Testing for non-linearity in multivariate stochastic processes

Marián Vávra

Testing for non-linearity in multivariate stochastic processes[1]

Working paper NBS

Marián Vávra[2]

Abstract

Two well known multivariate non-linearity tests are modified using a principal component analysis. The Monte Carlo results show that the proposed principal component-based tests do provide a remarkable dimensionality reduction without any systematic power loss. It can be concluded that using linear dynamic economic models is in sharp contrast with our empirical findings.

[2] Marián Vávra, Advisor to the Governor and Research Department of the NBS, marian.vavra@nbs.sk.

1. INTRODUCTION

Non-linear time series analysis has become a progressively growing part of statistics in the last decades. The main reason for its popularity lies in the fact that non-linear models are capable to capture characteristic features observed in stochastic processes (e.g. regime switching, time-varying volatility, etc.) which cannot be adequately accounted for by any linear models, see Tong (1990) and Granger and Teräsvirta (1993), among others. Although the mainstream literature has focused mainly on univariate non-linear models, there are situations where a set of variables are dependent in a non-linear way. Therefore, increasing attention has been paid to non-linear multivariate time series models. Many examples can be found in economics and finance, see Sims and Zha (2006), Rudebusch and Swanson (2008), or Engle and Kroner (1995), among others. Despite recent advances in the computer science, the identification, estimation, and forecasting from multivariate non-linear models is still very computationally intensive.[3] Therefore, it is desirable to test for non-linearity in the first place.

However, it is by no means easy to do so in multivariate systems. Two problems immediately arise. First, although there exist many univariate non-linearity tests in the literature, see Tong (1990) for a survey, there is no guarantee that the univariate tests can adequately capture non-linearity in multivariate processes. Therefore, the use of multivariate tests seems to be more appropriate. Second, unfortunately, there are only a few multivariate tests available in the literature. And those existing tests often suffer from the dimensionality problem. It means that, due to the construction of the tests, they require the large number of observations, which is not feasible to get in practice.[4]

This article brings three contributions to the literature of testing for non-linearity. First, the dimensionality issue of two selected multivariate tests (i.e. the TSAY and ARCH tests) is bypassed by means of a principal component analysis. Second, although it will be shown that principal components can reduce, or even completely eliminate, the dimensionality problem, there is still an ultimate question of how many components to retain for the tests. Therefore, special attention is paid to the finite sample properties of the new principal component-based multivariate tests under different stopping rules determining the number of components to retain. Third, we show, analytically and by means of Monte Carlo experiments, that univariate tests can completely fail when testing for non-linearity in multivariate systems.

The paper is organized as follows. A brief description of two multivariate non-linearity tests is

---

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

[4]Examples are provided in the next section.

given in Section 2. The Monte Carlo setup and results are discussed in Sections 3 and 4. An empirical example is provided in Section 5.

2. NON-LINEARITY TESTING

2.1 WHY MULTIVARIATE TESTS?

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

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

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

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

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

Obviously, the system is linear since $\boldsymbol{\Upsilon}_0 = \mathbf{0}$ and $\boldsymbol{\Upsilon}_1 = \boldsymbol{\xi}_1$ in our case. Now suppose that non-linearity is not tested for the whole vector $\mathbf{x}_t$, but all variables are treated individually using an univariate version of the definition above. For example, let us consider the variable X_{1t} and suppose that an appropriate filter is, only for simplicity of an explanation, an AR(1) process. Although it holds that $\mathbb{E}(X_{1t}|X_{1t-1}) \neq \mathbb{E}(X_{1t}|\mathbf{x}_{t-1})$, it also holds that $\mathbb{E}(X_{1t}|X_{1t-1}) = \xi_{11}X_{1t-1} + \xi_{12}\mathbb{E}(X_{2t-1}|X_{1t-1}) = (\xi_{11} + \xi_{12}\rho)X_{1t-1} = \beta_1 X_{1t-1}$. Therefore, the incorrect filtration (conditioning) does not have to necessarily lead to a size distortions under the null hypothesis of linearity.[5]

---

[5]However, it is worth noting that some size distortion can be expected in large-dimensional systems (i.e. $k > 5$). The author is very grateful to Professor Timo Teräsvirta from the Aarhus University for making this point.

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

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

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

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

Assumption 1 *Let us assume the following stationary real-valued finite-order linear VAR model under the null hypothesis of (conditional) linearity*

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

where $\mathbf{x}_t$ denotes a $(k \times 1)$ vector, $\{\mathbf{a}_t : t \in \mathbb{Z}\}$ is a sequence of multivariate WN($\mathbf{0}, \boldsymbol{\Sigma}$) innovations with zero means and the variance-covariance matrix $\boldsymbol{\Sigma}$, which is symmetric and positive definite, such that $\mathbb{E}(\|\mathbf{a}_t\|^4) < \infty$. Let $\boldsymbol{\beta} = (\boldsymbol{\xi}_0', \mathrm{vec}(\boldsymbol{\xi}_1)', \dots, \mathrm{vec}(\boldsymbol{\xi}_p)')'$ be a $(k^2 P + k \times 1)$ parameter vector, which is assumed to lie in the interior of the parameter space given by

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

$\square$

The assumption ensures that a given linear process is stationary, parameters do not lie on the

boundary, and all moment conditions are satisfied. These conditions are sufficient to ensure consistency of the estimated parameters in β, the estimated residuals, and subsequently, the non-linearity test statistics. Note that the null hypothesis about linearity can be extended to VARMA models as well.[6]

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

2.2 MULTIVARIATE TSAY TEST

Harvill and Ray (1999) proposed a multivariate version of the TSAY test, which can be considered as a generalization of the univariate case, see Tsay (1986). The test is based on running the following auxiliary equation

$$\hat{\mathbf{a}}_t = \mathbf{b}_0 + \mathbf{B}_1 \mathbf{z}_t + \mathbf{B}_2 \mathbf{v}_t + \mathbf{u}_t, \tag{2}$$

where $\hat{\mathbf{a}}_t$ is a $(k \times 1)$ vector of residuals from a particular VAR(P) filter, $\mathbf{z}_t = (\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-P})'$ denotes a $(kP \times 1)$ vector of aggregated predetermined variables, $\mathbf{v}_t = \text{vech}(\mathbf{z}_t \otimes \mathbf{z}'_t)$ represents an $(s \times 1)$ vector of predetermined variables consisting of all square and cross-product elements $(s = kP(kP + 1)/2)$. $\mathbf{b}_0$ denotes a $(k \times 1)$ vector of constants, $\mathbf{B}_1$ represents a $(k \times kP)$ matrix of parameters, and finally, $\mathbf{B}_2$ is a $(k \times s)$ matrix of parameters. The null hypothesis of linearity of the vector $\mathbf{x}_t$ is given by: $H_0 : \mathbf{B}_2 = \mathbf{0}$ versus $H_1 : \mathbf{B}_2 \neq \mathbf{0}$. The null hypothesis can be tested both by the LM and LR-based test statistics.[7] The appropriate LR-type test statistic is given by[8]

$$MTSAY(org) = (T - \tau)(\log(|\hat{\boldsymbol{\Sigma}}_r|) - \log(|\hat{\boldsymbol{\Sigma}}_u|)) \xrightarrow{d} \chi^2(sk), \tag{3}$$

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

---

[6]Recall that the above mentioned tests are neglected non-linearity test, where non-linearity is tested from estimated residuals after filtering conditional mean out of the system. However, identifying and estimating VARMA models is computationally expensive. For this reason, main attention is paid to VAR and non-linear VAR models in this paper only.

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

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

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

It is worth mentioning, however, that testing non-linearity using the above defined MTSAY(org) test statistic displays at least two shortcomings: (i) Due to the large number of terms in the vector $\mathbf{v}_t$, the original test requires the large number of observations; (ii) In addition, terms in the vector $\mathbf{v}_t$ are highly collinear, which significantly increases the degrees of freedom of the test, but actually does not improve the fit in (2). As a results, multicollinearity in the vector $\mathbf{v}_t$ can reduce the power of the multivariate tests. It is clear from the dimension of the $\mathbf{v}_t$ vector that, for a given set of k variables and the lag order P of a VAR filter , the test requires $T > kP(kP+1)/2$ observations, but $sk = k^2 P(kP+1)/2$ degrees of freedom, which implies that, for a given set of observations T, the power of the original MTSAY test reduces with increasing k and/or P. For example, consider a small model consisting of $k = 5$ different variables and the moderate lag order $P = 4$ of a VAR filter. As a result, the MTSAY(org) test requires $T > 210$ observations. Table 1 depicts some other examples of the degrees of freedom and the number of observations required by the original TSAY(org) test.

Table 1: Requirements of the original multivariate TSAY test

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

A dimensionality problem of the MTSAY test can be efficiently diminished using a principal component analysis, see next section for details. The modified multivariate TSAY test is based on running the following auxiliary equation

$$\hat{\mathbf{a}}_t = \mathbf{c}_0 + \mathbf{C}_1 \mathbf{z}_t + \mathbf{C}_2 \mathbf{w}_t + \mathbf{u}_t, \tag{4}$$

where $\hat{\mathbf{a}}_t$ is a $(k \times 1)$ vector of residuals from a particular filter, $\mathbf{c}_0$ is a $(k \times 1)$ vector of constants, $\mathbf{C}_1$ is an appropriate $(k \times kP)$ matrix of coefficients, and $\mathbf{w}_t$ is an $(n \times 1)$ vector of principal components, such that $k \leq n \leq s$. The null hypothesis about linearity of the vector $\mathbf{x}_t$, is given

by: $H_0 : \mathbf{C}_2 = \mathbf{0}$ versus $H_1 : \mathbf{C}_2 \neq \mathbf{0}$. The appropriate LR based test statistic is given by

$$TSAY = (T - \tau)(\log(|\mathbf{\Sigma}_r|) - \log(|\mathbf{\Sigma}_u|)) \xrightarrow{d} \chi^2(nk), \tag{5}$$

where $|\cdot|$ denotes the determinant of a square matrix, T stands for the sample size, $\tau = (k + n + 1)/2$ is a small sample correction term.

2.3 MULTIVARIATE ARCH TEST

Lütkepohl (2005, Ch. 16) considers a multivariate version of the ARCH test, which can be considered as a generalization of the univariate ARCH test, see Engle (1982).[9] The test is based on running the following auxiliary equation

$$\text{vech}(\hat{\mathbf{a}}_t \otimes \hat{\mathbf{a}}_t') = \mathbf{b}_0 + \mathbf{B}\mathbf{v}_t + \mathbf{u}_t, \tag{6}$$

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

$$MARCH(org) = (T - \tau)(\log(|\hat{\mathbf{\Sigma}}_r|) - \log(|\hat{\mathbf{\Sigma}}_u|)) \xrightarrow{d} \chi^2(sm), \tag{7}$$

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

It is worth mentioning, however, that testing the conditional variance using the above defined MARCH(org) test statistic displays several shortcomings as well. As in the previous case, the main problem does lie both in the number of observations directly required and in the rapidly in-

---

[9]Lee (1991) shows that a simple ARCH test is equivalent to the LM-based test for GARCH models as well. It makes sense to assume analogical results for multivariate GARCH models as well.

creasing degrees of freedom of the test. It is clear from the dimension of the $\mathbf{v}_t$ vector that, for a given set of k residuals, the test requires $T > Qk(k+1)/2$ observations, but $sm = Qk^2(k+1)^2/4$ degrees of freedom. It seems to be clear that, for a given set of observations T, the power of the original MARCH(org) test reduces with increasing k and Q. For example, consider a small model consisting of $k = 5$ different variables) and the moderate lag order $Q = 4$ of the MARCH test. As a result, the test requires $T > 60$ observations, but the degrees of freedom surge to 900. Table 2 depicts some other examples of the degrees of freedom and number of observations required by the original MARCH(org) test.[10]

Table 2: Requirements of the original multivariate ARCH test

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

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

$$\mathrm{diag}(\hat{\mathbf{a}}_t \otimes \hat{\mathbf{a}}_t') = \mathbf{c}_0 + \mathbf{C}_1\mathbf{w}_t + \mathbf{u}_t, \qquad (8)$$

where $\mathrm{diag}(\hat{\mathbf{a}}_t \otimes \hat{\mathbf{a}}_t')$ is a $(k \times 1)$ vector of diagonal elements, $\mathbf{c}_0$ is a $(k \times 1)$ vector of constants, $\mathbf{C}_1$ is an appropriate $(k \times n)$ matrix of coefficients, and $\mathbf{w}_t$ is an $(n \times 1)$ vector of principal components, such that $k \le n \le s$. The null hypothesis of homoscedasticity of the vector $\mathbf{a}_t$, is

---

[10]Alternatively, one can apply a multivariate Q test. Nevertheless, the implementation of the multivariate Q test would require even a larger number of the degrees of freedom due to the fact that the Q test is based on the autocorrelation structure, whereas the ARCH test on the partial autocorrelation structure. Both tests are, however, asymptotically equivalent, see Granger and Teräsvirta (1993, Ch. 6).

given by: $H_0 : \mathbf{C}_1 = \mathbf{0}$ versus $H_1 : \mathbf{C}_1 \neq \mathbf{0}$. The appropriate LR-based test statistic is given by

$$MARCH = (T - \tau)(\log(|\mathbf{\Sigma}_r|) - \log(|\mathbf{\Sigma}_u|)) \xrightarrow{d} \chi^2(nk), \tag{9}$$

where $|\cdot|$ denotes the determinant of a square matrix, T stands for the sample size, $\tau = (k + n + 1)/2$ is a small sample correction term.

Although principal component analysis can reduce the dimensionality problem, in practice, there is still an ultimate question of how many components to retain. Details about a principal component analysis and various stopping rules for determining the number of principal components are discussed in the Section 2.4.

2.4 PRINCIPAL COMPONENT ANALYSIS

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

$$w_{jt} = \mathbf{e}_j' \mathbf{v}_t, \quad \text{for} \quad j = 1, \ldots, s, \qquad t = 1, \ldots, T, \tag{10}$$

where w_{jt} is the jth-principal component at time t, $\mathbf{e}_j$ is a particular eigenvector associated with the eigenvalue λ_j estimated from the variance-covariance or correlation matrix. It is important to point out that there is no one-to-one mapping between the roots calculated from the variance-covariance matrix and correlation matrix. A problem is that, unlike the correlation matrix, the variance-covariance matrix is not scale invariant and, hence, neither the calculated roots. Therefore, comfortable or not, the use of the correlation matrix is often recommended, especially for heterogenous data sets and/or indicators originally measured in different units, see Jackson (1991, 64–65) for details. For this reason, the correlation matrix is used in this paper unless otherwise stated. For instance, $\mathbf{v}_t$ takes the following form for the MTSAY test based on a VAR(P) filter

$$\mathbf{v}_t = \text{vech}(\mathbf{z}_t \otimes \mathbf{z}_t'),$$

where $\mathbf{z}_t = (\mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}')'$ is a vector of predetermined variables. Note that the vector $\mathbf{v}_t$ for the MARCH test is defined in a similar way, see the previous section.

A characteristic feature of PCA is that components are uncorrelated linear combinations of the original variables due to orthogonality of the estimated eigenvectors (i.e. $\mathbf{e}_i' \mathbf{e}_j = 0$ for all $i \neq j$). The advantage of this approach is that principal components basically eliminate

multicollinearity from a testing procedure.[11] Another advantage of this approach is that, at least for testing purposes, no interpretation of the calculated principal components is required, which significantly simplifies the use of PCA. Another interesting property is that $\text{var}(w_j) = \mathbf{e}_j' \text{var}(\mathbf{v}_t) \mathbf{e}_j = \lambda_j$, which immediately implies that

$$\sum_{j=1}^{s} \text{var}(w_j) = \sum_{j=1}^{s} \text{var}(v_j) = \sum_{j=1}^{s} \lambda_j. \tag{11}$$

It is clear from (11) that s components are required to reproduce the total variability of the original data set. In practice, however, most of the variability can be accounted for by just a small number of the first components, say n.

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

We omit all the statistical rules since they are problematic in the context of time series. For example, let us consider a Bartlett test of the equivalence of the last $s - n$ roots calculated from the variance-covariance matrix. There are three shortcomings of the Bartlett test. The main disadvantage is that the limiting distribution of Bartlett test is based on the assumption of multivariate normality of the original observations in the vector $\mathbf{v}_t$, see Jolliffe (2005, p. 53–54), which is rather difficult to justify in the context of a time series analysis and/or non-linearity testing. Another disadvantage is that the limiting distribution of the test is no longer χ^2, provided that the roots are calculated from the correlation matrix, see Jackson (1991, p. 99–101) for details. Finally, due to the problematic empirical distribution of the test, the Bartlett test overestimates the actual number of components in practice. Of course, there are some other formal statistical rules to determine the number of components to retain, many of them, unfortunately, suffer from similar problems like the Bartlett test, see Jolliffe (2005, p. 118–126) for a discussion.

We also skip useful, yet relatively subjective, graphical methods such as a scree plot. A scree plot is a figure depicting the sample eigenvalues plotted in descending order against the order number. Provided that just a few first components dominate in magnitude and the rest of

---

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

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

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

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

2. The variance rule: Another popular way of selecting the number of principal components is to use the first n components attributing $100\gamma\%$ of total variance of the original set of variables. The usually recommended proportion of total variance recommended in multivariate analysis is $\gamma = 0.9$.

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

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

Nevertheless, the use of PCA in time series is not without problems. First, PCA is actually built on the independence assumption of the original variables (i.e. variables in the vector $\mathbf{v}_t$) over time. Obviously, this assumption is not fully satisfied even under the null hypothesis since we allow for correlation among model innovations, see Assumption 1 for details. However, if PCA is used entirely just for descriptive purposes, not inferential, then weak dependence and/or other non-IID features in the original vector $\mathbf{v}_t$ do not seriously affect the main objective, see Jolliffe (2005, Ch.12) and Jackson (1991, Ch. 4) for details. Second, PCA critically depends on the properties of the variance-covariance or correlation structure of the original observations. This assumption might be problematic as well. It is nowadays well known that non-linear feature can be to some extent interchanged with outliers and/or structural breaks, see Koop and Potter

(2001) or Carrasco (2002). In such a case, the standard Pearson correlation matrix might not be the best choice and some robust way of calculating the correlation matrix might be more appropriate. Although there exist many robust alternatives, see Croux and Haesbroeck (2000), Ma and Genton (2001), or Hubert et al. (2005), among others, they are not convenient for non-linearity testing purposes. The main problem is that robust techniques trim or downweight the outlying observations, which might be, however, associated with non-linear nature of a given stochastic process. For this reason, the Spearman rank correlation matrix seems to be a good choice here.

Finally, the following 6 alternative ways to determine the number of principal components n are used for both MTSAY and MARCH tests in this paper: (i) "P-BIC" stands for the test with automatically selected number of principal components n using the BIC approach and the Pearson correlation matrix; (ii) "P-0.9" denotes the test, where n is determined in such a way that the principal components cover at least 90 % of the variability of the original variables; (iii) "P-K" is the test, where n is set using the Kaiser rule with the cut off 1.0 for the eigenvalues calculated from the Pearson correlation matrix; (iv)"S-BIC" stands for the test with automatically selected number of principal components n using the BIC approach and the Spearman correlation matrix; (v) "S-0.9" denotes the test where n is determined in such a way that the principal components cover at least 90 % of the variability of the original variables; (vi) "S-K" is the test where n is set using the Kaiser rule with the cut off 1.0 for the eigenvalues calculated form the Spearman correlation matrix.

3. MONTE CARLO SETUP

The statistical properties of the original and modified multivariate and univariate non-linearity tests are examined using three sets of multivariate stationary data generating processes: (i) M-models: three linear VAR models; (ii) N-models: six non-linear conditional mean models; and finally (iii) H-models: six non-linear conditional heteroscedastic models. All data generating processes (DGPs) considered in this paper are summarized in Tables 3 – 5. Since a large number of multivariate models is considered in this paper, some comments on the selected DGPs are in order: M-models are used to assess the size properties of the univariate and multivariate TSAY and ARCH tests: model M1 represents a diagonal VAR(2) model with relatively high persistence, whereas models M2 and M3 are simple VAR(1) models with different persistence. H-models represent various conditional heteroscedastic and stochastic volatility models: H1 and H2 denote different versions of a diagonal VEC-GARCH model with different source of volatility; H3, H4, and H5 represent different versions of a BEKK-GARCH model also with different source of volatility; finally, H6 is a simple multivariate stochastic volatility model. N-models represent a set of non-linear conditional mean models: N1 – N4 denote different specifications of the first-order Taylor approximation of a non-linear VARMA model with a rather

distinct source of non-linearity; N5 is a vector TAR model, and N6 is a vector MSAR model. Special attention is paid to the following pairs of models: N1 and N2, H1 and H2, H3 and H4, which all have the same functional form, but non-linearity is governed in a slightly different way. The above mentioned pairs of models might be considered as a robustness check of the non-linearity tests. Note that all the selected DGPs are time series models borrowed from the literature, see Ling and Li (1997), Duchesne and Lalancette (2003), and Harvill and Ray (1999), among others. Although the list of multivariate time series models is definitely not exhaustive, we are strongly convinced that all the main classes of non-linear models are included. Note also that parameters of all DGPs are set in such a way that the generated series are stationary and and imply positive-definite variance-covariance matrices (i.e. H models). The moment condition of the tests is by no means easy to check for a given set of multivariate non-linear models. For this reason, the max-sum ratio procedure to check the approximate existence of moments is implemented here, see Embrechts et al. (2011, p. 309).

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

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

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

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

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

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

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

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

Table 3: List of multivariate time series models: part 1

M1:

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

M2:

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

M3:

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

Table 4: List of multivariate time series models: part 2

N1:

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

N2:

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

N3:

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

N4:

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

N5:

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

N6:

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

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

Table 5: List of multivariate time series models: part 3

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

$$\boldsymbol{\epsilon}_t = \sqrt{\mathbf{H}_t}\mathbf{a}_t,$$

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

H1:

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

H2:

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

H3:

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

H4:

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

H5:

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

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

$$\boldsymbol{\epsilon}_t = \mathbf{u}_t \odot \exp\{0.5\mathrm{diag}(\mathbf{H}_t),$$

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

H6:

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

4. MONTE CARLO RESULTS

4.1 SIZE PROPERTIES (M-MODELS)

For each data generating process, each configuration of the variance-covariance matrix Σ, and each sample size T, the average rejection frequency is calculated for both the univariate and multivariate tests as follows

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

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

The size results for MTSAY and MARCH tests are presented in Tables 7 – 8. The results suggest the following: (i) The size of the vast majority of the test statistics lies in the range between 0.03 and 0.07, regardless of the data generating process, the sample size T, or the variance-covariance matrix Σ. It can be concluded that all the univariate and multivariate non-linearity tests are reasonably well sized; (ii) No significant differences in the size are observed either between univariate and multivariate tests. However, differences might get noticeable in large-dimensional systems (i.e. $k > 5$)[12]; (iii) Nevertheless, from the detailed results follow that the size of the univariate ARCH test is slightly upward biased, whereas the size of the univariate TSAY test is slightly downward biased in general.

Special attention is paid to evaluating the multivariate tests with respect to the number of components selected by different stopping rules. For higher clarity, the results are depicted in Figure 1. The figure shows the relationship between the average number of principal components (x-axis) and the standard deviation of the number of principal components (y-axis) calculated over all Monte Carlo repetitions. The results suggest the following: (i) Significant differences in the number of principal components and its variability are observed for individual stopping rules; (ii) The BIC stopping rule is apparently the most robust and efficient rule determining almost a constant number of principal components (i.e. $n \approx 2$), regardless of the data generating process, the sample size T, and the configuration of the variance-covariance matrix Σ; (iii) The Kaiser stopping rule performs significantly better than the variance rule, but worse than the BIC

---

[12]The author is very grateful to Professor Timo Teräsvirta from the Aarhus University for making this point.

Figure 1: Statistical properties of the stopping rules of the multivariate tests: M-models
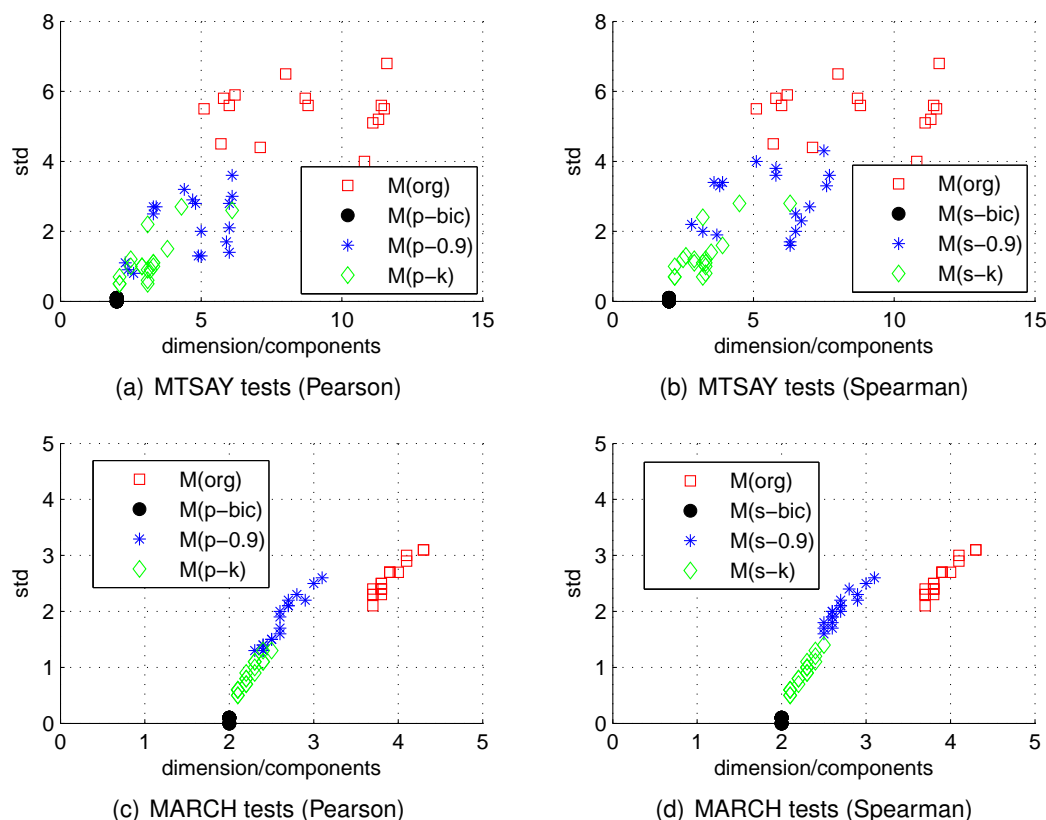
(a) MTSAY tests (Pearson)

(b) MTSAY tests (Spearman)

(c) MARCH tests (Pearson)

(d) MARCH tests (Spearman)

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

rule; (iv) No noticeable differences are observed between the results of the multivariate tests based on the Pearson and Spearman correlation matrices; (v) An interesting difference in the statistical properties of the stopping rules is observed between the MTSAY and MARCH tests. Much higher heterogeneity in the average number of components is observed for the MTSAY tests as compared to the MARCH tests. The reason for that lies in the fact that parameters of the MTSAY tests are directly linked to the parameters of a filter (i.e. the lag order P of a VAR model in our case), whereas the parameters of the MARCH tests lack this link. This interesting property might be also used for further improvements of the multivariate non-linearity tests.[13]

4.2 POWER PROPERTIES (N-MODELS)

The power results of TSAY and ARCH tests for the N-models are presented in Tables 9. Since the ARCH tests have relatively low power against non-linear conditional mean models, the main focus is on the TSAY tests in this section. The ARCH tests results are reported for complete-

---

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

ness only. The results suggest the following: (i) Significant differences are observed in the power of the multivariate and univariate TSAY tests, unless the sample size is sufficiently large (i.e. $T = 1000$); (ii) The power of the univariate TSAY test, applied to individual series, is significantly lower, especially for N2, N3 and N4 model configurations. Put differently, the univariate TSAY test completely fails, provided that non-linearity in the conditional mean is generated by cross-bilinear terms (N2 case), cross-quadratic terms (N3 case), or cross-product terms (N4 case).

Figure 2: Statistical properties of the stopping rules of the multivariate tests: N-models
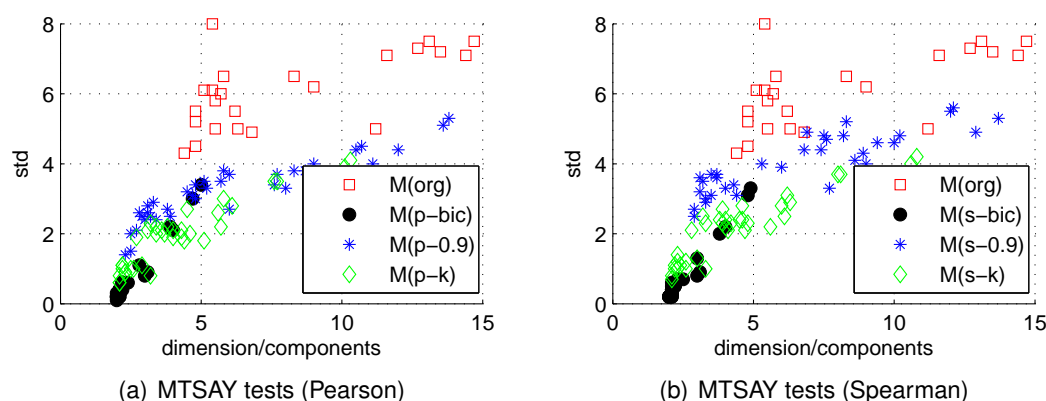
| (a) MTSAY tests (Pearson) | (b) MTSAY tests (Spearman) |

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

Special attention is paid to evaluating the multivariate tests with respect to the number of components selected by different rules. For higher clarity, the results are depicted in Figure 2. The figure depicts the relationship between the average number of principal components (x-axis) and the standard deviation of the number of principal components (y-axis). The results suggest the following: (i) Significant differences in the number of principal components and its variability are observed for individual stopping rules; (ii) The BIC stopping rule is apparently the most robust and efficient rule, regardless of the DGPs, the sample size T, and the variance-covariance matrix Σ; (iii) The Kaiser stopping rule performs slightly better than the variance rule, but worse than the BIC rule; (iii) No differences are observed between the results of the multivariate tests based on the Pearson and Spearman correlation matrices.

## 4.3 POWER PROPERTIES (H-MODELS)

The power results of TSAY and ARCH tests for the H-models are presented in Tables 10. Since the TSAY tests have very low power against conditional heteroscedastic models, the main focus is on the ARCH tests in this section. The TSAY tests results are reported for completeness only. The results suggest the following: (i) Significant differences are observed in the power of the

multivariate and univariate ARCH tests, whereas no differences exist among the multivariate ARCH tests, regardless the DGPs, the sample size T, and the configuration of the variance-covariance matrix Σ; (ii) The power of the univariate ARCH test, applied to individual series, is significantly lower in some cases, especially for H2 and H5 configurations. Put differently, the univariate ARCH test suffers from a serious power distortion, provided that the conditional volatility is generated by cross-heteroscedasticity terms rather than individual autoregressive terms (H2 case).

Figure 3: Power properties of the multivariate tests: H-models
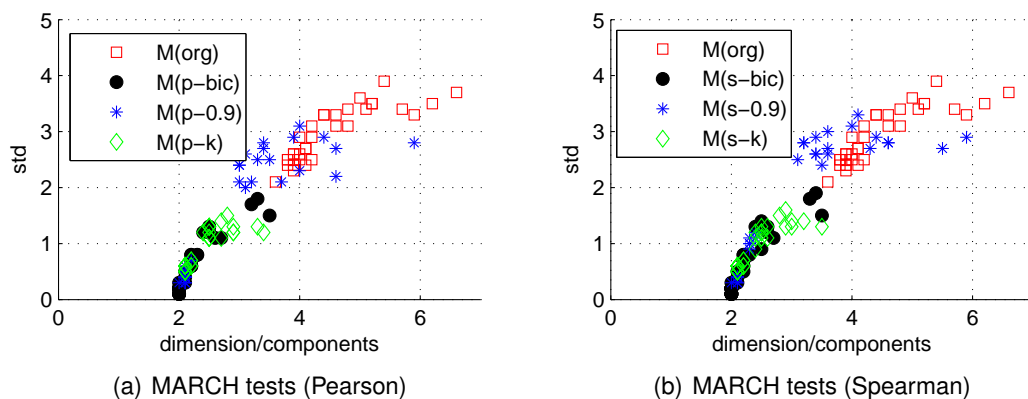
| (a) MARCH tests (Pearson) | (b) MARCH tests (Spearman) |

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

Special attention is paid to evaluating the multivariate tests with respect to the number of components selected by different rules. For higher clarity, the results are depicted in Figure 3 as well. The figure depicts the relationship between the average number of principal components (x-axis) and the standard deviation of the number of principal components (y-axis) calculated over all Monte carlo repetitions. The results suggest the following: (i) In contrast to MTSAY tests results, the average number of principal components of the MARCH tests is much more concentrated, regardless of the DGPs, the sample size T, and the configuration of the variance-covariance matrix Σ. The reason for that is discussed in Section 4.1; (ii) There seems not to be a clear cutoff between the BIC approach and the Kaiser rule in selecting the number of principal components. Both rules produce almost identical results. Keeping in mind the simplicity, the Kaiser rule might be slightly preferred. As in the previous cases, the variance rule is inefficient and produces similar results to the original MARCH test; (iii) No differences are observed between the results of the multivariate tests based on the Pearson and Spearman correlation matrices.

5. EMPIRICAL EXAMPLE

Having analyzed the behaviour of the proposed tests using extensive Monte Carlo experiments, we can turn our attention to applying the new principal component-based multivariate non-linearity tests to economic indicators. Following Cho and Moreno (2003), a simple rational expectations model with New Kayenesian features (e.g. the habit formation, the price index-ation, etc) is considered in this section. The model describes the behaviour of three agents in the economy (i.e. households, firms, and government). It is assumed that the economy is populated by infinitely lived households who consume and supply labour to firms. Households are assumed to maximize an intertemporal (constant relative risk aversion) utility function sub-ject to a budget constraint. Firms, completely owned by households, produce output using a simple production function. The log-linearized equilibrium (first-order) conditions can be written as follows

$$\pi_t = \bar{\pi} + \mu_1 \mathbb{E}_t(\pi_{t+1}) + (1 - \mu_1)\pi_{t-1} + \kappa y_t + x_t^\pi,$$

$$y_t = \mu_2 \mathbb{E}_t(y_{t+1}) + (1 - \mu_2)y_{t-1} - \omega^{-1}(r_t - \mathbb{E}_t(\pi_{t+1})) + x_t^y,$$

$$r_t = \bar{r} + \phi\pi_t + \gamma y_t + x_t^r,$$

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

$$\mathbf{G}\mathbf{y}_t = \mathbf{c} + \mathbf{F}\mathbb{E}_t(\mathbf{y}_{t+1}) + \mathbf{H}\mathbf{y}_{t-1} + \mathbf{L}\mathbf{x}_t,$$

$$\mathbf{x}_t = \mathbf{N}\mathbf{x}_{t-1} + \mathbf{a}_t,$$

where $\mathbf{y}_t = (\pi_t, y_t, r_t)'$ is a vector of dependent variables and $\mathbf{x}_t = (x_t^\pi, x_t^y, x_t^r)'$ is a vector of (unobserved) exogenous variables. Uhlig (1995) proposed a simple solution of the model based on undetermined coefficients. It can be shown that the unobserved (latent) shock variables may be substituted out of the system. The resulting representation is then a simple 3-variate VAR(2) model given by

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{y}_{t-2} + \boldsymbol{\Theta}\mathbf{a}_t. \tag{14}$$

Three quarterly US economic indicators are considered for $\mathbf{y}$: the output variable y is approxi-mated by the growth rate of the real US GDP series (GDP); the inflation rate π is approximated by the US CPI inflation rate (CPI), and the policy rate r by the 3M treasure bill rate (IR). The economic time series span the period 1961Q1 − 2005Q4.[14] The series are depicted in Fig-ure 4. The appropriate filter for the model in (14) seems to be a VAR(2) model. However, it is

---

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

important to emphasize that the order of the filter is affected by an ad-hoc decision that shock variables follow an AR(1) process. This assumption may not be correct in general. For this reason, an automatic lag order selection procedure to determine the lag order of a VAR filter is implemented as well. The procedure indicates $P = 4$ (rather than $P = 2$) and $Q = 2$.

Figure 4: US macroeconomic time series
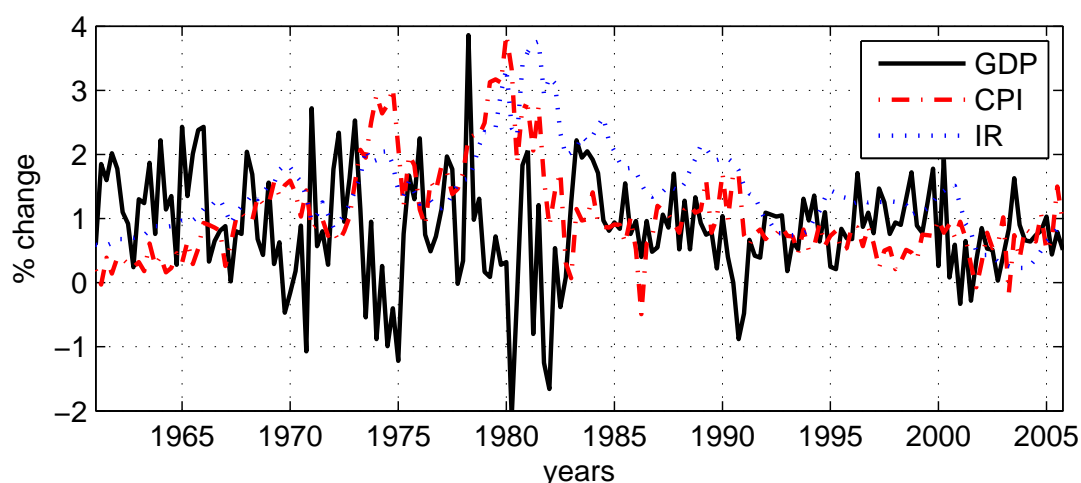
The results are presented in Table 6. These suggest the following: (i) The null hypothesis is rejected by almost all multivariate non-linearity tests even at the significance level 0.10, regardless of the estimation of the correlation matrix or the determination of the number of principal components; (ii) The automatically selected number of components varies from 3 to 8, depending on the test, the correlation matrix and the stopping rule. For example, the BIC rule indicates to retain 3 principal components for the MTSAY(P-BIC) test, which is a tremendous dimensionality reduction as compared to the original MTSAY test requiring 78 additional variables; (iii) The univariate and multivariate tests produce ambiguous results. For example, the univariate TSAY test does reject the null hypothesis of linearity for both the inflation rate (CPI) and the nominal interest rate (IT) but not for the GDP growth rates (GDP) at the significance level 0.10, whereas the null is rejected by all configurations of the multivariate TSAY test.

The above results are based on the whole sample consisting of $T = 180$ observations. It might be also interesting to assess the robustness of the tests against the sample size. For this purpose, a rolling-window approach is applied to check the stability of the tests. The method is based on splitting the original sample into 81 overlapping sub-samples. Each sub-sample consists of only 100 consecutive observations. It means that the first sub-sample (window) span the period 1961Q1 − 1985Q4, the second one 1961Q2 − 1985Q2 etc. The results are depicted in a graphical form in Figures 5 − 6. They suggest the following: (i) We can observe much higher instability of the univariate tests in sub-periods as compared to the multivariate tests; (ii) In both cases, the principal component based multivariate tests do provide a significant dimensionality

Table 6: P-values of the non-linearity tests

filter	univariate tests			multivariate tests	
	AR(2)	AR(4)	AR(6)	VAR(4)	
tests/variables	GDP	CPI	IR	$\mathbf{y} =$(CPI,GDP,IR)	n
TSAY	0.914	0.000	0.000	–	–
MTSAY(org)	–	–	–	0.000	78
MTSAY(P-BIC)	–	–	–	0.026	3
MTSAY(P-0.9)	–	–	–	0.013	7
MTSAY(P-k)	–	–	–	0.005	9
MTSAY(S-bic)	–	–	–	0.057	3
MTSAY(S-0.9)	–	–	–	0.001	6
MTSAY(S-k)	–	–	–	0.001	8
ARCH	0.037	0.060	0.000	–	–
MARCH(org)	–	–	–	0.000	12
MARCH(P-bic)	–	–	–	0.005	4
MARCH(P-0.9)	–	–	–	0.002	8
MARCH(P-K)	–	–	–	0.001	5
MARCH(S-BIC)	–	–	–	0.244	3
MARCH(S-0.9)	–	–	–	0.000	9
MARCH(S-K)	–	–	–	0.062	4

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

[b] "M(org)" denotes the original multivariate TSAY test, "M(P-)" denotes the multivariate test based on the Pearson correlation matrix, whereas "M(S-)" on the Spearman correlation matrix. "M(-BIC)" denotes the multivariate test with the number of principal components selected by the BIC approach, whereas "M(-0.9)" by the variance rule with the cut-off variance 0.9, and "M(-K)" by the Kaiser rule.

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

6. CONCLUSION

Our Monte Carlo results confirm that univariate tests might not be adequate for testing non-linearity in multivariate processes. The univariate tests can suffer from a serious power distortion, and, thus, lead to misleading inference. In contrast, all the modified principal component-based non-linearity tests do exhibit very good size and power properties. The statistical properties of the multivariate tests are robust against: (i) The configuration of the variance-covariance

structure of model innovations; (ii) The correlation matrix used for calculating the roots. In addition, the principal component-based tests do offer a remarkable dimensionality reduction (in average about 70 %) without any systematic power distortion. Nevertheless, special care should be exercised to stopping rules determining the number of components. The Monte Carlo results suggest that the BIC rule performs best as compared to the variance rule and/or the Kaiser rule, although the results are case dependent to some extent.

REFERENCES

Adolfson, M., S. Laséen, J. Lindé, and M. Villani (2007). Bayesian estimation of an open economy DSGE model with incomplete pass-through. *Journal of International Economics 72*(2), 481–511.

Adolfson, M., S. Laséen, J. Lindé, and M. Villani (2008). Empirical Properties Of Closed-And Open-Economy Dsge Models Of The Euro Area. *Macroeconomic Dynamics 12*, 2–19.

Adolfson, M., S. Laséen, J. Lindé, and M. Villani (2008). Evaluating an estimated new keynesian small open economy model. *Journal of Economic Dynamics and Control 32*(8), 2690–2721.

Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley.

Bauwens, L., S. Laurent, and J. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics 21*, 79–109.

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

Carrasco, M. (2002). Misspecified structural change, threshold, and markov-switching models. *Journal of econometrics 109*(2), 239–273.

Cho, S. and A. Moreno (2003). A structural estimation and interpretation of the New Keynesian macro model. *Columbia University Working Paper 14*.

Croux, C. and G. Haesbroeck (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika 87*(3), 603–618.

Davidson, R. and J. G. MacKinnon (1999). *Econometric Theory and Methods*. Oxford University Press.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007). On the fit of new keynesian models. *Journal of Business & Economic Statistics 25*(2), 123–143.

Deschamps, P. (1993). Joint tests for regularity and autocorrelation in allocation systems. *Journal of Applied Econometrics 8*(2), 195–211.

Deschamps, P. (1996). Monte carlo methodology for lm and lr autocorrelation tests in multivariate regression. *Annales d'Economie et de Statistique*, 149–169.

Duchesne, P. and S. Lalancette (2003). On testing for multivariate arch effects in vector time series models. *The Canadian Journal of Statistics 31*(3), 275–292.

Embrechts, P., C. Klüppelberg, and T. Mikosch (2011). *Modelling extremal events: for insurance and finance*, Volume 33. Springer.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*(4), 987–1007.

Engle, R. and F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*, 122–150.

Godfrey, L. G. (1988). *Misspecification tests in econometrics*. Cambridge University Press.

Granger, C. and T. Teräsvirta (1993). *Modelling Nonlinear Economic Relationships*. Oxford University Press.

Harvill, J. and B. Ray (1999). Testing for nonlinearity in a vector time series. *Biometrika 89*(3), 728–834.

Hubert, M., P. Rousseeuw, and K. Branden (2005). Robpca: a new approach to robust principal component analysis. *Technometrics 47*(1), 64–79.

Izenman, A. (2008). *Modern Multivariate Statistical Techniques*. Springer.

Jackson, J. (1991). *A User's Guide to Principal Components*. Wiley.

Johnson, R. and D. Wichern (2007). *Applied Multivariate Statistical Analysis*. Prentice Hall.

Jolliffe, I. (2005). *Principal Component Analysis*. Wiley.

Koop, G. and S. Potter (2001). Are apparent findings of nonlinearity due to structural instability in economic time series? *The Econometrics Journal 4*(1), 37–55.

Lee, J. (1991). A Lagrange multiplier test for GARCH models. *Economics Letters 37*(3), 265–271.

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.

Ling, S. and W. Li (1997). Diagnostic checking of non-linear multivariate time series with multivariate arch errors. *Journal of Time Series Analysis 18*(5), 447–464.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

Ma, Y. and M. Genton (2001). Highly robust estimation of dispersion matrices. *Journal of Multivariate Analysis 78*(1), 11–36.

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.

Peres-Neto, P., D. Jackson, and K. Somers (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis 49*(4), 974–997.

Rudebusch, G. and E. Swanson (2008). Examining the bond premium puzzle with a dsge model. *Journal of Monetary Economics 55*, 111–126.

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

Sims, C. and T. Zha (2006). Were there regime switches in us monetary policy? *The American Economic Review 96*(1), 54–81.

Smets, F. and R. Wouters (2002). Openness, imperfect exchange rate pass-through and monetary policy. *Journal of Monetary Economics 49*(5), 947–981.

Smets, F. and R. Wouters (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association 1*(5), 1123–1175.

Smets, F. and R. Wouters (2007). Shocks and Frictions in US Business Cycles: a Bayesian DSGE Approach. *American Economic Review 97*(3), 586–606.

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.

Uhlig, H. (1995). A toolkit for analyzing nonlinear dynamic stochastic models easily. *Federal Reserve Bank of Mineapolis, Discussion Paper* (101).

7. APPENDIX A: TABLES

Table 7: Size properties of the TSAY tests: M-models

matrix	test	T=200			T=500			T=1000		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
Σ_1	M(org)	0.059	0.040	0.051	0.034	0.055	0.044	0.059	0.046	0.044
	M(P-BIC)	0.057	0.037	0.050	0.045	0.045	0.047	0.048	0.052	0.039
	M(P-0.9)	0.042	0.042	0.047	0.041	0.041	0.040	0.042	0.044	0.039
	M(P-K)	0.053	0.042	0.046	0.047	0.047	0.051	0.050	0.052	0.034
	M(S-BIC)	0.048	0.045	0.056	0.051	0.048	0.043	0.037	0.051	0.032
	M(S-0.9)	0.049	0.043	0.047	0.042	0.052	0.038	0.043	0.039	0.044
	M(S-K)	0.048	0.045	0.051	0.051	0.043	0.049	0.052	0.056	0.037
	X_1	0.041	0.050	0.041	0.038	0.064	0.053	0.046	0.051	0.041
	X_2	0.037	0.046	0.070	0.050	0.048	0.060	0.056	0.061	0.047
Σ_2	M(org)	0.058	0.041	0.052	0.057	0.041	0.046	0.053	0.056	0.050
	M(P-BIC)	0.053	0.053	0.057	0.062	0.042	0.044	0.054	0.044	0.047
	M(P-0.9)	0.042	0.049	0.053	0.061	0.044	0.043	0.050	0.045	0.045
	M(P-K)	0.034	0.048	0.052	0.064	0.042	0.042	0.052	0.042	0.040
	M(S-BIC)	0.051	0.051	0.051	0.055	0.044	0.044	0.051	0.049	0.051
	M(S-0.9)	0.042	0.042	0.057	0.052	0.048	0.039	0.048	0.044	0.039
	M(S-K)	0.039	0.049	0.049	0.061	0.045	0.037	0.049	0.046	0.037
	X_1	0.034	0.049	0.058	0.055	0.058	0.066	0.039	0.054	0.054
	X_2	0.021	0.050	0.053	0.045	0.054	0.054	0.052	0.057	0.058

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

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

Table 8: Size properties of the ARCH tests: M-models

matrix	test	T=200			T=500			T=1000		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
Σ_1	M(org)	0.063	0.067	0.062	0.050	0.059	0.055	0.068	0.053	0.050
	M(P-BIC)	0.066	0.070	0.067	0.047	0.051	0.058	0.048	0.059	0.061
	M(P-0.9)	0.062	0.072	0.063	0.050	0.057	0.057	0.057	0.052	0.060
	M(P-K)	0.064	0.074	0.063	0.052	0.053	0.057	0.050	0.056	0.061
	M(S-BIC)	0.066	0.072	0.072	0.043	0.047	0.064	0.051	0.054	0.063
	M(S-0.9)	0.062	0.070	0.065	0.047	0.051	0.061	0.054	0.056	0.065
	M(S-K)	0.056	0.070	0.061	0.045	0.046	0.063	0.051	0.057	0.065
	X_1	0.049	0.055	0.069	0.054	0.049	0.048	0.052	0.044	0.059
	X_2	0.047	0.050	0.064	0.056	0.061	0.058	0.055	0.057	0.066
Σ_2	M(org)	0.050	0.068	0.061	0.056	0.050	0.058	0.065	0.046	0.045
	M(P-BIC)	0.050	0.063	0.068	0.060	0.055	0.054	0.060	0.047	0.056
	M(P-0.9)	0.052	0.066	0.057	0.055	0.053	0.051	0.059	0.048	0.056
	M(P-K)	0.049	0.060	0.057	0.055	0.049	0.049	0.056	0.046	0.051
	M(S-BIC)	0.052	0.066	0.068	0.059	0.053	0.052	0.062	0.050	0.055
	M(S-0.9)	0.049	0.062	0.060	0.058	0.054	0.052	0.061	0.049	0.054
	M(S-K)	0.049	0.059	0.056	0.057	0.047	0.048	0.058	0.049	0.051
	X_1	0.051	0.050	0.046	0.052	0.057	0.044	0.063	0.056	0.065
	X_2	0.058	0.042	0.051	0.052	0.054	0.049	0.052	0.047	0.061

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

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

Table 9: Power properties of the TSAY tests: N-models

matrix	test	T=200						T=500						T=1000					
		N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6	N1	N2	N3	N4	N5	N6
Σ_1	M(org)	0.34	0.36	0.81	0.47	1.00	0.49	0.76	0.77	0.99	0.87	1.00	0.53	0.98	0.97	1.00	0.99	1.00	0.58
	M(P-BIC)	0.15	0.20	0.81	0.47	0.99	0.33	0.28	0.33	0.98	0.84	1.00	0.34	0.42	0.40	1.00	0.97	1.00	0.35
	M(P-0.9)	0.29	0.30	0.81	0.46	1.00	0.36	0.66	0.69	0.99	0.88	1.00	0.41	0.96	0.94	1.00	0.99	1.00	0.45
	M(P-K)	0.23	0.23	0.81	0.44	0.99	0.35	0.49	0.53	0.99	0.80	1.00	0.38	0.80	0.77	1.00	0.97	1.00	0.41
	M(S-BIC)	0.16	0.17	0.81	0.47	0.99	0.30	0.26	0.30	0.98	0.84	1.00	0.32	0.38	0.38	1.00	0.98	1.00	0.34
	M(S-0.9)	0.34	0.34	0.82	0.46	1.00	0.40	0.74	0.74	0.99	0.87	1.00	0.44	0.98	0.97	1.00	0.99	1.00	0.49
	M(S-K)	0.23	0.23	0.81	0.44	1.00	0.35	0.47	0.53	0.98	0.81	1.00	0.38	0.78	0.77	1.00	0.98	1.00	0.41
	X_1	0.34	0.06	0.07	0.10	0.99	0.21	0.72	0.06	0.06	0.17	1.00	0.25	0.94	0.09	0.07	0.32	1.00	0.27
	X_2	0.26	0.06	0.05	0.10	0.32	0.38	0.63	0.11	0.05	0.19	0.53	0.43	0.92	0.14	0.06	0.31	0.74	0.48
Σ_2	M(org)	0.39	0.40	0.81	0.25	1.00	0.49	0.83	0.81	0.99	0.54	1.00	0.51	0.99	0.99	1.00	0.80	1.00	0.59
	M(P-BIC)	0.24	0.22	0.79	0.23	1.00	0.37	0.38	0.38	0.96	0.43	1.00	0.33	0.33	0.37	1.00	0.56	1.00	0.36
	M(P-0.9)	0.34	0.36	0.81	0.23	1.00	0.39	0.73	0.73	0.99	0.51	1.00	0.40	0.97	0.96	1.00	0.82	1.00	0.46
	M(P-K)	0.33	0.32	0.81	0.23	1.00	0.36	0.73	0.73	0.99	0.51	1.00	0.36	0.97	0.96	1.00	0.83	1.00	0.42
	M(S-BIC)	0.21	0.20	0.78	0.24	1.00	0.34	0.32	0.31	0.93	0.46	1.00	0.32	0.24	0.25	0.98	0.65	1.00	0.36
	M(S-0.9)	0.35	0.35	0.81	0.23	1.00	0.41	0.74	0.75	0.99	0.52	1.00	0.43	0.98	0.97	1.00	0.84	1.00	0.48
	M(S-K)	0.32	0.32	0.81	0.22	1.00	0.35	0.73	0.71	0.99	0.51	1.00	0.36	0.97	0.96	1.00	0.84	1.00	0.42
	X_1	0.33	0.05	0.06	0.09	0.99	0.19	0.69	0.06	0.07	0.13	1.00	0.24	0.91	0.15	0.12	0.29	1.00	0.26
	X_2	0.29	0.07	0.06	0.10	0.98	0.41	0.62	0.11	0.08	0.14	1.00	0.43	0.91	0.21	0.12	0.29	1.00	0.51

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

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

Table 10: Power properties of the ARCH tests: H-models

| matrix	test	T=200						T=500						T=1000					
		H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6
Σ_1	M(org)	0.74	0.72	0.36	0.34	0.41	0.38	0.99	0.98	0.72	0.69	0.81	0.62	1.00	1.00	0.94	0.96	0.98	0.82
	M(P-BIC)	0.72	0.67	0.40	0.40	0.47	0.35	0.98	0.95	0.76	0.74	0.85	0.59	1.00	1.00	0.95	0.98	0.99	0.81
	M(P-0.9)	0.73	0.69	0.40	0.39	0.47	0.34	0.97	0.97	0.76	0.74	0.85	0.58	1.00	1.00	0.95	0.98	0.99	0.81
	M(P-K)	0.72	0.67	0.40	0.40	0.47	0.32	0.97	0.96	0.76	0.74	0.85	0.58	1.00	1.00	0.96	0.98	0.99	0.81
	M(S-BIC)	0.72	0.70	0.40	0.40	0.47	0.37	0.98	0.96	0.76	0.74	0.85	0.60	1.00	1.00	0.95	0.98	0.99	0.82
	M(S-0.9)	0.71	0.70	0.39	0.39	0.47	0.36	0.97	0.96	0.76	0.75	0.85	0.60	1.00	1.00	0.96	0.98	0.99	0.82
	M(S-K)	0.70	0.68	0.40	0.40	0.47	0.34	0.97	0.96	0.76	0.75	0.85	0.59	1.00	1.00	0.96	0.98	0.99	0.82
	X_1	0.49	0.16	0.38	0.35	0.11	0.24	0.89	0.28	0.74	0.68	0.21	0.44	0.99	0.42	0.95	0.94	0.32	0.65
	X_2	0.54	0.17	0.38	0.38	0.28	0.25	0.86	0.30	0.76	0.69	0.56	0.44	0.99	0.39	0.94	0.93	0.82	0.63
Σ_2	M(org)	0.72	0.72	0.60	0.60	0.87	0.54	0.99	0.98	0.94	0.96	1.00	0.74	1.00	1.00	1.00	1.00	1.00	0.94
	M(P-BIC)	0.74	0.72	0.66	0.65	0.91	0.48	0.99	0.98	0.96	0.97	1.00	0.73	1.00	1.00	1.00	1.00	1.00	0.94
	M(P-0.9)	0.76	0.74	0.66	0.65	0.91	0.47	0.99	0.98	0.97	0.97	1.00	0.73	1.00	1.00	1.00	1.00	1.00	0.94
	M(P-K)	0.74	0.73	0.66	0.65	0.91	0.46	0.98	0.98	0.97	0.97	1.00	0.72	1.00	1.00	1.00	1.00	1.00	0.93
	M(S-BIC)	0.74	0.73	0.66	0.65	0.91	0.51	0.99	0.98	0.96	0.97	1.00	0.75	1.00	1.00	1.00	1.00	1.00	0.95
	M(S-0.9)	0.76	0.75	0.66	0.65	0.91	0.50	0.99	0.98	0.96	0.97	1.00	0.74	1.00	1.00	1.00	1.00	1.00	0.95
	M(S-K)	0.74	0.73	0.65	0.65	0.91	0.49	0.99	0.98	0.97	0.97	1.00	0.74	1.00	1.00	1.00	1.00	1.00	0.95
	X_1	0.70	0.22	0.66	0.60	0.19	0.33	0.98	0.36	0.95	0.95	0.38	0.52	1.00	0.56	1.00	1.00	0.61	0.74
	X_2	0.32	0.24	0.72	0.72	0.47	0.35	0.64	0.39	0.98	0.97	0.85	0.60	0.90	0.66	1.00	1.00	0.99	0.85

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

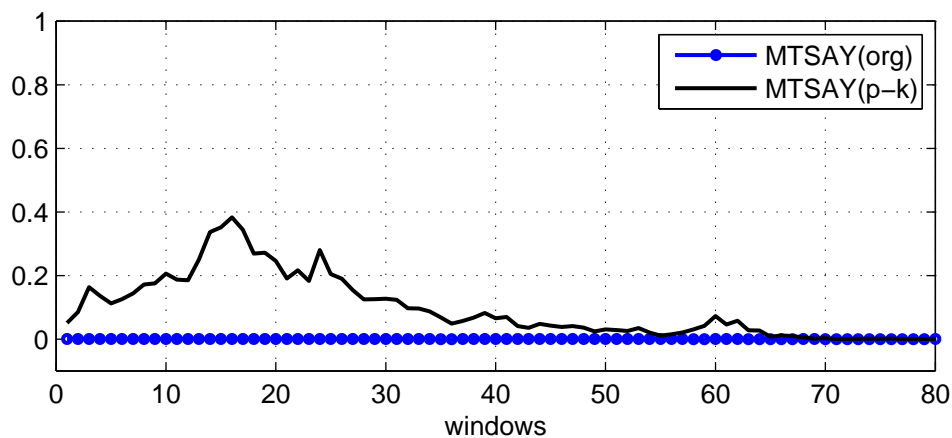[b] The significance level is set to $\alpha = 0.05$.
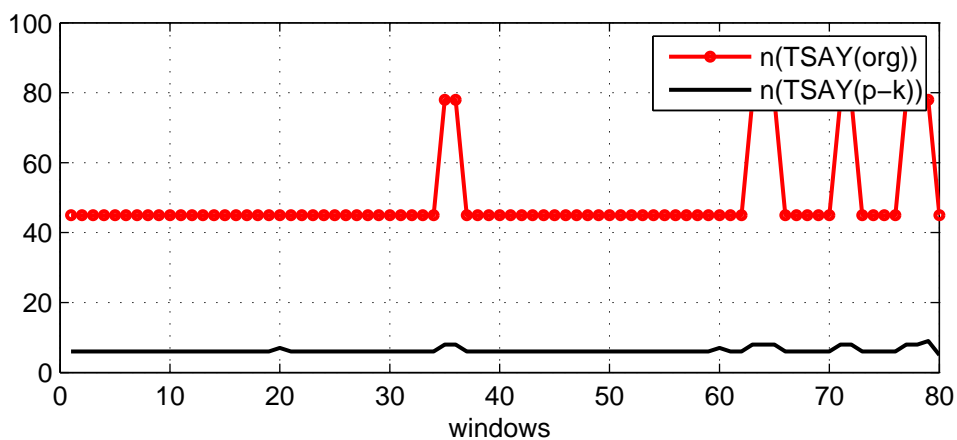
8. APPENDIX B: FIGURES

Figure 5: Empirical results of the TSAY tests: 25 year rolling windows
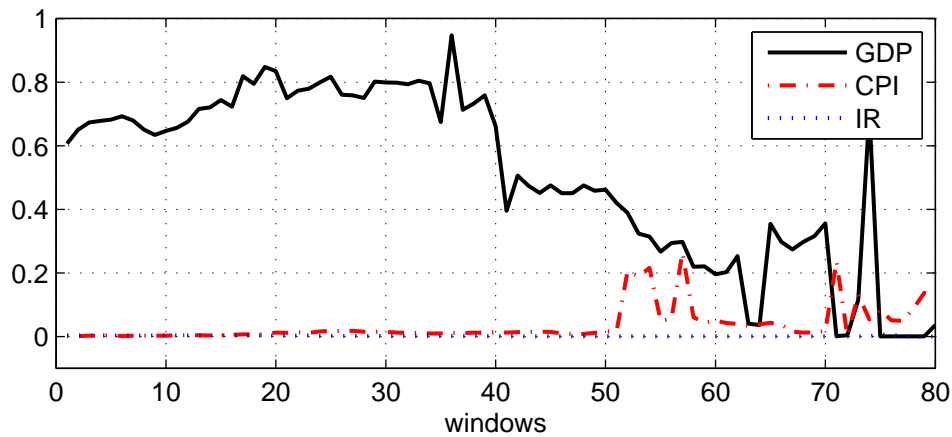
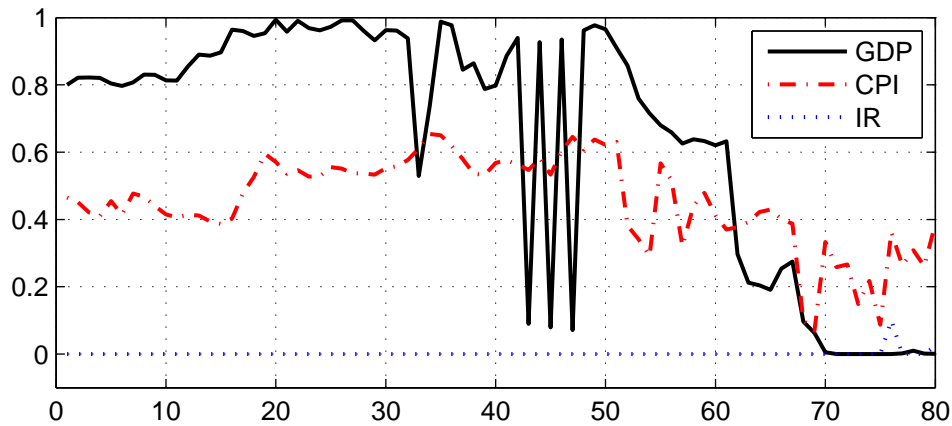(a) p-values of the univariate TSAY tests

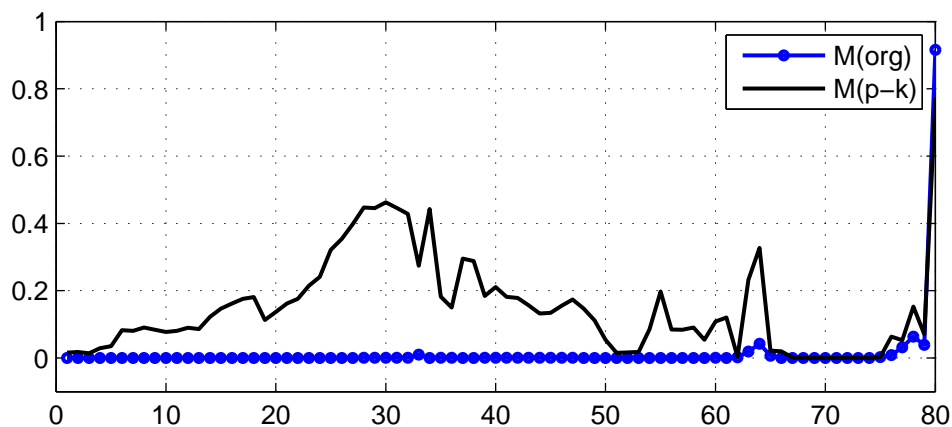

(b) p-values of the multivariate TSAY tests

(c) dimension/number of components of the multivariate TSAY tests

Note: "M(org)" denotes the average rejection frequency of the original multivariate TSAY test, "M(P-BIC)" stands for the multivariate TSAY test based on the automatically selected number of principal components, whereas "GDP", "CPI", and "IR" denote the results obtained from univariate TSAY tests for individual series. "NBER" stands for a fraction of the recession observations in a particular 25-year window.
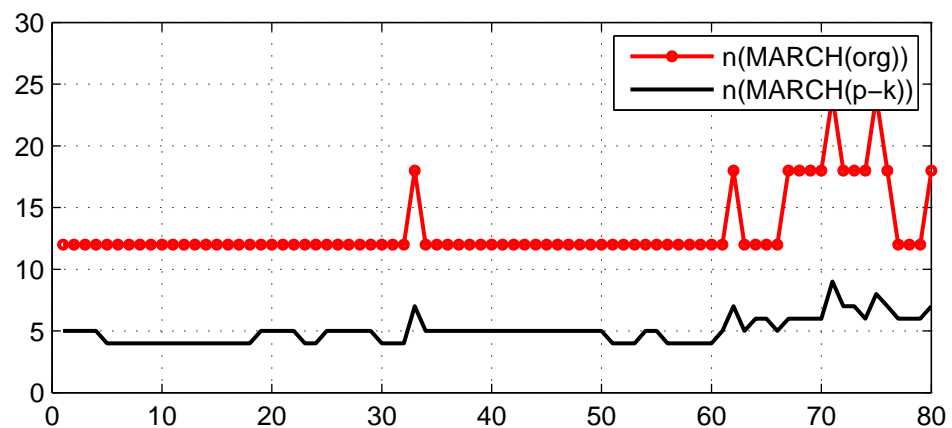
Figure 6: Empirical results of the ARCH tests: 25 year rolling windows

(a) p-values of the univariate ARCH tests

(b) p-values of the multivariate ARCH tests

(c) dimension/number of components of the multivariate ARCH tests

Note: 'M(org)" denotes the average rejection frequency of the original multivariate ARCH test, "M(P-BIC)" stands for the multivariate ARCH test based on the automatically selected number of principal components, whereas "GDP", "CPI", and "IR" denote the results obtained from univariate ARCH tests for individual series. "NBER" stands for a fraction of the recession observations in a particular 25-year window.