problem. The authors proposed a unified approach that can be used to select an appropriate time series model from a given subset of potentially non-nested non-linear candidates for each observation of a given stochastic process. Although the reported results are encouraging, some problems still remain open. For instance, the main disadvantage of this approach is that the user is supposed to select: (i) A subset of potential candidates for a given process (e.g. a TAR versus MSAR model or a TAR versus Structural Break model); (ii) Extremely strong priors. However, as shown by Koop and Potter (2001) the correct specification of a subset of the right competing models and/or priors might not be an easy task even for experts proficient in a time series analysis. The authors demonstrate, by means of Monte Carlo experiments, that a model with time-varying parameters can be mistakenly interchanged with a TAR model. Similar results, but for structural break models, are obtained by Carrasco (2002); (iii) In addition to that, the current Bayesian procedure is computationally expensive even for relatively simple univariate processes. No similar procedure is known, at least to the best of our knowledge, for multivariate time series.

For this reason, we hold the view that further research should be devoted to developing a computationally efficient Bayesian procedure for discriminating among the main classes of univariate and multivariate non-linear models based on reasonable ("flat") priors.

Bibliography

Acemoglu, D. and A. Scott (1997). Asymmetric business cycles: Theory and time-series evidence. *Journal of Monetary Economics 40*, 501–533.

Adler, R., R. Feldman, and C. Gallagher (1998). Analysing stable time series. *In: A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, 133–58.

Adolfson, M., S. Laséen, J. Lindé, and M. Villani (2007). Bayesian estimation of an open economy DSGE model with incomplete pass-through. *Journal of International Economics 72*, 481–511.

Adolfson, M., S. Laséen, J. Lindé, and M. Villani (2008). Empirical Properties Of Closed-And Open-Economy Dsge Models Of The Euro Area. *Macroeconomic Dynamics 12*, 2–19.

Adolfson, M., S. Laséen, J. Lindé, and M. Villani (2008). Evaluating an estimated new keynesian small open economy model. *Journal of Economic Dynamics and Control 32*, 2690–2721.

Amendola, A., M. Niglio, and C. Vitale (2006). The moments of SETARMA models. *Statistics and Probability Letters 76*, 625–633.

Anderson, H. (1997). Transaction costs and nonlinear adjustment towards equilibrium in the US Treasury Bill market. *Oxford Bulletin of Economics and Statistics, 59*, 465–484.

Anderson, T. (1991). The asymptotic distributions of autoregressive coefficients. Technical report, Stanford University.

Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis.* Wiley.

Anderson, T. and A. Walker (1964). On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process. *The Annals of Mathematical Statistics 35*, 1296–1303.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Ang, A. and G. Bekaert (2002). Short rate nonlinearities and regime switches. *Journal of Economic Dynamics and Control 26*, 1243–1274.

Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business and Economic Statistics 23*, 49–60.

Bali, T. (2003). An extreme value approach to estimating volatility and value at risk. *The Journal of Business 76*, 83–108.

Balke, N. and T. Fomby (1994). Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series. *Journal of Applied Econometrics 9*, 181–200.

Bauwens, L., S. Laurent, and J. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics 21*, 79–109.

Beirlant, J. (2004). *Statistics of Extremes: theory and applications.* Wiley.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy 81*, 637–654.

Blake, A. and G. Kapetanios (2003). A radial basis function artificial neural network test for neglected nonlinearity. *Econometrics Journal 6*, 357–373.

Boldin, M. (1999). Should policy makers worry about asymmetries in the business cycle? *Studies in Nonlinear Dynamics and Econometrics 3*, 203–220.

Bollerslev, T. (1990). Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics 72*, 498–505.

Bowman, K. and L. Shenton (1975). Tables of moments of the skewness and kurtosis statistics in non-normal sampling. *Union Carbide Nuclear Divison Report*.

Box, G., G. Jenkins, and G. Reinsel (2008). *Time Series Analysis: forecasting and control*. Wiley.

Box, G. and D. Pierce (1970). Distributions of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Assciation 65*, 1509–1525.

Brock, W., D. Hsieh, and B. LeBaron (1991). *Nonlinear Dynamics, Chaos, and Instability: statistical theory and economic evidence*. MIT Press.

Brock, W., J. Scheinkman, W. Dechert, and B. LeBaron (1996). A test for independence based on the correlation dimension. *Econometric Reviews 15*, 197–235.

Brockwell, P. and R. Davis (1991). *Time Series: theory and methods*. Springer.

Burns, A. and W. Mitchell (1946). *Measuring Business Cycles*. National Bureau of Economic Research.

Cai, Z. and G. G. Roussas (1992). Uniform strong estimation under α-mixing, with rates. *Statistics & probability letters 15*, 47–55.

Carrasco, M. (2002). Misspecified structural change, threshold, and markov-switching models. *Journal of econometrics 109*, 239–273.

Casella, G. and R. Berger (2001). *Statistical Inference*. Duxbury Press.

Chanda, K. (1974). Strong mixing properties of linear stochastic processes. *Journal of Applied Probability 11*, 401–408.

Chen, C., R. McCulloch, and R. Tsay (1997). A unified approach to estimating and modeling linear and nonlinear time series. *Statistica Sinica 7*, 451–472.

Chen, R., J. Liu, and R. Tsay (1995). Additivity tests for nonlinear autoregression. *Biometrika 82*, 369–383.

Cho, S. and A. Moreno (2003). A structural estimation and interpretation of the New Keynesian macro model. *Columbia University Working Paper 14*.

Christensen, J., F. Diebold, and G. Rudebusch (2011). The affine arbitrage-free class of Nelson-Siegel term structure models. *Journal of Econometrics 164*, 4–20.

Clements, M. and H. Krolzig (2003). Business Cycle Asymmetries. *Journal of Business and Economic Statistics 21*, 196–211.

Conine, T. and M. Tamarkin (1981). On diversification given asymmetry in returns. *Journal of Finance*, 1143–1155.

Corrado, C. and T. Su (1996). Skewness and kurtosis in S&P 500 index returns implied by option prices. *Journal of Financial research 19*, 175–192.

Croux, C. and G. Haesbroeck (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika 87*, 603–618.

Davidson, R. and J. G. MacKinnon (1999). *Econometric Theory and Methods*. Oxford University Press.

Davig, T. (2007). Phillips Curve Instability and Optimal Monetary Policy. *Federal Reserve Bank of Kansas City Working Paper Series 4*.

Davis, R. and T. Mikosch (2000). The sample autocorrelations of financial time series models. *In: Nonlinear and nonstationary signal processing*, 247–274.

Davis, R. and S. Resnick (1986). Limit theory for the sample covariance and correlation functions of moving averages. *The Annals of Statistics 14*, 533–558.

De Jong, R. (2000). A strong consistency proof for heteroskedasticity and autocorrelation consistent covariance matrix estimators. *Econometric Theory 16*, 262–268.

de Lima, P. (1997). On the robustness of nonlinearity tests to moment condition failure. *Journal of Econometrics 76*, 251–280.

Dehling, H. and W. Philipp (2002). Empirical process techniques for dependent data. In H. Dehling, T. Mikosch, and M. Sørensen (Eds.), *Empirical Process Techniques for Dependent Data*. Springer.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007). On the fit of new keynesian models. *Journal of Business & Economic Statistics 25*, 123–143.

Deschamps, P. (1993). Joint tests for regularity and autocorrelation in allocation systems. *Journal of Applied Econometrics 8*, 195–211.

Deschamps, P. (1996). Monte Carlo methodology for LM and LR autocorrelation tests in multivariate regression. *Annales d'Economie et de Statistique 43*, 149–169.

Diebold, F., G. Rudebusch, and S. Borag?an Aruoba (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of econometrics 131*, 309–338.

Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics 130*, 337–364.

Ding, Z., C. Granger, and R. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance 1*, 83–106.

Doukhan, P. (1994). *Mixing: properties and examples*. Springer.

Duchesne, P. and S. Lalancette (2003). On testing for multivariate arch effects in vector time series models. *The Canadian Journal of Statistics 31*, 275–292.

Embrechts, P., C. Klüppelberg, and T. Mikosch (2011). *Modelling Extremal Events*. Springer.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007.

Engle, R. and F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*, 122–150.

Escanciano, J. and I. Lobato (2009). An automatic Portmanteau test for serial correlation. *Journal of Econometrics 151*, 140–149.

Escribano, Á. and O. Jordá (2001). Testing nonlinearity: Decision rules for selecting between logistic and exponential STAR models. *Spanish Economic Review 3*, 193–209.

Francq, C. and J. Zakoïan (2006). Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process. *Econometric Theory 22*, 815–834.

French, M. and D. Sichel (1993). Cyclical patterns in the variance of economic activity. *Journal of Business & Economic Statistics 11*, 113–119.

Fuller, W. (1996). *Introduction to Statistical Time Series*. Wiley.

Godfrey, L. G. (1988). *Misspecification tests in econometrics*. Cambridge University Press.

Granger, C. and A. Andersen (1978). *An Introduction to Bilinear Time Series Models*. Gottingen.

Granger, C. W. J. and T. Teräsvirta (1993). *Modelling Nonlinear Economic Time Series*. Oxford University Press.

Groeneveld, R. and G. Meeden (1984). Measuring skewness and kurtosis. *The Statistician 33*, 391–399.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2007). The us treasury yield curve: 1961 to the present. *Journal of Monetary Economics 54*, 2291–2304.

Hacker, R. and A. Hatemi-J (2005). A test for multivariate arch effects. *Applied Economics Letters 12*, 411–417.

Hall, P., S. Lahiri, and Y. Truong (1995). On bandwidth choice for density estimation with dependent data. *The Annals of Statistics 23*, 2241–2263.

Hansen, B. (1992). Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica 60*, 967–972.

Hansen, B. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica 64*, 413–430.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory 24*, 726.

Harvill, J. and B. Ray (1999). Testing for nonlinearity in a vector time series. *Biometrika 89*, 728–834.

Hinich, M. and D. Patterson (1985). Evidence of nonlinearity in daily stock returns. *Journal of Business & Economic Statistics 3*, 69–77.

Hinkley, D. (1975). On power transformations to symmetry. *Biometrika 62*, 101–111.

Horowitz, J., I. Lobato, J. Nankervis, and N. Savin (2006). Bootstrapping the Box-Pierce Q test: A robust test of uncorrelatedness. *Journal of Econometrics 133*, 841–862.

Hsieh, D. (1989). Testing for nonlinear dependence in daily foreign exchange rates. *Journal of Business 62*, 339–368.

Hubert, M., P. Rousseeuw, and K. Branden (2005). Robpca: a new approach to robust principal component analysis. *Technometrics 47*, 64–79.

Ibragimov, I. and Y. Linnik (1971). Independent and stationary sequences of random variables.

Izenman, A. (2008). *Modern Multivariate Statistical Techniques*. Springer.

Jackson, J. (1991). *A User's Guide to Principal Components*. Wiley.

Jansen, D. and C. D. Vries (1991). On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics 73*, 18–24.

Jarrow, R. and A. Rudd (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics 10*, 347–369.

Jiang, J. (2010). *Large Sample Techniques for Statistics*. Springer.

Johnson, N., S. Kotz, and N. Balakrishnan (1994). *Continuous Univariate Distributions*. Wiley.

Johnson, R. and D. Wichern (2007). *Applied Multivariate Statistical Analysis*. Prentice Hall.

Jolliffe, I. (2005). *Principal Component Analysis*. Wiley.

Jones, M., J. Marron, and S. Sheather (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association 91*, 401–407.

Jorion, P. (2007). *Value at risk*. McGraw-Hill.

Kearns, P. and A. Pagan (1997). Estimating the density tail index for financial time series. *Review of Economics and Statistics 79*, 171–175.

Keenan, D. (1985). A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika 72*, 39–44.

Kendall, M. and J. Ord (1973). *Time Series*. Griffin London.

Koedijk, K., M. Schafgans, and C. de Vries (1990). The tail index of exchange rate returns. *Journal of International Economics 29*, 93–108.

Konishi, S. (1978). An approximation to the distribution of the sample correlation coefficient. *Biometrika 65*, 654–656.

Koop, G. and S. Potter (2001). Are apparent findings of nonlinearity due to structural instability in economic time series? *The Econometrics Journal 4*, 37–55.

Kozicki, S. and P. Tinsley (2002). Term Premia: Endogenous Constraints on Monetary Policy. *FRB of Kansas City Research Working Paper*.

Kwan, A. and A. Sim (1996a). On the finite-sample distribution of modified portmanteau tests for randomness of a Gaussian time series. *Biometrika 83*, 938–943.

Kwan, A. and A. Sim (1996b). Portmanteau tests of randomness and Jenkins' variance-stabilizing transformation. *Economics Letters 50*, 41–49.

Lahiri, S. and S. Sun (2009). A Berry–Esseen theorem for sample quantiles under weak dependence. *The Annals of Applied Probability 19*, 108–126.

Lawrance, A. and P. Lewis (1985). Modelling and residual analysis of nonlinear autoregressive time series in exponential variables. *Journal of the Royal Statistical Society 47*, 165–202.

Lawrance, A. and P. Lewis (1987). Higher-order residual analysis for nonlinear time series with autoregressive correlation structures. *International Statistical Review 55*, 21–35.

Lee, J. and M. King (1993). A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *Journal of Business & Economic Statistics 11*, 17–27.

Lee, T., H. White, and C. Granger (1993). Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics 56*, 269–290.

Lehmann, E. (1999). *Elements of Large-sample Theory*. Springer.

Li, W. (1992). On the asymptotic standard errors of residual autocorrelations in nonlinear time series modelling. *Biometrika 79*, 435–437.

Li, W. and T. Mak (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis 15*, 627–636.

Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and their Applications 65*, 69–80.

Liebscher, E. (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *Journal of Time Series Analysis 26*, 669–689.

Lin, J. and A. McLeod (2008). Portmanteau tests for arma models with infinite variance. *Journal of Time Series Analysis 29*, 600–617.

Ling, S. and W. Li (1997). Diagnostic checking of non-linear multivariate time series with multivariate arch errors. *Journal of Time Series Analysis 18*, 447–464.

Liu, Z., D. Waggoner, and T. Zha (2009). Asymmetric expectation effects of regime shifts in monetary policy. *Review of Economic Dynamics 12*, 284–303.

Ljung, G. and G. Box (1978). On a measure of fit in time series models. *Biometrika 65*, 297–303.

Lobato, I. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association 96*, 1066–1076.

Lobato, I., J. Nankervis, and N. Savin (2002). Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory 18*, 730–743.

Loretan, M. and P. Phillips (1994). Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance 1*, 211–248.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988). Testing linearity in univariate time series models. *Scandinavian Journal of Statistics 15*, 161–175.

Ma, Y. and M. Genton (2001). Highly robust estimation of dispersion matrices. *Journal of Multivariate Analysis 78*, 11–36.

Maravall, A. (1983). An application of nonlinear time series forecasting. *Journal of Business & Economic Statistics 1*, 66–74.

Markowitz, H. (1968). *Portfolio selection: Efficient diversification of investments*. Yale University Press.

McCulloch, J. (1997). Measuring tail thickness to estimate the stable index α: A critique. *Journal of Business & Economic Statistics 15*, 74–81.

McLeod, A. and W. Li (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis 4*, 269–273.

McQueen, G. and S. Thorley (1993). Asymmetric business cycle turning points. *Journal of Monetary Economics 31*, 341–362.

Monti, A. (1994). A proposal for a residual autocorrelation test in linear models. *Biometrika 81*, 776–780.

Neftci, S. (1984). Are economic time series asymmetric over the business cycle? *The Journal of Political Economy 92*, 307–328.

Newey, W. and K. West (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies 61*, 631–653.

Ng, S. and P. Perron (2005). A note on the selection of time series models. *Oxford Bulletin of Economics and Statistics 67*, 115–134.

Peel, D. and A. Speight (1994). Testing for non-linear dependence in inter-war exchange rates. *Weltwirtschaftliches Archiv 130*, 391–417.

Peel, D. and A. Speight (1998a). Modelling business cycle nonlinearity in conditional mean and conditional variance: some international and sectoral evidence. *Economica 65*, 211–229.

Peel, D. and A. Speight (1998b). Threshold nonlinearities in output: some international evidence. *Applied Economics 30*, 323–333.

Peiró, A. (1999). Skewness in financial returns. *Journal of Banking & Finance 23*, 847–862.

Peres-Neto, P., D. Jackson, and K. Somers (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis 49*, 974–997.

Pérez, A. and E. Ruiz (2003). Properties of the sample autocorrelations of nonlinear transformations in long-memory stochastic volatility models. *Journal of Financial Econometrics 1*, 420–444.

Petruccelli, J. and S. Woolford (1984). A threshold AR (1) model. *Journal of Applied Probability 21*, 270–286.

Prakash, A., C. Chang, and T. Pactwa (2003). Selecting a portfolio with skewness: Recent evidence from us, european, and latin american equity markets. *Journal of Banking & Finance 27*, 1375–1390.

Premaratne, G. and A. Bera (2005). A test for symmetry with leptokurtic financial data. *Journal of Financial Econometrics 3*, 169.

Psaradakis, Z. (2003). A bootstrap test for symmetry of dependent data based on a kolmogorov–smirnov type statistic. *Communications in Statistics 32*, 113–126.

Psaradakis, Z. and M. Sola (2003). On detrending and cyclical asymmetry. *Journal of Applied Econometrics 18*, 271–289.

Psaradakis, Z. and N. Spagnolo (2002). Power properties of nonlinearity tests for time series with Markov regime. *Studies in Nonlinear Dynamics and Econometrics 6*, 1–14.

Ramsey, J. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society 31*, 350–371.

Randles, R., M. Fligner, G. Policello, and D. Wolfe (1980). An asymptotically distribution-free test for symmetry versus asymmetry. *Journal of the American Statistical Association 75*, 168–172.

Rao, C. (1973). *Linear Statistical Inference.* Wiley.

Romano, J. and L. Thombs (1996). Inference for Autocorrelations under Weak Assumptions. *Journal of the American Statistical Association 91*, 590–600.

Rudebusch, G. and E. Swanson (2008). Examining the bond premium puzzle with a dsge model. *Journal of Monetary Economics 55*, 111–126.

Runde, R. (1997). The asymptotic null distribution of the Box-Pierce Q-statistic for random variables with infinite variance an application to German stock returns. *Journal of Econometrics 78*, 205–216.

Sarno, L., M. Taylor, and I. Chowdhury (2004). Nonlinear dynamics in deviations from the law of one price: a broad-based empirical study. *Journal of International Money and Finance 23*, 1–25.

Schmitt-Grohe, S. and M. Uribe (2004). Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control 28*, 755–775.

Schott, J. (2005). *Matrix Analysis for Statistics*. Wiley.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*, 461–464.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.

Sharipov, O. and M. Wendler (2013). Normal limits, nonnormal limits, and the bootstrap for quantiles of dependent data. *Statistics & Probability Letters 83*, 1028–1035.

Sichel, D. (1993). Business cycle asymmetry: a deeper look. *Economic Inquiry 31*, 224–224.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC.

Sims, C. and T. Zha (2006). Were there regime switches in us monetary policy? *The American Economic Review 96*, 54–81.

Smets, F. and R. Wouters (2002). Openness, imperfect exchange rate pass-through and monetary policy. *Journal of Monetary Economics 49*, 947–981.

Smets, F. and R. Wouters (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association 1*, 1123–1175.

Smets, F. and R. Wouters (2007). Shocks and Frictions in US Business Cycles: a Bayesian DSGE Approach. *American Economic Review 97*, 586–606.

Speight, A. and D. McMillan (1998). Testing for asymmetries in uk macroeconomic time series. *Scottish Journal of Political Economy 45*, 158–170.

Sun, S. and S. Lahiri (2006). Bootstrapping the sample quantile of a weakly dependent sequence. *Sankhyā: The Indian Journal of Statistics 68*, 130–166.

Teräsvirta, T. and H. Anderson (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics 7*, 119–136.

Tong, H. (1990). *Non-linear Time Series: a dynamical system approach.* Oxford University Press.

Tsay, R. (1986). Nonlinearity tests for time series. *Biometrika 73*, 461–466.

Uhlig, H. (1995). A toolkit for analyzing nonlinear dynamic stochastic models easily. *Federal Reserve Bank of Mineapolis, Discussion Paper 101.*

Vahamaa, S. (2003). Skewness and kurtosis adjusted black-scholes model: a note on hedging performance. *Finance Letters 1*, 6–12.

van Dijk, D. and P. Franses (1999). Testing for smooth transition nonlinearity in the presence of outliers. *Journal of Business & Economic Statistics 17*, 217–235.

Wei, W. (1990). *Time Series Analysis: univariate and multivariate methods.* Addison-Wesley.

White, H. (1987). Specification testing in dynamic models. *Advances in Econometrics 1*, 1–58.

White, H. (1989). An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. *Proceedings of the International Joint Conference on Neural Networks 2*, 451–5.

White, H. (2001). *Asymptotic Theory for Econometricians.* Academic Press.

Woodford, M. (2011). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.

Yang, M. and R. Bewley (1995). Moving average conditional heteroskedastic processes. *Economics Letters 49*, 367–372.

Yao, Q. and P. Brockwell (2006). Gaussian maximum likelihood estimation for ARMA models. *Journal of Time Series Analysis 27*, 857–875.