Birkbeck Working Papers in Economics & Finance

ISSN 1745-8587

Birkbeck
UNIVERSITY OF LONDON

School of Economics, Mathematics and Statistics

BWPEF 1205

Robustness of Power Properties of Non-linearity Tests

Marian Vavra
Birkbeck, University of London

March 2012

▪ Birkbeck, University of London ▪ Malet Street ▪ London ▪ WC1E 7HX ▪

Robustness of Power Properties of Non-linearity Tests

Marian Vavra *†
Department of Economics, University of London, UK

March 26, 2012

JEL classification: C15, C22.
Key words: non-linearity testing, Monte Carlo experiments.

Abstract

The paper examines the robustness of the size and power properties of the standard non-linearity tests under different conditions such as moment failure and asymmetry of innovations. Our results reveal the following. First, there seems not to be a direct link between moment condition failure and the power variation of non-linearity tests. Second, the power of the tests is very sensitive to asymmetry of innovations compared to moment condition failure. Third, although we evaluate 9 non-linear time series models using 8 standard non-linearity tests, some non-linear models remain completely undetected.

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

*Correspondence to: Marian Vavra, Department of Economics, University of London, Malet Street, WC1E 7HX, London, UK. E-mail: marian.vavra@gmail.com

†This paper is a part of my Ph.D. thesis. I would like to thank Professor Zacharias Psaradakis and Professor Ron Smith for very useful comments and suggestions leading to significant improvements of the paper. All remaining errors are only mine.

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

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

where $I(\cdot)$ is a standard indicator function taking 1 if $|Y_{t-1}| \leq 1$ 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, we present three realizations of a simple threshold autoregressive (TAR) model with different parameters, see Figure 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 paper examines the robustness of standard non-linearity tests against a parameter configuration.

Second, another problem is that the parameter specification in many research papers do 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

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

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 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 about 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 process adequately. See Figure 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 $\kappa \in (1, 2)$. 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 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 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

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

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 (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 (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), the situation is even more peculiar for non-linear time series models with an infinite variance. Authors show that the rate of convergence of the sample autocorrelations of some non-linear models (e.g. a BL and ARCH model) 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 autocorrelation functions of some non-linear models (e.g. a stochastic volatility model) 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.[1] For this reason, the second part of this paper 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 necessary 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 in date generating processes (DGPs). 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 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 paper focuses on the robustness of non-linearity tests against asymmetry of innovations.

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

The main task of this chapter is to fill the gap in the literature and assess the robustness of selected non-linearity tests to: (i) a parameter variation of the data generating process; (ii) moment condition failure of innovations; (iii) asymmetry of innovations.

The following conclusions emerge from the results. First, the non-linearity tests considered in this chapter are all sensitive to the parameter configuration of DGPs. In particular, the tests produce robust results about non-linearity testing only in less than 50 % of all cases. Second, there seems not to be a direct link between a power variation of the tests and moment condition failure of model innovations. Third, and rather surprisingly, a power variation of the tests is significantly higher in the case of asymmetric innovations compared to innovations with moment failure.

The paper 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.

2 Nonlinearity Tests

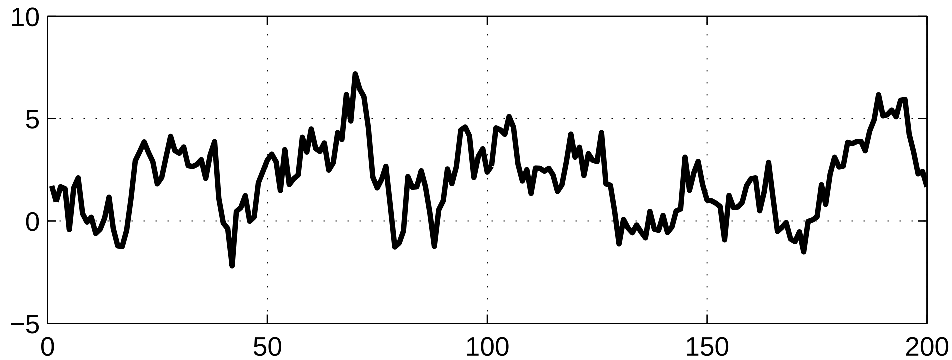
2.1 Null Hypothesis

In all cases discussed below, we assume that under the null hypothesis of linearity, the sequence $\{Y_t : t \in \mathbb{Z}\}$ is a realization from a simple AR(p) process given by

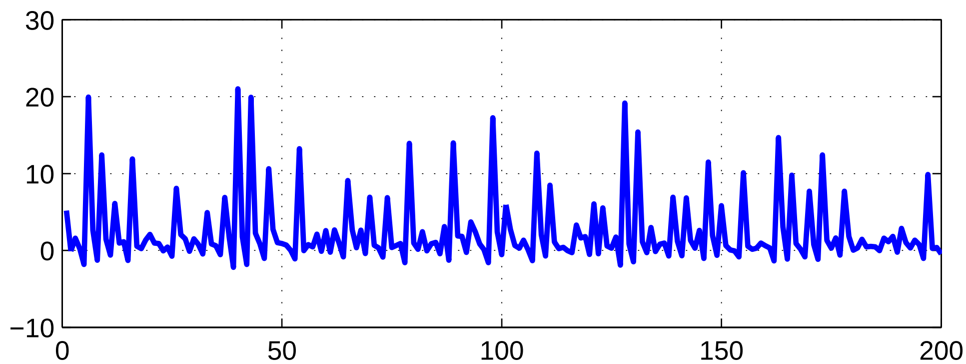
$$Y_t = \xi_0 + \xi_1 Y_{t-1} + \cdots + \xi_p Y_{t-p} + a_t = \boldsymbol{\xi}' \mathbf{X}_t + a_t, \tag{2}$$

where $\{a_t : t \in \mathbb{Z}\}$ is a sequence of IID(0,σ^2) innovations, $\mathbf{X}_t = (1, Y_{t-1}, \ldots, Y_{t-p})'$ is a $(p + 1 \times 1)$ vector of predetermined variables and $\boldsymbol{\xi} = (\xi_0, \xi_1, \ldots, \xi_p)'$ is a $(p + 1 \times 1)$ vector of unknown parameters. Moreover, we assume that all roots of $\xi(z) = 1 - \sum_{i=1}^{q} \xi_i z^i$ polynomial lie outside the unit circle. It is worth noting that the null hypothesis can be easily extended also to a linear ARMA model, or a 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) and $\hat{\sigma}^2$ is the sample variance of residuals, unless otherwise stated.

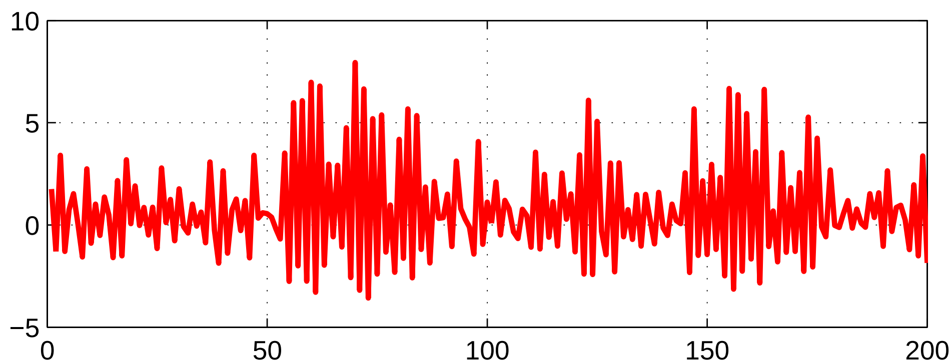
Figure 1: Different realizations of a TAR model: $N(0,1)$ innovations



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



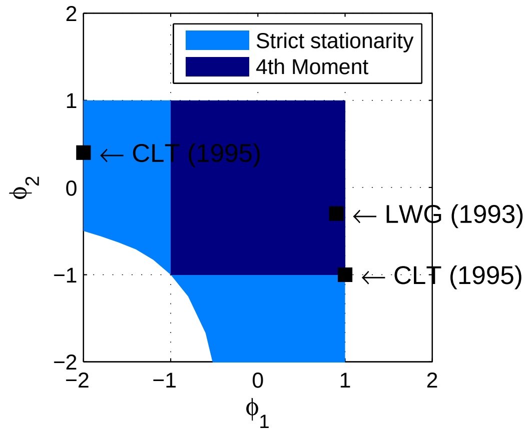
(b) $\phi_1 = 0.1$, $\phi_2 = -10.0$



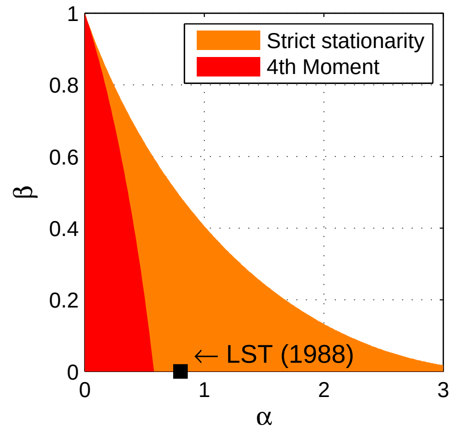
(c) $\phi_1 = 0.4$, $\phi_2 = -2.0$

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

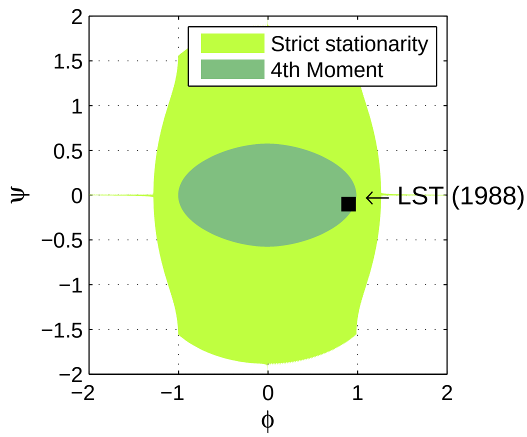
Figure 2: Moment failure of non-linear models



(a) TAR model



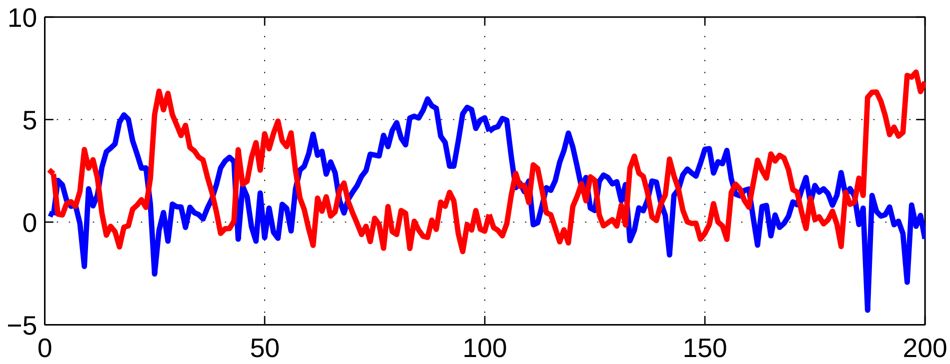
(b) GARCH model



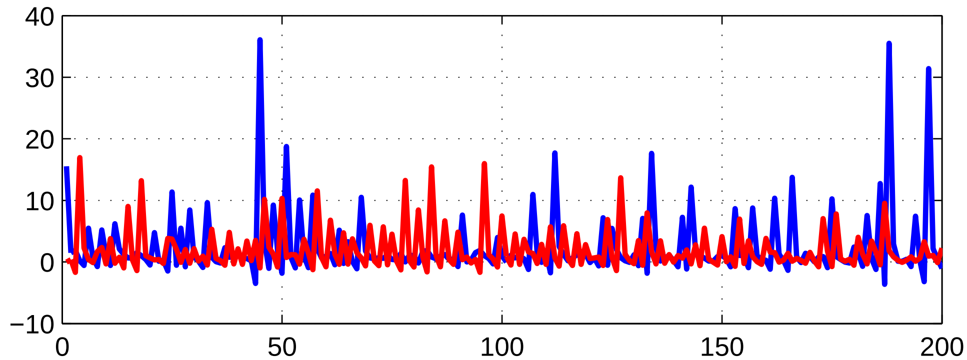
(c) BL model

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 $a \sim N(0,1)$, if necessary, whereas 4th-Moment regions represent an 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: $Y_t = \phi_1 Y_{t-1} I(Y_{t-1} > 0) + \phi_2 Y_{t-1} I(Y_{t-1} \leq 0) + a_t$, (b) a GARCH(1,1) model: $Y_t = a_t \sqrt{h_t} = \epsilon_t, h_t = \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1}$, (c) a BL(1,0,1,1) model: $Y_t = \phi Y_{t-1} + \theta a_{t-1} + \psi Y_{t-1} a_{t-1} + a_t$. We assume that $\{a_t : t \in \mathbb{Z}\}$ is a sequence of NID(0,1) innovations in all models.

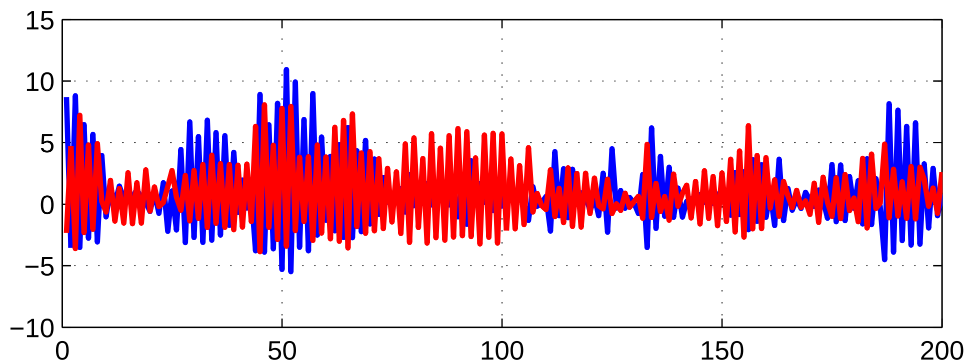
Figure 3: Different realizations of a TAR model: asymmetric innovations



(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: $Y_t = \phi_1 Y_{t-1} I(Y_{t-1} > 0) + \phi_2 Y_{t-1} I(Y_{t-1} \leq 0) + a_t$, where $\{a_t : t \in \mathbb{Z}\}$ is a sequence of IID innovations drawn from a GLD family: the blue line corresponds to A2(+) specification (skewness = 1.5, kurtosis = 7.5), and the red line to A2(−) specification (skewness = −1.5, kurtosis = 7.5), see Table 4 for details. Particular model parameters come from Petruccelli and Woolford (1984).

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 [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).

2.2.1 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

$$BDS(n,\epsilon) = \sqrt{N}\left(\frac{C(n,\epsilon) - C(1,\epsilon)^n}{\sqrt{\sigma^2(n,\epsilon)}}\right) \xrightarrow{d} N(0,1), \tag{3}$$

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

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

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

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 on the parameter ϵ for moderate values of the n–history; (b) for the moderate n–history, the standard normal distribution is relatively a good asymptotic approximation. 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).

2.2.2 McLeod–Li Test

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

$$MLQ(m) = T(T+2)\sum_{j=1}^{m}\frac{\hat{\rho}_j^2}{T-j} \xrightarrow{d} \chi^2(m), \tag{4}$$

where T is the sample size, m is the lag order of the test, and $\hat{\rho}_j$ is the j-th sample correlation 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

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

where $\hat{a}_t$ is the estimated residual, $\hat{\sigma}$ is the estimated standard error of residuals. 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 on the lag order specification m. For this reason, we implement the Q test with the lag order m automatically selected by a procedure developed in Escanciano and Lobato (2009).

2.2.3 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

$$MQ(m) = T(T+2)\sum_{j=1}^{m}\frac{\hat{\pi}_j^2}{T-j} \xrightarrow{d} \chi^2(m), \tag{5}$$

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, we run the test with the lag order m automatically selected by a procedure developed in Escanciano and Lobato (2009).

2.2.4 Tsay Test

In order to improve the power of 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 denotes a half-stacking operator. The LM version of the test statistic is defined as

$$TSAY(p) = TR^2 \xrightarrow{d} \chi^2(p(p+1)/2), \tag{6}$$

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).

2.2.5 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

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

where $G(\cdot)$ is the so called transition function, $\mathbf{X}_t = (1, Y_{t-1}, \ldots, Y_{t-p})$ 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 approximation of the transition function $G(\cdot)$ by a Taylor approximation. Then, an auxiliary equation is given by

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

where $\mathbf{X}_t = (1, Y_{t-1}, \ldots, Y_{t-p})$ is a $(p+1 \times 1)$ vector of predetermined variables, $\mathbf{Y}_t = (Y_{t-1}, \ldots, Y_{t-p})$ is a $(p \times 1)$ vector of predetermined variables, $\mathbf{Z}_t = (Y_{t-1}^3, \ldots, Y_{t-p}^3)$ 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 diagonal matrix. The main advantage of the Taylor approximation of a given transition function 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

$$STAR(p) = TR^2 \xrightarrow{d} \chi^2(p(p+1)/2 + p), \tag{7}$$

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

2.2.6 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), 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 $l_t(\cdot)$ is 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 is a stacking operators 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

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

where $\mathbf{X}_t = (1, Y_{t-1}, \ldots, Y_{t-p})'$ 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 test takes the form of a simple Lagrange multiplier test and the relevant test statistic is given by

$$WHITE(q) = TR^2 \xrightarrow{d} \chi^2(q), \tag{8}$$

where T denotes a sample size, and R^2 is a 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 $\hat{\mathbf{s}}_t$.

2.2.7 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 Y 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 $\boldsymbol{\psi}_t = (\psi_1(\mathbf{X}_t'\boldsymbol{\gamma}_1), \ldots, \psi_k(\mathbf{X}_t'\boldsymbol{\gamma}_k))'$, where the individual squashing functions ψ_j are defined as follows

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

In order to eliminate the identification problem, the authors recommend to use randomly generated real-valued parameter vectors $\boldsymbol{\gamma}_j$, for $j = 1, \ldots, k$, 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 with collinearity in the model. The number of principal component is set to $k^* = 2p$. The test is based on running an auxiliary regression

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

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

$$NN(2p) = TR^2 \xrightarrow{d} \chi^2(2p), \tag{9}$$

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 bootstrap, see Blake and Kapetanios (2003). This approach would be very computationally intensive in our case.

3 Time Series Models and Monte Carlo Setup

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, it includes some of the most commonly used non-linear time series models. The models are summarized in Table 1.

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

Table 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:

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

M9: BL model:

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

M10: NLMA model:

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

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 model and RCAR models, see Table 2 for particular parameter configurations. Gaussian innovations are consider for all time series models when inspecting the robustness of the power properties of the tests against parameter configurations. Note that parameters of all non-linearity models 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 exception are S3, S4, S5, S6, A3 specifications of model innovations, for which the 6th-moment stationarity is not satisfied. The parameter configurations satisfying stationarity and 6th-moment stationarity and/or invertibility conditions, denoted as "Monte Carlo", are graphically depicted in Figure 4.

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, 6 different specifications from $t(3)$ to $t(8)$ are considered, see Table 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: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, for $0 \leq u \leq 1$. In particular, we consider 6 specifications of asymmetric distributions, which differ in the magnitude of asymmetry, see Table 4. All generated innovations are normalized to have zero mean and unit variance.

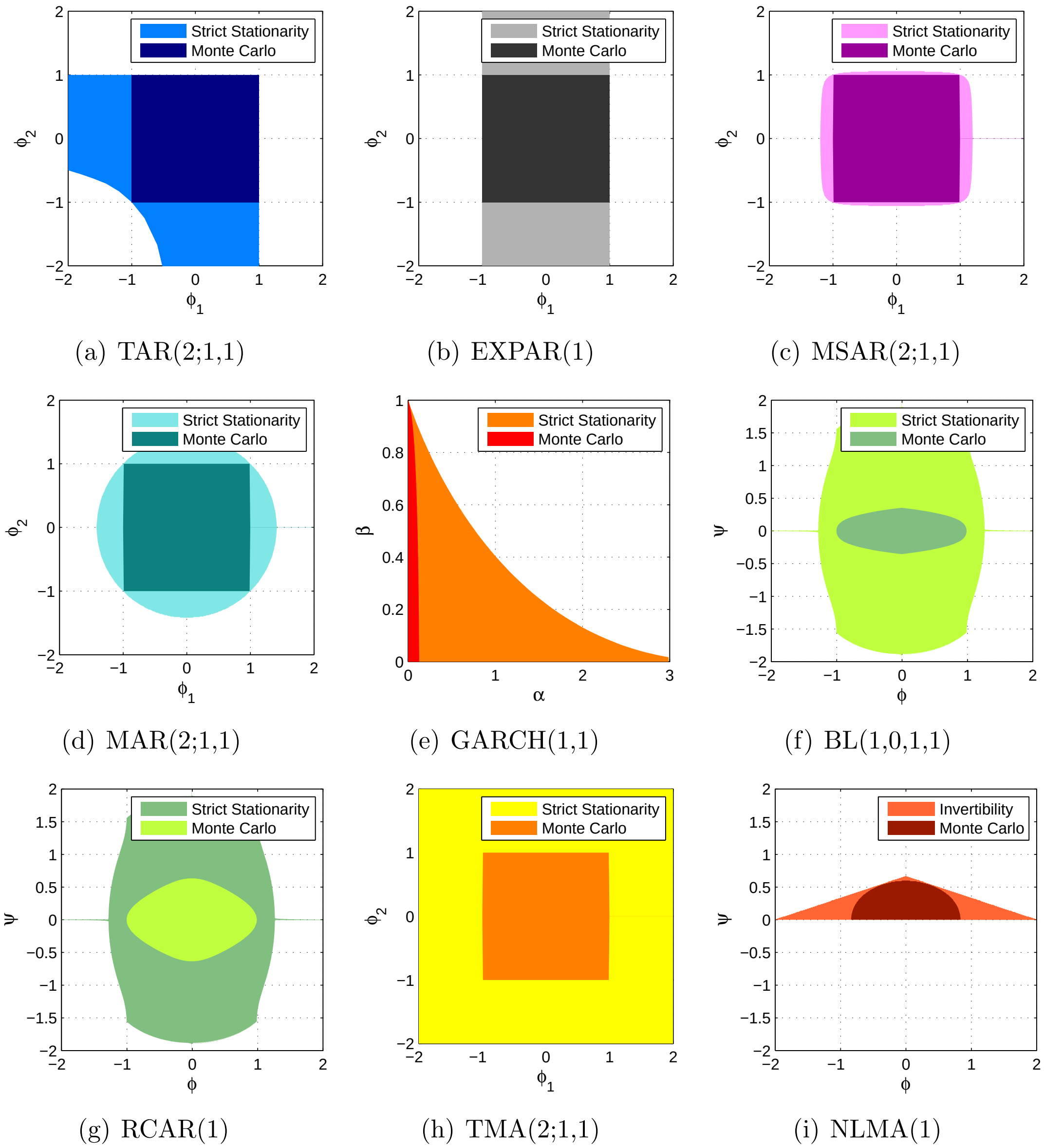
3.3 Monte Carlo Setup

We simulate originally T+100 observations 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 $R = 1000$. 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

Table 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, EXPRA	$c_1 = -0.25$, $c_2 = 0.25$ (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)
	$(\phi_1, \phi_2) \in \left\{ \begin{matrix} (-0.8, -0.8) & (-0.8, -0.5) & (-0.8, -0.2) & (-0.8, 0.2) & (-0.8, 0.5) \\ (-0.8, 0.8) & (-0.5, -0.8) & (-0.5, -0.5) & (-0.5, 0.5) & (-0.5, 0.8) \\ (-0.2, -0.8) & (-0.2, 0.8) & (0.2, -0.8) & (0.2, 0.8) & (0.5, -0.8) \\ (0.5, -0.5) & (0.5, 0.5) & (0.5, 0.8) & (0.8, -0.8) & (0.8, -0.5) \\ (0.8, -0.2) & (0.8, 0.2) & (0.8, 0.5) & (0.8, 0.8) & \end{matrix} \right\}$
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)
	$(\phi_1, \phi_2) \in \left\{ \begin{matrix} (-0.8, -0.8) & (-0.8, -0.5) & (-0.8, -0.2) & (-0.8, 0.2) & (-0.8, 0.5) \\ (-0.8, 0.8) & (-0.5, -0.8) & (-0.5, -0.5) & (-0.5, 0.5) & (-0.5, 0.8) \\ (-0.2, -0.8) & (-0.2, 0.8) & (0.2, -0.8) & (0.2, 0.8) & (0.5, -0.8) \\ (0.5, -0.5) & (0.5, 0.5) & (0.5, 0.8) & (0.8, -0.8) & (0.8, -0.5) \\ (0.8, -0.2) & (0.8, 0.2) & (0.8, 0.5) & (0.8, 0.8) & \end{matrix} \right\}$
GARCH	$c = 1$
	$\phi = 0.5$
	$\sigma^2 = 1$
	$(\alpha, \beta) \in \left\{ \begin{matrix} (0.05, 0.2) & (0.05, 0.3) & (0.05, 0.4) & (0.05, 0.5) & (0.05, 0.6) \\ (0.05, 0.7) & (0.05, 0.8) & (0.05, 0.9) & (0.10, 0.2) & (0.10, 0.3) \\ (0.10, 0.4) & (0.10, 0.5) & & & \end{matrix} \right\}$
TMA	$c_1 = -0.25$, $c_2 = 0.25$
	$\sigma^2 = 1$
	$(\phi_1, \phi_2) \in \left\{ \begin{matrix} (-0.8, -0.8) & (-0.8, -0.5) & (-0.8, -0.2) & (-0.8, 0.2) & (-0.8, 0.5) \\ (-0.8, 0.8) & (-0.5, -0.8) & (-0.5, -0.5) & (-0.5, 0.5) & (-0.5, 0.8) \\ (-0.2, -0.8) & (-0.2, 0.8) & (0.2, -0.8) & (0.2, 0.8) & (0.5, -0.8) \\ (0.5, -0.5) & (0.5, 0.5) & (0.5, 0.8) & (0.8, -0.8) & (0.8, -0.5) \\ (0.8, -0.2) & (0.8, 0.2) & (0.8, 0.5) & (0.8, 0.8) & \end{matrix} \right\}$
BL, RCAR	$c = 1$
	$\sigma^2 = 1$
	$(\phi, \psi) \in \left\{ \begin{matrix} (-0.8, -0.2) & (-0.6, -0.2) & (-0.4, -0.2) & (-0.2, -0.2) & (-0.2, 0.2) \\ (-0.4, 0.2) & (-0.6, 0.2) & (-0.6, 0.4) & (-0.8, 0.2) & (0.2, -0.2) \\ (0.2, 0.2) & (0.4, -0.2) & (0.4, 0.2) & (0.6, -0.2) & (0.6, -0.4) \\ (0.6, 0.2) & (0.8, -0.2) & (0.8, 0.2) & & \end{matrix} \right\}$
NLMA	$c = 1$
	$\sigma^2 = 4$
	$(\phi, \psi) \in \left\{ \begin{matrix} (-0.20, 0.20) & (-0.20, 0.40) & (-0.40, 0.20) & (-0.40, 0.40) & (-0.60, 0.20) \\ (-0.60, 0.40) & (0.20, 0.20) & (0.20, 0.40) & (0.40, 0.20) & (0.40, 0.40) \\ (0.60, 0.20) & (0.60, 0.40) & & & \end{matrix} \right\}$

Figure 4: Parameter configurations of time series models



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

Table 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

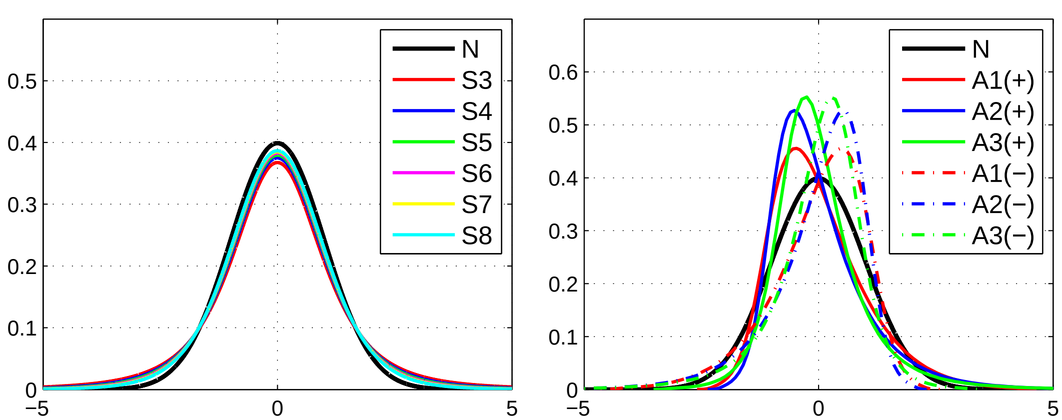
[a] The maximum exponent of a given distribution.

Table 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 maximum exponent of a given distribution.

Figure 5: Distributions of model innovations



(a) Symmetric densities (b) Asymmetric densities 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 for AR(p) models is estimated by the following simple rule: $\hat{p} = \min_{l\in\{1,\ldots,L\}}(BIC_l)$.

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

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

where Q_l is a value of the Q tests, π_l is a penalization function, $c = 2.4$ is a correction constant recommended by Escanciano and Lobato (2009) based on Monte Carlo experiments. Finally, the lag order m for the Q tests is determined by the following simple rule: $\hat{m} = \max_{l\in\{1,\ldots,L\}}(Q_l^*)$.

4 Monte Carlo Results

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

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

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

where $\mathcal{P}_{i,j}$ is the rejection frequency of the test for a given parameter configuration $i \in \{1, \ldots, K\}$ and distribution of innovations $j \in \{1, \ldots, 13\}$. The following set of 13 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. The sample size is $T \in \{200, 500, 1000\}$.

The rejection frequency $\mathcal{P}_{i,j}$ is given by

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

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 a parameter configuration of DGPs is assessed using a modified coefficient of variation. For a given test, a given DGP, and Gaussian innovations, the coefficient of variation is calculated as follows

$$cv = \frac{\max_i(\mathcal{P}_{i,1}) - \min_i(\mathcal{P}_{i,1})}{avg_1}, \tag{10}$$

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. The coefficient of variation is denoted as $cv(N)$ in tables below.

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

$$cv = \frac{\max_j(avg_j) - \min_j(avg_j)}{avg_1}, \tag{11}$$

where avg_1 represents the average rejection frequency calculated based on Gaussian innovations over all parameter K parameter configurations of the 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.

4.2 Monte Carlo Results: Parameters

Size: The size results of the tests are presented in Table 7. The results reveal that all non-linearity tests considered in the paper 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 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, 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 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 also worth noting that a BL model is easily recognized by all 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 the test from each group.

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 simple coefficient of variation, calculated according to (10), is reported in Table 8 as well. One can logically expect that the more powerful the test is, which means the higher the average frequency, the lower the coefficient of variation. Our results clearly show the expected relationship between the average rejection frequency and its variation holds, and is insensitive to the sample size, but only for the average rejection frequency exceeding 0.6, see Figure 6(a). Note, however, that even tests with a very high average rejection frequency exceeding 0.9 can suffer from a relatively high

[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 $\mathbf{S}$.

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

coefficient of variation close to 1, see see Figure 6(a). Unfortunately, no clear conclusion can be make for the tests with the average rejection frequency less than 0.4. Our results clearly indicate that, although the selected non-linearity do exhibit power against a given set of DGPs, the power results are very sensitive to the parameter specification of DGPs.

In order to make this point clear, the individual Monte Carlo results are presented in the form of graphical images. 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, we use a color range (from black to white) indicating different magnitude of the statistical significance of the non-linearity tests, see Figure 7 and 8 for the case of Gaussian innovations. 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 of all images). However, completely opposite results are obtained for a GARCH model, where almost all the estimated p-values are much larger than the significance level $\alpha = 0.05$, but the results are not sensitive on the parameter configuration (see the orange color of all images). These two outcomes, although completely opposite, are in favor of given test statistics since they give us clear and reliable information about non-linearity testing. 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 annealing color of the image).

All in all, nine non-linear models are examined using a battery of eight standard non-linearity test. The presented graphical images suggest that 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 the cases. Our results suggest that one should interpret the results about non-linearity testing with caution, since linearity does not have to be rejected only because of a particular parameter configuration of a purely non-linear process.

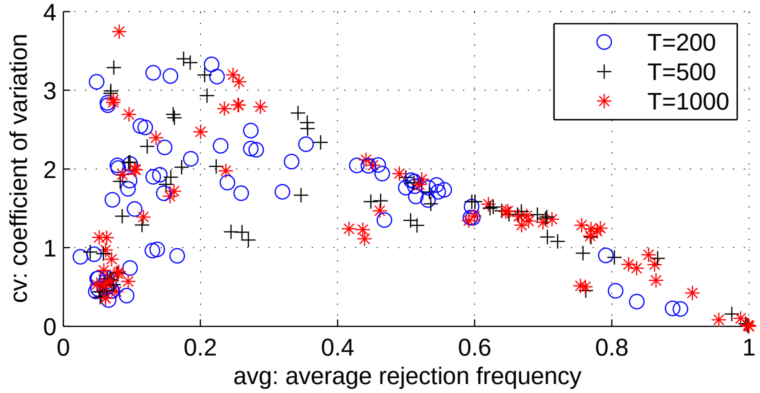
4.3 Monte Carlo Results: Moments and Asymmetry

The power results of the tests based on moment condition failure and asymmetry of innovations are presented in Tables 9 – 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 of a given distribution is even higher than a given test statistic requires.

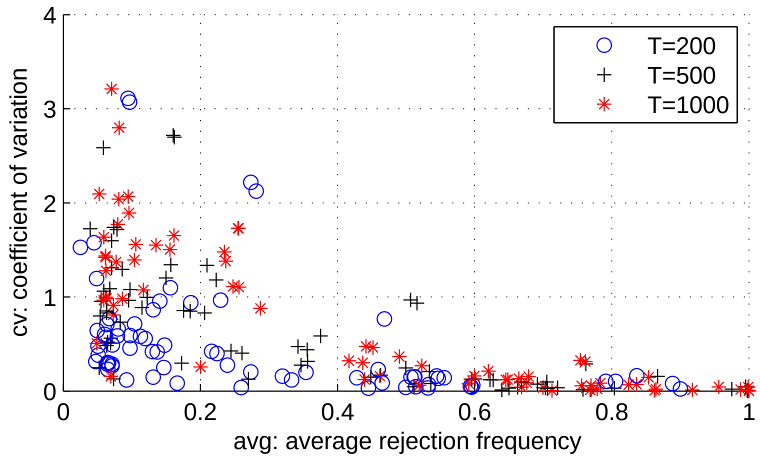
Moment condition failure: Our results suggest that the power variation of the non-linearity tests is extremely model rather than test dependent. For example, the BDS test, although it requires the existence of only the second moment of estimated residuals, suffers from high variability of the average 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 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 regardless the sample size) 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 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 regardless the sample size) for MAR, MSAR, and GARCH models, but large variation for TMA and NLMA models, see the last column in Table 10. Figure 6(b) depicts 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. 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 other hand, 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

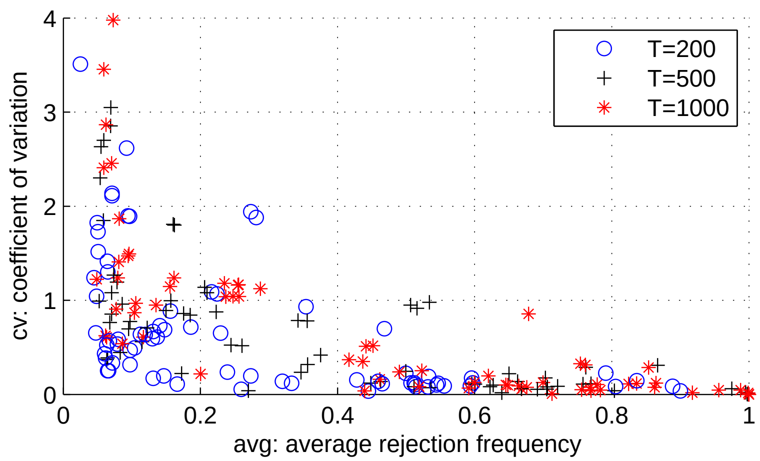
Figure 6: Power variation of the tests



(a) Gaussian innovations



(b) moment failure



(c) asymmetry

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, and six specifications from a Student t distribution, see Table 3, and also six specifications from a generalized lambda distribution, see Table 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 5.

The results indicate that there seems not to be a clear cut off between the moment requirement of a given test and the robustness of its power against either moment condition failure or asymmetry of innovations. For example, the null hypothesis about no change (i.e. $H_0 : avg_j = avg_1$) is 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 power variability in general. 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. Moreover, the results clearly confirm that the power of the tests under moment condition failure of model innovations is inflated upwards.
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 5. The results indicate that the average rejection frequencies of the tests based on asymmetric innovations suffer from even a higher variation compared to moment condition failure. The only exception is 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 frequencies of the tests are statistically significantly inflated upward in the case of asymmetric innovations.

[5] Moreover, since we 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.

Table 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.

5 Summary and Conclusion

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

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 power against a parameter configuration 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 are presented in Table 25. 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 6. The main reason for using the 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).[6]

Table 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.

The results reveal that the best non-linearity tests with the highest average rejection frequency across all nine non-linear time series models are the BDS, MLQ tests. On the other hand, 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 much more sophisticated non-linearity tests such as the $WHITE$ or NN tests. On the other hand, the Monte Carlo results reveal that more powerful tests suffer from high variability of the power. 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. The ordering of the non-linearity tests according to moment failure and asymmetry of innovations gives rather different results. In the case of moment failure, the lowest variability of the power is obtained for the BDS, NN, $WHITE$ and MQ tests. The worst test, suffering from the highest power variation,

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

seems to be the $TSAY$ test. On the other hand, in the case of asymmetric innovations, the lowest variability of the power is obtained for the $STAR$, $WHITE$, and NN tests. The worst test, suffering from the highest power variation, seems to be the BDS test.

All in all, since we face a problem of heavy-tailed time series rather than asymmetric ones in economics and finance, we put subjective priority to the results from a symmetric Student t distribution compared to asymmetric innovations. Based on this decision, the best tests, according to their power and its variability, are the following: the BDS, NN, followed by MLQ and $STAR$ tests.

References

Adler, R., R. Feldman, and C. Gallagher (1998). Analysing stable time series. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, 133–58.

Blake, A. and G. Kapetanios (2003). A radial basis function artificial neural network test for neglected nonlinearity. *Econometrics Journal 6*(2), 357–373.

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*(1), 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*(3), 197–235.

Casella, G. and R. Berger (2001). *Statistical Inference.* Duxbury Press.

Chen, R., J. Liu, and R. Tsay (1995). Additivity tests for nonlinear autoregression. *Biometrika 82*(2), 369–383.

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*(2), 533–558.

de Lima, P. (1997). On the robustness of nonlinearity tests to moment condition failure. *Journal of Econometrics 76*(1-2), 251–280.

Escanciano, J. and I. Lobato (2009). An automatic Portmanteau test for serial correlation. *Journal of Econometrics 151*(2), 140–149.

Hansen, B. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica 64*(2), 413–430.

Hsieh, D. (1989). Testing for nonlinear dependence in daily foreign exchange rates. *Journal of Business 62*(3), 339–368.

Kearns, P. and A. Pagan (1997). Estimating the density tail index for financial time series. *Review of Economics and Statistics 79*(2), 171–175.

Keenan, D. (1985). A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika 72*(1), 39–44.

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*(3), 269–290.

Lin, J. and A. McLeod (2008). Portmanteau tests for arma models with infinite variance. *Journal of Time Series Analysis 29*(3), 600–617.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988). Testing linearity in univariate time series models. *Scandinavian Journal of Statistics 15*(3), 161–175.

McCulloch, J. (1997). Measuring tail thickness to estimate the stable index α: A critique. *Journal of Business & Economic Statistics 15*(1), 74–81.

McLeod, A. and W. Li (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis 4*(4), 269–273.

Monti, A. (1994). A proposal for a residual autocorrelation test in linear models. *Biometrika 81*(4), 776–780.

Ng, S. and P. Perron (2005). A note on the selection of time series models. *Oxford Bulletin of Economics and Statistics 67*(1), 115–134.

Petruccelli, J. and S. Woolford (1984). A threshold AR (1) model. *Journal of Applied Probability 21*(2), 270–286.

Psaradakis, Z. and N. Spagnolo (2002). Power properties of nonlinearity tests for time series with Markov regime. *Studies in Nonlinear Dynamics and Econometrics 6*(3), 1–14.

Ramsey, J. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society 31*(2), 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*(369), 168–172.

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*(2), 205–216.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464.

Tong, H. (1990). *Non-linear Time Series: a dynamical system approach.* Oxford University Press.

Tsay, R. (1986). Nonlinearity tests for time series. *Biometrika 73*(2), 461–466.

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.

A Appendix A: Tables

Table 7: Size of the non-linearity tests: AR (#8), N(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 optimal lag order 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 $T = 1000$. The lag order m of the Q tests is determined by an optimal 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 8: Power of the non-linearity tests: N(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 optimal lag order 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 $T = 1000$. The lag order m of the Q tests is determined by an optimal 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 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 optimal 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. The significance level is set to $\alpha = 0.05$.

[d] "cv(S)" denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations.

[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 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 optimal lag selection procedure discussed in Ng and Perron (2005). The lag order m of the Q tests is determined by an optimal 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. The significance level is set to $\alpha = 0.05$.

[d] "cv(S)" denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations.

[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 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 optimal lag selection procedure discussed in Ng and Perron (2005). The lag order m of the Q tests is determined by an optimal 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. The significance level is set to $\alpha = 0.05$.

[d] "cv(S)" denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations.

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

[d] "cv(S)" denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations.

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

[d] "cv(S)" denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations.

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

[d] "cv(S)" denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations.

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

[d] "cv(S)" denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations.

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

[d] "cv(S)" denotes a coefficient of variation calculated from test statistics using symmetric (S) innovations.

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

[d] "cv(A)" denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations.

[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 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 optimal lag selection procedure discussed in Ng and Perron (2005). The lag order m of the Q tests is determined by an optimal 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. The significance level is set to $\alpha = 0.05$.

[d] "cv(A)" denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations.

[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 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 optimal lag selection procedure discussed in Ng and Perron (2005). The lag order m of the Q tests is determined by an optimal 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. The significance level is set to $\alpha = 0.05$.

[d] "cv(A)" denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations.

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

[d] "cv(A)" denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations.

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

[d] "cv(A)" denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations.

[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 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 optimal 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. The significance level is set to $\alpha = 0.05$.
[d] "cv(A)" denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations.
[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 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**	**rc(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 optimal 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. The significance level is set to $\alpha = 0.05$.

[d] "cv(A)" denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations.

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

[d] "cv(A)" denotes a coefficient of variation calculated from test statistics using asymmetric (A) innovations.

[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 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

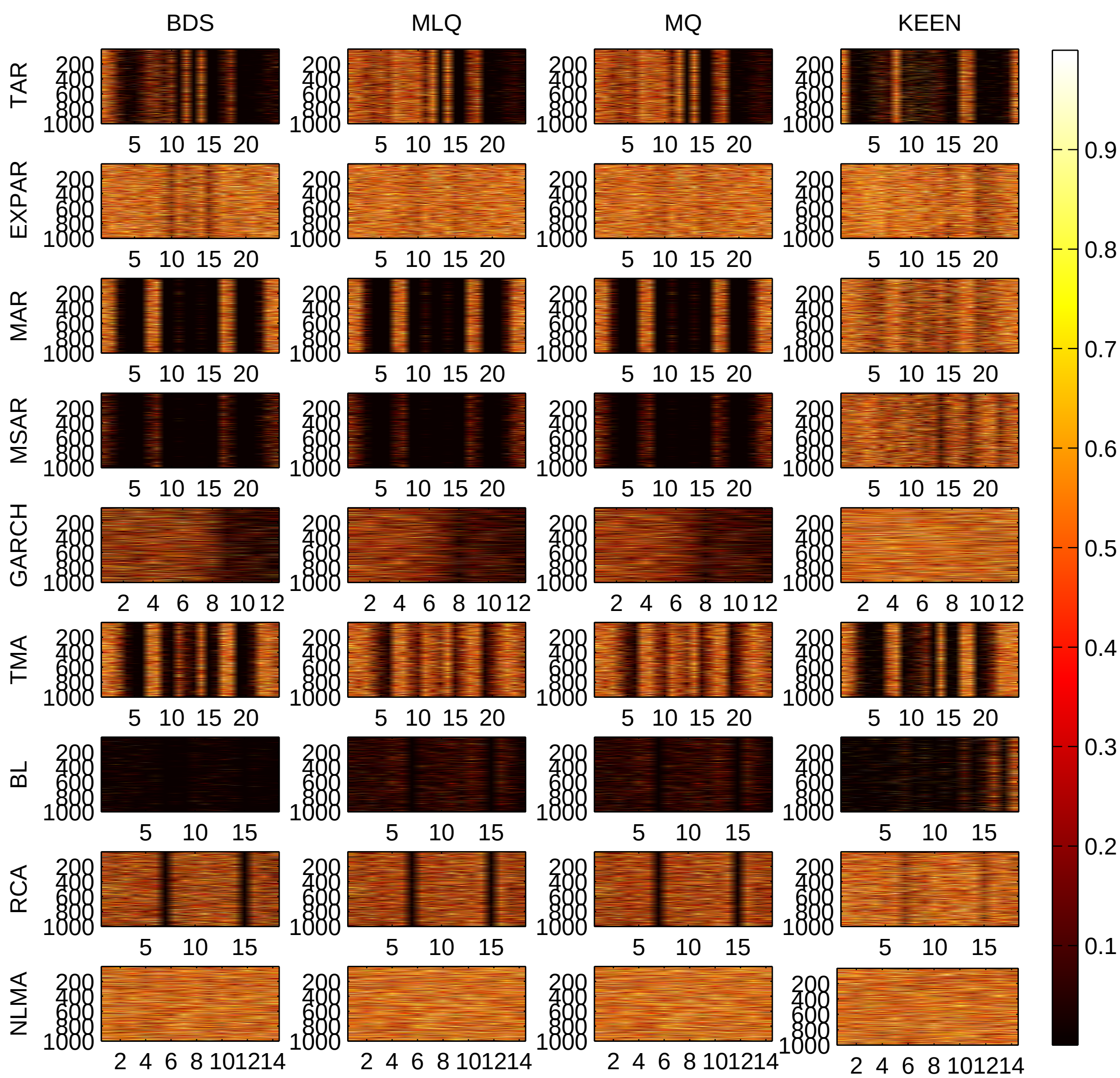
	TSAY				**STAR**				**WHITE**				**NN**			
	avg(N)	cv(N)	cv(S)	cv(A)	avg(N)	cv(N)	cv(S)	cv(A)	avg(N)	cv(N)	cv(S)	cv(A)	avg(N)	cv(N)	cv(S)	cv(A)
TAR	3	4	4	4	2	2	1	2	4	3	2	3	1	1	3	1
EXPAR	5	8	7	8	2	4	2	2	3	5	3	3	1	3	1	1
MAR	6	6	8	8	4	7	5	5	5	8	6	6	8	1	4	4
MSAR	7	7	7	8	6	5	5	1	4	4	4	2	5	6	6	3
GARCH	5	4	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 the non-linearity tests based on Gaussian innovations and all parameter configurations for a given non-linear model, "cv(N)" stands for the coefficient of variation of a given test statistic 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.

[b] "median" represents a median ordering using individual results about ordering for all non-linear models.

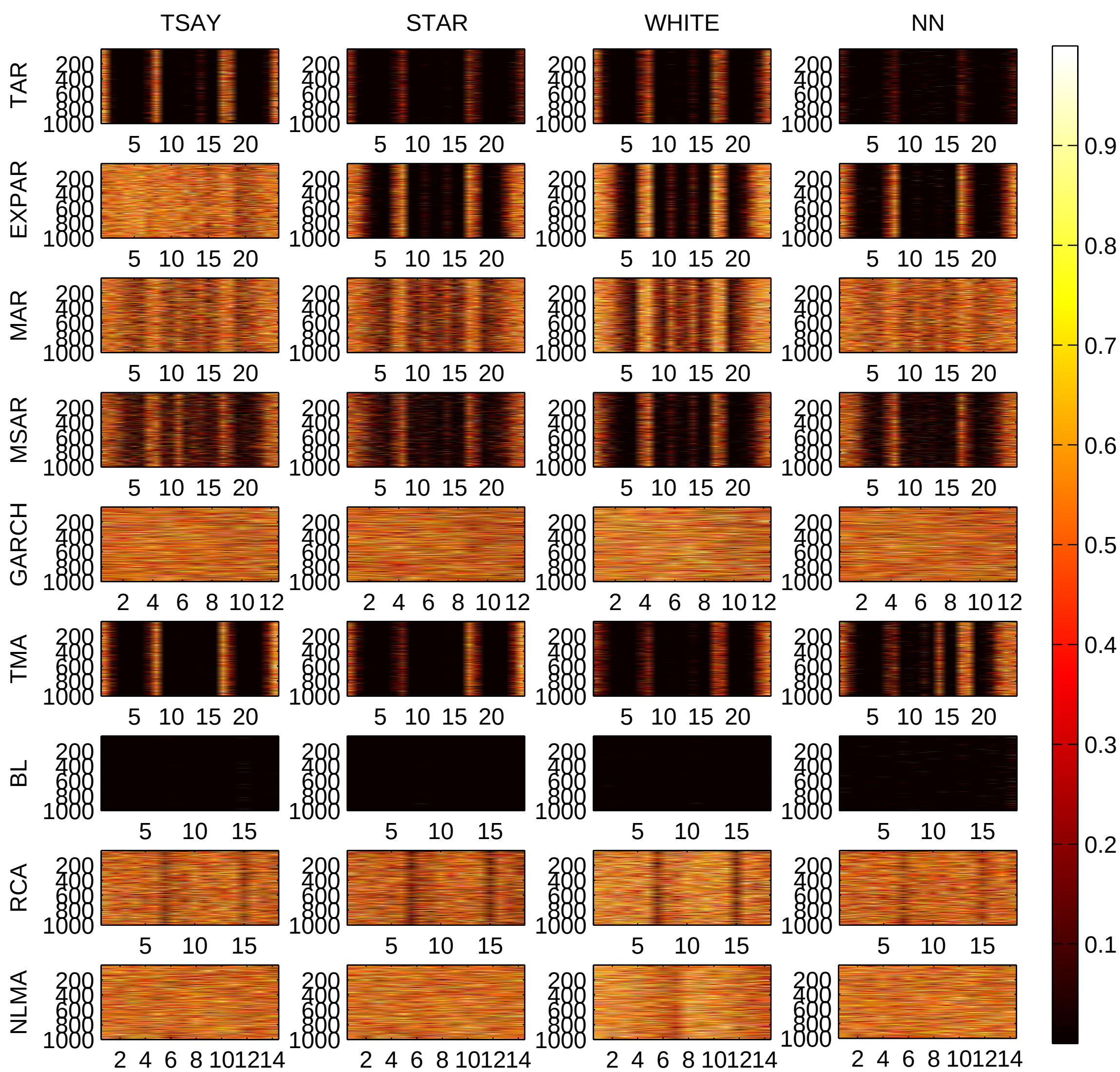
B Appendix B: Figures

Figure 7: Power images of non-linearity tests: part 1



Note: the x axes denotes the number of parameter configurations K of a given time series model and the y axes denotes the number of repetitions R.

Figure 8: Power images of non-linearity tests: part 2



Note: the x axes denotes the number of parameter configurations K of a given time series model and the y axes denotes the number of repetitions R.