Electronic Journal of Statistics
Vol. 5 (2011) 507–533
ISSN: 1935-7524
DOI: 10.1214/11-EJS617

Testing linear causality in mean when the number of estimated parameters is high

Hamdi Raïssi

EQUIPPE-GREMARS, Université Lille 3
and
IRMAR-INSA Rennes
20, avenue des buttes de Coësmes, CS 14315, F-35043 Rennes Cedex, France
e-mail: hamdi.raissi@insa-rennes.fr

Abstract: This paper investigates the problem of testing for linear Granger causality in mean when the number of parameters is high with the possible presence of nonlinear dynamics. Dependent innovations are taken into account by considering tests which asymptotic distributions is a weighted sum of chi-squares and tests with modified weight matrices. Wald, Lagrange Multiplier (LM) and Likelihood Ratio (LR) tests for linear causality in mean are studied. It is found that the LM tests based on restricted estimators significantly improve the analysis of linear Granger causality in mean relations when the dimension is high or when the autoregressive order is large. We also see that the tests based on a modified asymptotic distribution have a better control of the error of first kind when compared to the tests with modified statistic in finite samples. An application to international finance data is proposed to illustrate the robustness to the presence of nonlinearities of the studied tests.

AMS 2000 subject classifications: Primary 62M10; secondary 91B84.
Keywords and phrases: Linear causality in mean, VAR models, high dimensional processes, large autoregressive order, causality in variance, weak errors.

Received May 2010.

Contents

1. Introduction

Since the paper by Granger [18], the use of linear causality in mean to study relations between subsets of variables is a common practice (see [33, 35], or [19] among other references). This can be explained by the fact that linear causality in mean is based on linear predictors, and can be easily tested by considering tests of zero restrictions on the parameters of VAR models with iid innovations. However when investigating linear causal relations, researchers often consider a large number of lags to correctly take into account the linear temporal dynamics of the variables. For instance [4, 34] or [14] use large autoregressive orders to investigate the money-income causality relation. Note also that linear causality in mean tests are often carried out in a bivariate framework. Nevertheless omitting relevant variables from the model can lead to erroneous conclusions (see e.g. [28]). Then researchers sometimes consider VAR models of higher dimension for the test of linear causality in mean (see e.g. [9, 34] and references therein). In this paper it is shown that the standard Wald test for linear causality in mean perform poorly in these cases. We propose Lagrange Multiplier (LM) tests, based on restricted estimators, which give significant improvements for the analysis of linear Granger causality in mean relations when a large number have to be estimated.

Processes with dependent innovations can arise from many situations (see e.g. [6, 2] or [25]). [20] tested linear Granger causality in mean between stock returns and percentage volume changes with evidence of nonlinear causality in the residuals. In addition note that numerous models produce processes with conditional heteroscedasticity, as for instance hidden Markov, all-pass or GARCH models (see e.g. [1])). In such cases it is common to use Wald tests with corrected statistics. If the error process is assumed to be a martingale difference the White [38] correction matrix is used. When the innovation process is only assumed uncorrelated a Heteroscedasticity Autocorrelation Consistent (HAC) weight matrix is used. However [3] pointed out that using HAC estimators in the tests statistics may lead to over-rejections when the (linear) temporal dependence of the series is strongly marked. The test for linear Granger causality in mean in presence of nonlinearities has become of increased interest. For instance linear causality in mean is studied by [37] in the case of Markov Switching VAR models. Vilasuso [36] studied the test of linear causality in mean in presence of causality in variance, and considered tests with corrected statistics by considering HAC and [38] covariance estimation methods. An important output of this work is that, in certain cases, these tests may suffer from severe size distortion in presence of causality in variance. This effect of nonlinear dynamics on the test of linear causality in mean is also raised in [11]. In this paper tests are built in the context of weak VAR models, i.e. VAR models with errors only assumed uncorrelated to take into account a wide range of nonlinearities. More precisely we use the asymptotic normality of the Quasi Maximum Likelihood Estimator (QMLE) obtained in [15]. Different approaches for testing the linear Granger causality in mean when the number of estimated parameters is high are proposed and compared in various situations. The tests are developed assuming no particular structure for the error process, and are then quite general. It is found that the use of the tests with data driven critical values are preferable when a high number of parameters are estimated in case of dependent errors.

The structure of the paper is as follows. In Section 2 the test of linear causality in mean in the weak VAR framework is discussed. In Section 3 the QMLE is derived and its asymptotic behaviour is stated. Modified tests for linear causality in mean are proposed in Section 4. In Section 5 Monte Carlo experiments are performed. We study the linear causality in mean from the daily log returns of the exchange rate of U.S. Dollars to one British Pound (USD/BP hereafter) to the daily log returns of the exchange rate of U.S. Dollars to one New Zealand Dollar (USD/NZD hereafter) using the standard and modified tests in Section 6. The proofs are relegated to the appendix.

The following notations will be used throughout in the paper. The generic term of a matrix A is denoted $A(i,j)$. We denote by $A \otimes B$ the Kronecker product of two matrices A and B, and $\operatorname{vec} A$ denotes the vector obtained by stacking the columns of A. The convergence in probability is denoted by $\xrightarrow{P}$. The symbol $\Rightarrow$ denotes the convergence in distribution and the almost surely convergence is denoted by $\xrightarrow{a.s.}$.

2. Testing for linear causality in mean in weak VAR models

The test of linear causality in mean is often conducted within the context of VAR models. Indeed such models do not need a priori restrictions on the parameters. They also allow to avoid some problems of identification as for instance in the case of Vector AutoRegressive Moving Average (VARMA) models and account for linear intertemporal dynamics between variables. We consider the following VAR model

$$X_t = \sum_{i=1}^{p_0} A_{0i} X_{t-i} + \epsilon_t \qquad \text{for all } t \in \{0, \pm 1, \pm 2, \dots\} \tag{2.1}$$

where the X_t's are d-dimensional vectors. The matrices A_{0i} are of dimension $d \times d$ and such that $\det A_0(z) \neq 0$ for all $|z| \leq 1$, where $A_0(z) = I_d - \sum_{i=1}^{p_0} A_{0i} z^i$. The error process (ϵ_t) is commonly assumed to be iid Gaussian with positive definite covariance matrix Σ_ϵ and such that $E\epsilon_t = 0$. In this case it is said that (2.1) is a *strong VAR* model. However the strong white noise assumption is often considered to be too restrictive. Indeed there are processes which do not satisfy the strong assumption of iid innovations as for instance the multivariate GARCH models (see e.g. [5] for the MGARCH models). In addition there are some cases where the error process is not a martingale difference as for instance when we consider the causal representation of a non causal VAR process (i.e. when roots of the AR polynomial are inside the unit circle). Many applications for non causal models are given in [7]. Numerous situations where the iid gaussian assumption on the innovation process is not hold are presented in [32] or [16] among other references. More generally using the Wold decomposition theorem, it can be shown that a large set of processes can be approximated by VAR models (see [27]). However since the innovations are only supposed uncorrelated in the Wold decomposition theorem, assuming a strong VAR model for an observed process is very restrictive in many cases. Thus we consider the test for linear causality in mean in the framework given by the following assumption which allows for a large set of dynamics for the error process.

Assumption A1. *The process* (ϵ_t) *is strictly stationary ergodic with positive definite covariance matrix* Σ_ϵ *and such that* $E\epsilon_t = 0$, $Cov(\epsilon_t, \epsilon_{t-h}) = 0$ *for all* $t \in \{0, \pm 1, \pm 2, \dots\}$ *and all* $h \neq 0$.

Note that A1 is clearly weaker than the iid standard assumption. Indeed an error process which verifies A1 can be a martingale difference or can even be such that $E(\epsilon_t \mid \epsilon_{t-1}, \dots) \neq 0$. If we suppose that the *weak white noise* assumption A1 holds, it is said that (2.1) is a *weak VAR* model. Therefore it is easy to see from A1 that the set of weak VAR processes can take into account a large set of temporal dynamics.

In this part we introduce the tested hypotheses in our framework. Let us first rewrite model (2.1) as follows

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \sum_{i=1}^{p_0} \begin{pmatrix} A_{0i,11} & A_{0i,12} \\ A_{0i,21} & A_{0i,22} \end{pmatrix} \begin{pmatrix} X_{1t-i} \\ X_{2t-i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix},$$

where X_{1t} is of dimension d_1 and the matrices $A_{0i,mj}$ have appropriate size. It is said that (X_{2t}) does not linearly Granger-cause (X_{1t}) in mean if we have

$$EL(X_{1t} \mid X_{1t-1}, \dots) = EL(X_{1t} \mid X_{1t-1}, X_{2t-1}, \dots),$$

where $EL(X_{1t} \mid \dots)$ is the linear conditional expectation. It is well known that (X_{2t}) does not linearly Granger-cause (X_{1t}) in mean if and only if $A_{0i,12} = 0$ for all $i \in \{1, \dots, p_0\}$ (see [27]). Therefore when testing the null hypothesis that (X_{2t}) does not linearly cause (X_{1t}) in mean versus the alternative that (X_{2t}) linearly causes (X_{1t}) in mean, we consider the following pair of hypotheses

$$H_0 \ : \ A_{0i,12} = 0 \quad \text{for all} \quad i \in \{1, \dots, p_0\} \quad \text{v.s.} \quad H_1 \ : \ A_{0i,12} \neq 0$$

for at least one $i \in \{1, \dots, p_0\}$. For our study we also introduce the concept of causality in variance. It is said that (X_{2t}) does not Granger-cause (X_{1t}) in variance if we have

$$E(X_{1t}^2 \mid X_{1t-1}, \dots) = E(X_{1t}^2 \mid X_{1t-1}, X_{2t-1}, \dots).$$

Weak VAR models allow for conditional heteroscedasticity, so that testing linear causality in mean in presence of causality in variance can be considered as a particular case of our framework. We will consider causality in variance in the Monte Carlo experiments because of the considerable attention has been paid for this kind of nonlinear causality relation in the literature (see e.g. [11] or [30]). In addition we note that most of the nonlinear models produce processes with conditional heteroscedasticity, as for instance hidden Markov models (see e.g. [1]). Furthermore models which produce processes with conditional heteroskedasticity are much employed in the literature (see e.g. [8, 12, 29]). In conclusion note that the strong assumption of iid errors is often considered to be not realistic. Thus taking into account for nonlinear dynamics in the error process when building tests for linear causality is important. Since there is no possible confusion, we refer to the linear Granger-causality in mean as linear causality in mean in the rest of the paper.

3. Asymptotic behaviour of the QMLE

Let us first re-write model (2.1) as follow

$$X_t = (\tilde{X}_{t-1}' \otimes I_d)\theta_0 + \epsilon_t, \tag{3.1}$$

where $\tilde{X}_{t-1} = (X_{t-1}', \dots, X_{t-p_0}')'$ and $\theta_0 = \mathrm{vec}(A_{01}, \dots, A_{0p_0})$. We use the quasi maximum likelihood method for the estimation procedure since the error terms are no longer assumed Gaussian. Consider the observations $X_1, \dots, X_T$ and set $X_t = 0$ for $t < 1$ in the sequel. The QMLE is given by

$$\hat{\theta} = \mathrm{vec}(\hat{\Sigma}_{X_t, \tilde{X}_{t-1}} \hat{\Sigma}_{\tilde{X}_{t-1}}^{-1}), \tag{3.2}$$

where

$$\hat{\Sigma}_{X_t, \tilde{X}_{t-1}} = T^{-1} \sum_{t=1}^{T} X_t \tilde{X}_{t-1}' \quad \text{and} \quad \hat{\Sigma}_{\tilde{X}_{t-1}} = T^{-1} \sum_{t=1}^{T} \tilde{X}_{t-1} \tilde{X}_{t-1}',$$

(see e.g. [27] for more details on the derivation of the maximum likelihood estimator). We also obtain $\hat{\Sigma}_\epsilon = T^{-1} \sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t'$, where $\hat{\epsilon}_t = \epsilon_t(\hat{\theta}) = X_t - (\tilde{X}_{t-1}' \otimes I_d)\hat{\theta}$ are the residuals. Note that before testing the linear causality in mean, it is important to study the adequacy of the fitted model. In our framework of possibly dependent errors one can check that the autoregressive order is well fitted using portmanteau tests developed by [15]. Then we will suppose that the autoregressive order is well fitted.

We use results on the QMLE stated below to build statistical tools for testing linear causality in mean which can take into account the presence of nonlinearities. Let us first define the mixing coefficients

$$\alpha_a(h) = \sup_{A \in \sigma(a_u, u \le t), B \in \sigma(a_u, u \ge t+h)} |P(A \cap B) - P(A)P(B)|,$$

that measure the temporal dependence of the stationary process (a_t). We also define $\|a_t\|_q = (E\|a_t\|^q)^{1/q}$, where $\|.\|$ denotes the Euclidean norm with $E\|a_t\|^q < \infty$. In order to state our asymptotic results, we need to make the following assumption.

Assumption A2. *The process* (ϵ_t) *satisfies* $\|\epsilon_t\|_{4+2\nu} < \infty$, *and the mixing coefficients of the process* (ϵ_t) *are such that* $\sum_{h=0}^{\infty}\{\alpha_\epsilon(h)\}^{\nu/(2+\nu)} < \infty$ *for some* $\nu > 0$.

The moment assumption in A2 is slightly stronger than the standard moment assumption $\|\epsilon_t\|_4 < \infty$. It is interesting to note that the mixing assumption is valid for a large class of processes (see [31] or [10]). For instance note that **A2** is satisfied for exponential strongly mixing sequences. The following results are given in [15], making the mixing assumption on the observed process (X_t). More precisely under A1, the matrix $\hat{\Sigma}_{\tilde{X}_{t-1}}$ is almost surely invertible and we have

$$\hat{\theta} \xrightarrow{a.s.} \theta_0, \quad \hat{\Sigma}_\epsilon \xrightarrow{a.s.} \Sigma_\epsilon. \tag{3.3}$$

In addition if we suppose that A2 holds, we have

$$T^{\frac{1}{2}}\{\hat{\theta} - \theta_0\} \Rightarrow \mathcal{N}(0, J^{-1}IJ^{-1}), \tag{3.4}$$

where $J = \Sigma_{\tilde{X}} \otimes \Sigma_\epsilon^{-1}$ with $\Sigma_{\tilde{X}} = E(\tilde{X}_t\tilde{X}_t')$, and

$$I = \sum_{h=-\infty}^{\infty} E\left\{\tilde{X}_{t-1}\tilde{X}_{t-h-1}' \otimes \Sigma_\epsilon^{-1}\epsilon_t\epsilon_{t-h}'\Sigma_\epsilon^{-1}\right\},$$

If we suppose that the standard assumption on the innovation process hold we have $I = J$, so that we obtain in this case

$$T^{\frac{1}{2}}\{\hat{\theta} - \theta_0\} \Rightarrow \mathcal{N}(0, J^{-1}) \tag{3.5}$$

(see e.g. [27], p 74)). However the matrix I can be very different from the matrix J when the errors are dependent as illustrated in the following example.

Example 3.1. Consider an uncorrelated bivariate process $X_t = (X_{1t}, X_{2t})'$. If we want to study the linear causality in mean from (X_{2t}) to (X_{1t}), we fit an AR(1) model $X_t = A_0X_{t-1} + \epsilon_t$, while the underlying true value of A_0 is zero. We assume that the innovation process (ϵ_t) follows an ARCH(1) model with constant correlation given in [22]

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} = \begin{pmatrix} \sigma_{11,t} & 0 \\ 0 & \sigma_{22,t} \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}$$

where

$$\begin{pmatrix} \sigma_{11,t}^2 \\ \sigma_{22,t}^2 \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.2 \end{pmatrix} + \begin{pmatrix} 0.4 & 0 \\ 0.1 & 0.25 \end{pmatrix} \begin{pmatrix} \epsilon_{1t-1}^2 \\ \epsilon_{2t-1}^2 \end{pmatrix}.$$

Suppose for simplicity that (η_{1t}, η_{2t}) is Gaussian with variance I_2. In this case the matrices I and J are diagonal and we obtain $I = diag(2.538, 2.362, 0.664, 3.474)$ and $J = diag(1, \frac{3}{2}, \frac{2}{3}, 1)$.

It is clear that using the result (3.5) for the statistical analysis of VAR models, when the error process is dependent, can be quite misleading. On the other hand note that the tests build using (3.4) are valid for a large set of processes.

We define several estimators for the covariance matrices which appear in (3.4). We consider the following estimator for J

$$\hat{J} = \hat{\Sigma}_{\tilde{X}_{t-1}} \otimes \hat{\Sigma}_{\epsilon}^{-1}.$$

From the consistency of $\hat{\theta}$ and the ergodic theorem, we have $\hat{J} = J + o_p(1)$. To introduce a consistent estimator of the variance matrix in (3.4), we first note that

$$J^{-1}IJ^{-1} = \left(\Sigma_{\tilde{X}}^{-1} \otimes I_d\right) \left\{ \sum_{h=-\infty}^{\infty} E(\Upsilon_t \Upsilon'_{t-h}) \right\} \left(\Sigma_{\tilde{X}}^{-1} \otimes I_d\right) := \Lambda \Xi \Lambda,$$

where $\Upsilon_t = \tilde{X}_{t-1} \otimes \epsilon_t$. The matrix Λ can be consistently estimated by $\hat{\Lambda} = \hat{\Sigma}_{\tilde{X}_{t-1}}^{-1} \otimes I_d$. Now let us define $\hat{\Upsilon}_t = \tilde{X}_{t-1} \otimes \hat{\epsilon}_t$, and $\hat{\mathcal{A}}_q(z) = I_{d^2p} - \sum_{i=1}^{q} \hat{\mathcal{A}}_{q,i} z^i$, where $\hat{\mathcal{A}}_{q,1}, \ldots, \hat{\mathcal{A}}_{q,q}$ denote the coefficients of the LS regression of $\hat{\Upsilon}_t$ on $\hat{\Upsilon}_{t-1}$, $\ldots, \hat{\Upsilon}_{t-q}$. We denote by $\tilde{\epsilon}_{q,t}$ the residuals of this regression and $\hat{\Sigma}_{\tilde{\epsilon}_q} = T^{-1} \sum_{t=1}^{T} \tilde{\epsilon}_{q,t} \tilde{\epsilon}'_{q,t}$. In the framework of weak VAR models it is shown in [15] that

$$\hat{\Xi} := \hat{\mathcal{A}}_q^{-1}(1) \hat{\Sigma}_{\tilde{\epsilon}_q} \hat{\mathcal{A}}_q'^{-1}(1) \xrightarrow{P} \Xi \tag{3.6}$$

when $q = q(T) \to \infty$ and $q^3/T \to 0$ as $T \to \infty$. The order q can be chosen by considering an information criterion. Note that one may use other kinds of HAC estimators for the estimation of Ξ (see [13] or [17] for more details on this kind of covariance matrix estimation method).

In the case of martingale difference error processes the expression of Ξ simplifies into

$$\Xi = E\left(\Upsilon_t \Upsilon'_t\right), \tag{3.7}$$

and following [36] we will also consider in the sequel the consistent covariance estimator of [38]

$$\tilde{\Xi} = T^{-1} \sum_{t=1}^{T} \hat{\Upsilon}_t \hat{\Upsilon}'_t. \tag{3.8}$$

Nevertheless note that the simplification (3.7) is not available in general if we have $E(\epsilon_t \mid \epsilon_{t-1}, \ldots) \neq 0$.

4. Tests for linear causality in mean

We propose several approaches which can potentially give improvements for the test of linear causality in mean when the number of parameters we have to estimate is high. In such situations practitioners usually consider the standard Wald test or Wald tests with modified statistics introduced below. In this section we use the LM and Likelihood Ratio (LR) approaches to build tests, and tests with standard statistics and modified distribution. To introduce these modified tests let us define the block diagonal matrix $R = diag(C, \dots, C)$ of dimension $pd_1d_2 \times pd^2$, where C is a $d_1d_2 \times d^2$-dimensional matrix given by

$$C = \begin{pmatrix} 0_{d_1 \times d_1 d} & I_{d_1} & 0_{d_1 \times d_2} & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & I_{d_1} & 0_{d_1 \times d_2} \end{pmatrix},$$

so that, under the null hypothesis, we have $R\theta_0 = 0$. Andrews and Monahan [3] pointed out that the use of HAC estimators in the test statistics may yield to over-rejections. Therefore we propose modified tests which are based on standard statistics. The standard Wald, LM and LR statistics are respectively given by the expressions

$$Q_{WS} = T\hat{\theta}' R'(R\hat{J}^{-1}R')^{-1}R\hat{\theta},$$

$$Q_{LMS} = T^{-1}\mathcal{S}'R'(R\hat{J}^{-1}R')R\mathcal{S}$$

and

$$Q_{LR} = 2\left\{\mathcal{L}(\hat{\theta}, \hat{\Sigma}_\epsilon) - \mathcal{L}(\hat{\theta}^c, \hat{\Sigma}_\epsilon^c)\right\},$$

where $\mathcal{S} = \frac{\partial \mathcal{L}(\hat{\theta}^c, \hat{\Sigma}_\epsilon^c))}{\partial \theta}$ is the score vector and $\hat{\Sigma}_\epsilon^c = T^{-1}\sum_{t=1}^T \epsilon_t(\hat{\theta}^c)\epsilon_t(\hat{\theta}^c)$ with $\epsilon_t(\hat{\theta}^c) = X_t - (\tilde{X}'_{t-1} \otimes I_d)\hat{\theta}^c$. The log-likelihood is denoted by $\mathcal{L}$. The constrained estimator $\hat{\theta}^c$ is obtained using OLS estimation. Note that $\hat{\theta}^c$ is used for the estimation of J in Q_{LMS}. The following theorem gives the asymptotic behaviour of the standard statistics.

Theorem 4.1. *If we suppose that assumptions A1 and A2 hold, then under the null hypothesis, the standard statistics Q_{WS}, Q_{LMS} and Q_{LR} converge in distribution, as $T \to \infty$, to*

$$Z(\xi) = \sum_{i=1}^{pd_1d_2} \xi_i Z_i^2 \tag{4.1}$$

where $\xi = (\xi_1, \dots, \xi_{pd_1d_2})'$ is the vector of the eigenvalues of the matrix

$$\Omega = (RJ^{-1}R')^{-\frac{1}{2}}(RJ^{-1}IJ^{-1}R')(RJ^{-1}R')^{-\frac{1}{2}}$$

and the Z_i's are independent $\mathcal{N}(0,1)$ variables.

Using Theorem 4.1, we are now in position to define the tests for linear causality in mean we propose in this paper. These tests are based on modified critical values which can be obtained as follows. It is clear that using the consistent HAC estimator $\hat{\Xi}$ and $\hat{J}$ we obtain

$$\hat{\Omega} := (R\hat{J}^{-1}R')^{-\frac{1}{2}}(R\hat{\Lambda}\hat{\Xi}\hat{\Lambda}R')(R\hat{J}^{-1}R')^{-\frac{1}{2}} \xrightarrow{P} \Omega.$$

Now let us define by $\hat{\xi} = (\hat{\xi}_1, \dots, \hat{\xi}_{pd_1d_2})'$ the vector of the eigenvalues of the matrix $\hat{\Omega}$. At the asymptotic level v, the W_{md} test (resp. the LM_{md}, LR_{md}

tests) consists in rejecting the null hypothesis that (X_{2t}) does not linearly cause (X_{1t}) in mean when

$$P\{Z(\hat{\xi}) > Q_{WS}\} < \upsilon \quad (\text{resp.} \quad P\{Z(\hat{\xi}) > Q_{LMS}\} < \upsilon, \ P\{Z(\hat{\xi}) > Q_{LR}\} < \upsilon). \tag{4.2}$$

The subscript md stands for "modified distribution". In practice one can evaluate the p-values in (4.2) using the Imhof algorithm [21] or the saddlepoint approximation method (see e.g. [26]). Note that the use of standard statistics allows us to define a LR test which is adapted to our framework.

If we suppose that the error process is iid we have $I = J$, so that $\Omega = I_{pd_1d_2}$ in this case and we obtain the standard results

$$Q_{WS} \Rightarrow \chi^2_{pd_1d_2}, \quad Q_{LMS} \Rightarrow \chi^2_{pd_1d_2} \quad \text{and} \quad Q_{LR} \Rightarrow \chi^2_{pd_1d_2}. \tag{4.3}$$

It is clear from Example 3.1 that considering the critical values of the $\chi^2_{pd_1d_2}$ distribution while the errors are dependent can be quite misleading when we use the standard statistics. Indeed in this example (X_{2t}) does not cause (X_{1t}) in variance, while (X_{1t}) causes (X_{2t}) in variance. When testing if (X_{2t}) linearly causes (X_{1t}) in mean, we take $R = (0,0,1,0)$ so that $RJ^{-1}IJ^{-1}R' \approx RJ^{-1}R' = \frac{3}{2}$. If one want to test linear causality in mean from (X_{1t}) to (X_{2t}), we take $R = (0,1,0,0)$, so that we have $RJ^{-1}IJ^{-1}R' = 1.128$ and $RJ^{-1}R' = \frac{2}{3}$. Then for Example 3.1 one can use the standard distribution for testing the linear causality in mean from (X_{2t}) to (X_{1t}) since $E(X_{1t}^2 \mid X_{2t-1}, X_{1t-1} \dots) = E(X_{1t}^2 \mid X_{1t-1} \dots)$. However the test for linear causality in mean from (X_{1t}) to (X_{2t}) using the standard approach can be misleading. In the sequel the standard tests based on the results in (4.3) are denoted by W_s, LM_s and LR_s. The W_s test is the most commonly used test in the literature for testing linear causality in mean relations.

In the literature it is common to consider Wald tests with corrected statistics when nonlinear dynamics are suspected in the error process. This approach is studied by [36] for the test of linear Granger causality in mean in presence of heteroscedasticity. We also study these tests in the case where a large number of parameters have to be estimated. Let us consider the following modified statistics for the Wald test

$$Q_W = T\hat{\theta}' R'(R\hat{\Lambda}\hat{\Xi}\hat{\Lambda}R')^{-1}R\hat{\theta}.$$

Similarly we introduce the LM modified statistic

$$Q_{LM} = T^{-1}\mathcal{S}'R'(R\hat{J}^{-1}R')(R\hat{\Lambda}\hat{\Xi}\hat{\Lambda}R')^{-1}(R\hat{J}^{-1}R')R\mathcal{S}.$$

In the sequel we suppose that Ξ is invertible, so that $\hat{\Xi}$ is invertible at least asymptotically. The tests with modified statistics are based on the following result.

Theorem 4.2. *If we suppose that assumptions A1, A2 hold and under the null hypothesis, the modified statistics Q_W and Q_{LM} are asymptotically $\chi^2_{pd_1d_2}$ distributed.*

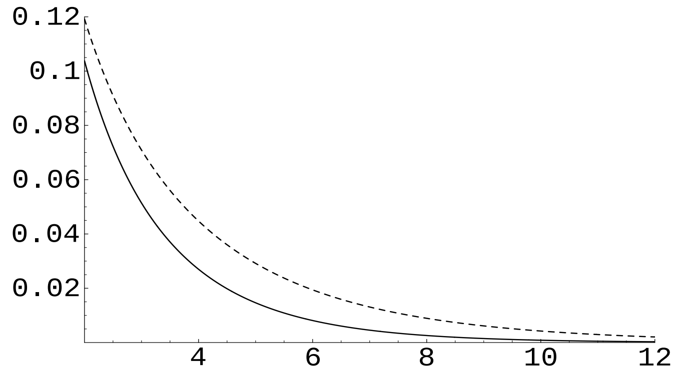

Fig 1. *The tails of the asymptotic distributions of the modified Wald statistic (full line) and the standard statistic (dotted line) in the case of the test of the linear Granger causality in mean from (X_{1t}) to (X_{2t}) in Example 3.1.*

At the asymptotic level υ, the W_{ms} test (resp. the LM_{ms} test) consists in rejecting the null hypothesis that (X_{2t}) does not linearly cause (X_{1t}) in mean when

$$Q_W > U_{1-\upsilon} \quad (\text{resp. } Q_{LM} > U_{1-\upsilon}),$$

where $U_{1-\upsilon}$ is such that $P\{\chi^2_{pd_1d_2} > U_{1-\upsilon}\} = \upsilon$. The subscript ms stands for "modified statistic". If we suppose that the error process is iid we have $I = J$, so that we obtain $Q_W \approx Q_{WS}$ and $Q_{LM} \approx Q_{LMS}$ for large enough sample size in this case. Nevertheless if we assume that the error process is dependent, Example 3.1 shows that the modified statistics we use can be quite different from the standard statistics even for large T. In Figure 1 we plotted the asymptotic distributions of the modified and standard Wald statistics in the case of testing linear Granger causality in mean from (X_{1t}) to (X_{2t}) in Example 3.1. In this case the asymptotic distribution of the standard Wald statistic is $1.58\chi^2_1$. It is found that the asymptotic distribution of the modified Wald statistic is quite different from the asymptotic distribution of the standard Wald statistic.

Vilasuso [36] also considered the following Wald statistic based on the White variance correction

$$\tilde{Q}_W = T\hat{\theta}' R'(R\hat{\Lambda}\tilde{\hat{\Xi}}\hat{\Lambda}R')^{-1}R\hat{\theta}.$$

Similarly we introduce the following corrected LM statistic

$$\tilde{Q}_{LM} = T^{-1}\mathcal{S}'R'(R\hat{J}^{-1}R')(R\hat{\Lambda}\tilde{\hat{\Xi}}\hat{\Lambda}R')^{-1}(R\hat{J}^{-1}R')R\mathcal{S}.$$

If we assume that the error process is a martingale difference, it can be shown that $\tilde{Q}_W \Rightarrow \chi^2_{pd_1d_2}$ and $\tilde{Q}_{LM} \Rightarrow \chi^2_{pd_1d_2}$. Using these results, one can consider the tests with modified statistics denoted by $\widetilde{W}_{ms}$ and $\widetilde{LM}_{ms}$ in a similar way to the W_{ms} and LM_{ms} tests. If we suppose that (ϵ_t) is a martingale difference, it can also be shown that the statistics Q_{WS}, Q_{LMS} and Q_{LR} converge in distribution to

$$Z\left(\tilde{\xi}\right) = \sum_{i=1}^{pd_1d_2} \tilde{\xi}_i Z_i^2$$

where $\tilde{\xi} = (\tilde{\xi}_1, \dots, \tilde{\xi}_{pd_1d_2})'$ is the vector of the eigenvalues of the matrix

$$\tilde{\Omega} = (RJ^{-1}R')^{-\frac{1}{2}} \left[R\Lambda\{E\left(\Upsilon_t\Upsilon_t'\right)\}\Lambda R'\right] (RJ^{-1}R')^{-\frac{1}{2}}.$$

As a consequence using again the White covariance estimator defined in (3.8), we introduce the tests with modified distributions denoted by $\widetilde{W}_{md}$, $\widetilde{LM}_{md}$ and $\widetilde{LR}_{md}$ in a similar way to the W_{md}, LM_{md} and LR_{md} tests.

In this section we have considered different approaches to build tests. The main advantage of considering tests with modified statistics is that we obtain tests with a standard chi-squared asymptotic distribution. However we have to invert $\hat{\Xi}$ or $\tilde{\Xi}$ to implement these tests, on the contrary to the modified tests based on the standard statistics. We also remark that the tests based on the HAC estimation have larger theoretical basis than the tests based on the White estimation method. Nevertheless the tests based on the White estimation are easier to implement than the tests based on the HAC estimation. Finally note that the use of the constrained estimator can potentially give some efficiency to the LM tests when there is a large number of parameters to estimate.

5. Monte Carlo experiments

The small sample properties of the tests presented in the previous section are compared in several situations. We consider the simple bivariate AR(1) parameters given by cases (a) and (c) in Table 1. We study the AR(1) case to illustrate the difference in the analysis between this simple case, and the cases where the number of parameters increase. To assess the behaviour of the tests under comparison when the autoregressive order increase, we also use the AR(2), AR(3) and AR(4) parameters given by cases (b), (d), (e) and (g). The finite sample performances of the tests when the dimension is high is investigated using the 4 and 5-dimensional parameters given by cases (h) and (f).

Independent and dependent error terms are considered. For experiments where the errors are iid, we let $\epsilon_t \sim \mathcal{N}(0, I_d)$. In order to illustrate the effect of ARCH innovations on the different tests we consider the model with constant correlation proposed by [22]. We have chosen this model because of its simplicity. In our simulations the process (ϵ_t) follows the Data Generating Process (DGP) given by

$$\begin{pmatrix} \epsilon_{1t} \\ \vdots \\ \epsilon_{dt} \end{pmatrix} = \begin{pmatrix} \sigma_{1t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{dt} \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \vdots \\ \eta_{dt} \end{pmatrix} \tag{5.1}$$

Table 1
Parameters used in the Monte Carlo experiments

(a) $A_{01} = \begin{pmatrix} 0.4 & \mathbf{a} \\ 0 & 0.4 \end{pmatrix}$ (b) $A_{01} = \begin{pmatrix} 0.4 & \mathbf{a} \\ 0 & 0.4 \end{pmatrix}$, $A_{02} = 0.2I_2$, $A_{03} = 0.1I_2$	
(c) $A_{01} = 0.4I_2$ (d) $A_{01} = 0.4I_2$, $A_{02} = 0.2I_2$ (e) $A_{01} = 0.4I_2$, $A_{02} = 0.2I_2$, $A_{03} = 0.1I_2$ (f) $A_{01} = 0.4I_5$ (g) $A_{01} = 0.4I_2$, $A_{02} = 0.2I_2$, $A_{03} = 0.1I_2$, $A_{04} = -0.1I_2$	
(h) $A_{01} = \begin{pmatrix} 0.6 & \mathbf{a} & \mathbf{a} & \mathbf{a} \\ 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 0.6 \end{pmatrix}$	(i) $B = \begin{pmatrix} 0.3 & 0 & 0 & 0.1 & 0.1 \\ 0 & 0.3 & 0 & 0.1 & 0.1 \\ 0 & 0 & 0.3 & 0.1 & 0.1 \\ 0 & 0 & 0 & 0.3 & 0.1 \\ 0 & 0 & 0 & 0 & 0.3 \end{pmatrix}$
(j) $B = 0.3I_2$ (k) $B = \begin{pmatrix} 0.3 & 0.2 \\ 0 & 0.3 \end{pmatrix}$ (l) $B = 0.3I_5$	
(m) $B = \begin{pmatrix} 0.3 & 0 & 0.1 & 0.1 \\ 0 & 0.3 & 0.1 & 0.1 \\ 0 & 0 & 0.3 & 0.1 \\ 0 & 0 & 0 & 0.3 \end{pmatrix}$	

where

$$\begin{pmatrix} \sigma_{1t}^2 \\ \vdots \\ \sigma_{dt}^2 \end{pmatrix} = \begin{pmatrix} \omega \\ \vdots \\ \omega \end{pmatrix} + \begin{pmatrix} b(1,1) & \dots & b(1,d) \\ \vdots & \ddots & \vdots \\ b(d,1) & \dots & b(d,d) \end{pmatrix} \begin{pmatrix} \epsilon_{1t-1}^2 \\ \vdots \\ \epsilon_{dt-1}^2 \end{pmatrix}.$$

The process $\eta_t = (\eta_{1t}, \dots, \eta_{dt})'$ is iid, such that $\eta_t \sim \mathcal{N}(0, I_d)$. We take $\omega = 0.1$. In the sequel we will denote by B the matrix of generic term $b(i,j) \geq 0$. In addition the parameters are chosen such that the stationarity conditions hold (see [22] for more details). The process defined by (5.1) presents conditional heteroscedasticity. We use parameters (i)-(m) in Table 1 for weak white noise (5.1). Note that for parameters (i), (k) and (m), we take $b(i,j) > 0$ for $i \in \{1, \dots, d_1\}$ and $j \in \{d_1 + 1, \dots, d\}$, such that $(X_{d_1+1\,t}, \dots, X_{dt})'$ is causal in variance for $(X_{1t}, \dots, X_{d_1\,t})'$.

The weak error process in (5.1) is such that the best predictor is linear. To illustrate the case where the best predictor is not linear, we consider the following bivariate DGP

$$\epsilon_{1t} - \phi\epsilon_{1t-1} = u_t - \phi^{-1}u_{t-1}, \tag{5.2}$$

with $\phi = 0.6$ and $u_t = \eta_t\epsilon_{2t}$. The iid standard Gaussian processes (ϵ_{2t}) and (η_t) are independent. The process (ϵ_{1t}) follows an all-pass model and is then uncorrelated, so that we write $EL(\epsilon_{1t} \mid \epsilon_{1t-1}, \dots) = 0$. However since (u_t) is not Gaussian we have $E(\epsilon_{1t} \mid \epsilon_{1t-1}, \dots) \neq 0$ and hence the best predictor is not linear on the contrary of the ARCH process (5.1). To see this note that we have $E\{\epsilon_{1t}(\epsilon_{1t-1} - \phi\epsilon_{1t-2})^3\} = (E(u_t^4) - 3)(1 - \phi^{-2})^2\phi$. Thereby we obtain

TABLE 2

Empirical size (in %) of the tests under comparison for testing linear non causality in mean in the bivariate case with autoregressive order $p_0 = 1$

Case	iid			no causality in variance			causality in variance		
T	100	300	1000	100	300	1000	100	300	1000
LM_s	4.5	5.4	5.8	4.8	5.7	5.7	**8.3**	**9.6**	**10.4**
LR_s	5.0	5.5	5.8	5.1	6.1	5.8	**9.2**	**9.7**	**10.4**
W_s	5.2	5.6	5.8	5.3	6.1	5.9	**9.3**	**9.7**	**10.5**
LM_{ms}	5.7	5.8	5.7	5.8	**6.7**	5.6	**6.6**	**7.2**	6.0
W_{ms}	**6.4**	6.0	5.8	**6.9**	**6.7**	5.7	**7.2**	**7.2**	6.0
$\widetilde{LM}_{ms}$	5.6	5.5	5.6	5.3	6.0	5.7	6.0	6.1	6.0
$\widetilde{W}_{ms}$	6.3	5.9	5.7	6.2	6.1	5.7	**7.0**	6.3	6.1
LM_{md}	5.7	5.8	5.7	5.8	**6.7**	5.6	**6.6**	**7.2**	6.0
LR_{md}	6.1	5.9	5.7	**6.5**	**6.7**	5.7	**7.1**	**7.2**	6.0
W_{md}	**6.4**	6.0	5.8	**6.9**	**6.7**	5.7	**7.2**	**7.2**	6.0
$\widetilde{LM}_{md}$	5.6	5.5	5.6	5.3	6.0	5.7	6.0	6.1	6.0
$\widetilde{LR}_{md}$	6.2	5.8	5.6	6.1	6.1	5.7	**7.0**	6.3	6.1
$\widetilde{W}_{md}$	6.3	5.9	5.7	6.2	6.1	5.7	**7.0**	6.3	6.1

$E\{E(\epsilon_{1t} \mid \epsilon_{1t-1}, \dots)(\epsilon_{1t-1} - \phi\epsilon_{1t-2})^3\} = (E(u_t^4) - 3)(1 - \phi^{-2})^2\phi \neq 0$ since $E(u_t^4) = 9$, so that the result follow. We chosen an all-pass model for our illustrations since it is well known that this kind of models can capture some features of nonlinear processes (see [7]).

In the sequel we simulate $n = 1000$ independent trajectories in each experiment. The null hypothesis that (X_{2t}) does not linearly cause (X_{1t}) in mean is tested for each simulated process $X_t = (X'_{1t}, X'_{2t})'$.

In this part we study the empirical size of the tests under comparison. The simulated processes are of lengths $T = 100$, $T = 300$ and $T = 1000$. The asymptotic nominal level 5% is used in all the experiments for the standard and modified tests. Note that since $n = 1000$ replications are performed and assuming that the finite sample size of the tests is 5%, the relative rejection frequencies should be between the significant limits 3.65% and 6.35% with probability 0.95. Then the relative rejection frequencies are displayed in bold type when they are outside the significant limits 3.65% and 6.35% in Tables 2-7.

We first study the case of simple bivariate AR(1) processes generated using parameter (c). We consider for the error process the iid case, the case of ARCH errors with no causality in variance using parameter (j) and the case of ARCH errors with causality in variance using parameter (k). The results are given in Table 2. We first analyze the results for the cases of iid errors and when there is no causality in variance. We find that the relative rejections frequencies of the different tests are close to the asymptotic nominal level in general. We only remark slight over-rejections of some of the modified Wald and LR tests when the samples are small. When (X_{2t}) cause (X_{1t}) in variance we note that the standard tests are clearly oversized, even when the samples are large. In presence of causality in variance the modified tests perform better than the standard tests. The relative rejection frequencies of the modified tests converge to the asymptotic nominal level as the samples increase. The results we found

Table 3
Empirical size (in %) of the tests under comparison for testing linear non causality in mean in the bivariate case with different autoregressive orders. The innovations are iid

Case	$p_0=2$			$p_0=3$			$p_0=4$		
T	100	300	1000	100	300	1000	100	300	1000
LM_s	4.7	5.9	5.0	4.5	5.0	3.9	4.0	5.0	4.1
LR_s	6.0	6.0	5.4	**7.0**	5.6	4.3	**8.2**	5.8	4.2
W_s	**6.6**	6.0	5.4	**7.7**	5.6	4.3	**9.0**	5.9	4.2
LM_{ms}	**6.4**	5.9	4.7	**11.9**	5.1	4.5	**43.2**	5.6	4.2
W_{ms}	**8.4**	**6.6**	4.9	**14.4**	5.8	4.7	**45.8**	7.0	4.7
$\widetilde{LM}_{ms}$	6.1	5.8	4.7	**6.6**	5.0	4.5	**7.4**	5.6	4.2
$\widetilde{W}_{ms}$	**7.6**	**6.5**	5.0	**10.1**	5.7	4.7	**11.9**	6.8	4.7
LM_{md}	5.3	6.0	5.0	5.8	5.1	4.0	**7.9**	4.8	4.1
LR_{md}	6.3	6.2	5.1	**7.5**	6.0	4.0	**11.2**	5.8	4.4
W_{md}	**6.6**	6.3	5.2	**8.3**	6.0	4.0	**12.0**	5.9	4.4
$\widetilde{LM}_{md}$	5.1	6.0	4.9	4.5	5.1	4.0	4.5	4.8	4.1
$\widetilde{LR}_{md}$	5.9	6.2	5.1	**6.7**	6.0	4.0	**8.1**	5.8	4.4
$\widetilde{W}_{md}$	**6.4**	6.2	5.2	**7.5**	5.9	4.0	**9.3**	5.9	4.4

are consistent with those obtained by [36] for the standard Wald tests and Wald tests with modified statistics. The standard tests are not able to distinguish between causality in variance and linear causality in mean. In general it appears from the results in Table 2 that there is no difference between the tests with modified distributions and tests with modified statistics in this simple case. We also find that the LM tests have better results in general than the Wald or LR tests for small samples ($T = 100$). In particular the $\widetilde{LM}_{md}$ test well control the error of first kind in all cases.

The finite sample properties of the tests when the autoregressive order p_0 increase is investigated by considering bivariate AR(2), AR(3) and AR(4) processes generated using parameters (d), (e) and (g). In Table 3 we give the results when the errors are iid. The results in Table 4 correspond to the case of ARCH errors with causality in variance using parameter (k). From Tables 3 and 4 we can note that as p_0 increase, the tests with modified statistics are severely size distorted when the samples are small. This could be explained by the fact that the temporal dependence is more marked when p_0 is large and that complicated estimators of the weight matrices are inverted. In this case the modified statistics likely lead to reject the null hypothesis too often as pointed out by [3], although the corresponding tests are intended to take into account for nonlinearities. It also emerges that the results of the tests with modified distributions are clearly better than those of the tests with modified statistics for small samples. These results may be explained by the fact that we use the standard statistics which do not require to invert complicated estimators of the weight matrices for the tests with modified distributions, while the possible presence of nonlinearities are taken into account by using the modified distributions. The standard test have similar results to the $\widetilde{LM}_{md}$, $\widetilde{LR}_{md}$ and $\widetilde{W}_{md}$ when the errors are iid. However the standard tests are not valid in presence of causality in variance as before. We also find that the tests built using the White covariance matrix esti-

TABLE 4
Empirical size (in %) of the tests under comparison for testing linear non causality in mean in the bivariate case with different autoregressive orders. The error process is such that (X_{2t}) is causal in variance for (X_{1t})

Case	$p_0 = 2$			$p_0 = 3$			$p_0 = 4$		
T	100	300	1000	100	300	1000	100	300	1000
LM_s	**8.8**	**10.2**	**10.1**	**8.8**	**9.4**	**9.7**	**7.3**	**9.6**	**8.4**
LR_s	**10.4**	**10.7**	**10.3**	**11.1**	**10.6**	**10.0**	**12.3**	**11.2**	**9.0**
W_s	**10.6**	**10.8**	**10.3**	**11.8**	**11.0**	**10.0**	**12.9**	**11.3**	**9.1**
LM_{ms}	**9.7**	7.3	5.6	**21.8**	**7.8**	5.7	**49.0**	**9.0**	5.5
W_{ms}	**11.7**	**8.0**	5.6	**25.1**	**8.7**	5.9	**51.7**	**10.3**	5.9
$\widetilde{LM}_{ms}$	**6.5**	6.3	4.7	**7.4**	6.2	4.5	**8.0**	**6.9**	4.7
$\widetilde{W}_{ms}$	**9.8**	**7.2**	5.0	**11.6**	**7.4**	4.7	**14.0**	**8.2**	4.8
LM_{md}	**6.5**	6.0	4.9	**8.6**	5.5	4.9	**8.1**	5.4	4.1
LR_{md}	**8.6**	**6.8**	5.2	**11.3**	5.6	4.9	**12.5**	**6.6**	4.3
W_{md}	**8.8**	**6.7**	5.3	**12.0**	5.9	5.1	**13.6**	**6.8**	4.3
$\widetilde{LM}_{md}$	5.4	5.2	4.3	5.2	5.0	5.0	5.2	5.4	4.3
$\widetilde{LR}_{md}$	**7.6**	5.7	4.6	**8.2**	5.7	5.3	**9.0**	**7.1**	4.6
$\widetilde{W}_{md}$	**8.2**	5.8	4.6	**9.0**	6.2	5.4	**10.2**	**7.3**	4.6

mation have better results than the tests built using the HAC covariance matrix in small samples ($T = 100$). This can be explained by the relative simplicity of computing the White covariance matrices. We note from Tables 3 and 4 that the LM tests perform better than the Wald and LR tests when p_0 increase. This can be explained by the fact that the number of parameters to estimate increase faster for the Wald and LR tests than the LM tests based on restricted estimators. In particular we again find that the $\widetilde{LM}_{md}$ test well control the error of first kind in all cases. Finally we can remark that the relative rejection of the modified tests are close to the asymptotic nominal level for large samples ($T = 1000$). When the errors are iid the relative rejection frequencies of the standard tests also converge to the asymptotic nominal level.

We also study the test of linear causality in mean in the case of high dimensional VAR processes. For this we consider 5-dimensional VAR(1) processes generated using parameter (f). We consider for the error process the iid case, the case of dependent errors with no causality in variance using parameter (l) and the case of ARCH errors with causality in variance using parameter (i). The null hypothesis that $(X_{3t}, X_{4t}, X_{5t})'$ does not linearly cause $(X_{1t}, X_{2t})'$ in mean is tested. We first analyze the results for the small samples in Table 5. We find that the tests with modified distribution perform better than the tests with modified statistics. In these experiments the estimators of Ξ we use to build the modified statistics are of high dimension and then may be difficult to invert when the sample is small. From Table 5 the standard tests have similar results to that of the $\widetilde{LM}_{md}$, $\widetilde{LR}_{md}$ and $\widetilde{W}_{md}$ tests when the errors are iid or when there is no causality in variance. However we again remark that the standard tests are not valid in presence of causality in variance contrary to the $\widetilde{LM}_{md}$, $\widetilde{LR}_{md}$ and $\widetilde{W}_{md}$ tests. It can also be noted that the tests based on the White approach have better results than the tests built using the HAC estimator of the weight matrix. A possible explanation is that the HAC estimation method can

Table 5

Empirical size (in %) of the tests under comparison for testing linear non causality in mean in the 5-dimensional case with autoregressive order $p_0 = 1$

Case	iid			no causality in variance			causality in variance		
T	100	300	1000	100	300	1000	100	300	1000
LM_s	4.9	5.6	5.9	4.7	5.9	5.8	**8.1**	**9.9**	**10.4**
LR_s	**8.3**	**6.8**	5.9	**7.6**	6.3	6.3	**12.8**	**11.6**	**10.6**
W_s	**9.1**	**6.7**	6.0	**8.3**	**6.4**	6.3	**13.3**	**11.6**	**10.8**
LM_{ms}	**47.2**	**7.5**	6.1	**48.4**	**7.5**	**7.0**	**46.0**	**9.1**	**7.1**
W_{ms}	**47.4**	**8.1**	6.3	**49.5**	**8.6**	**7.3**	**47.1**	**11.0**	**7.4**
$\widetilde{LM}_{ms}$	**9.7**	**7.5**	6.1	**8.8**	**7.4**	**6.4**	**9.9**	**7.8**	**6.4**
$\widetilde{W}_{ms}$	**12.4**	**8.0**	6.3	**12.2**	**8.2**	**6.8**	**14.5**	**9.4**	**6.6**
LM_{md}	**28.3**	5.7	5.3	**29.5**	5.9	6.1	**29.5**	5.7	5.6
LR_{md}	**30.6**	**6.6**	5.7	**31.8**	6.3	**6.4**	**32.3**	**6.6**	6.0
W_{md}	**30.7**	**6.8**	5.7	**31.6**	**6.4**	**6.6**	**32.5**	**6.9**	6.0
$\widetilde{LM}_{md}$	5.1	5.7	5.3	4.4	6.1	5.9	5.0	5.6	5.9
$\widetilde{LR}_{md}$	**7.9**	**6.6**	5.7	**7.3**	**6.5**	6.3	**8.2**	**6.7**	6.0
$\widetilde{W}_{md}$	**8.4**	**6.8**	5.7	**7.8**	**6.6**	**6.5**	**9.0**	**7.0**	6.1

be viewed to be too complicated when the dimension is high and the samples are small when compared to the White estimation. It also appears that the LM tests perform much better than the Wald and LR tests. Similarly to the case of large autoregressive orders, this is can be explained by the fact that the number of parameters to be estimated increase faster for the Wald and LR tests than the LM tests as the dimension increase. We also remark here that the $\widetilde{LM}_{md}$ test is well performing in all cases. For large samples ($T = 1000$) we find that the relative rejection frequencies of the modified tests are in general close to the asymptotic nominal level.

In this part we investigate the effect of the innovations variance on the test of the linear causality in mean. More precisely we study cases where the variance is relatively large beside the autoregressive parameter and explore the properties of the modified and standard tests in such situations. These results will be compared with the cases where the variance is relatively small when compared to the autoregressive parameter. Five dimensional VAR(1) processes are again used in these experiments with autoregressive parameter $A_{01} = aI_5$. We consider the dependent case with causality in variance and we again use model (5.1) with $\omega = 2$ and parameter (i) for the matrix B when the variance is relatively large beside the autoregressive parameter ($a = 0.1$). We also study the case where $\omega = 0.1$ using parameter (i) for the matrix B and $a = 0.6$. In this case the innovations variance is relatively small. The samples of the simulated processes are $T = 100$, $T = 300$ and $T = 1000$. From Table 6 we see that the standard tests are again unable to control the error of first kind when the innovations volatility is relatively large beside the autoregressive parameter even when the samples become large. It appears that the modified tests converge to the 5% asymptotic nominal level in such case. We also note that the results we obtain when the innovations variance is large and the autoregressive parameter is close to zero are in general similar to the results of the case where the innovations variance

TABLE 6

Empirical size (in %) of the tests under comparison for testing linear non causality in mean in 5-dimensional cases taking different variance and autoregressive parameter specifications with autoregressive order $p_0 = 1$

	$T = 100$		$T = 300$		$T = 1000$	
	$a = 0.1$ $\omega = 2$	$a = 0.6$ $\omega = 0.1$	$a = 0.1$ $\omega = 2$	$a = 0.6$ $\omega = 0.1$	$a = 0.1$ $\omega = 2$	$a = 0.6$ $\omega = 0.1$
LM_s	**8.1**	**7.7**	**11.2**	**10.8**	**9.9**	**9.5**
LR_s	**12.3**	**12.1**	**12.8**	**11.8**	**10.4**	**9.7**
W_s	**13.0**	**12.2**	**12.9**	**11.8**	**10.6**	**9.6**
LM_{ms}	**45.3**	**48.4**	**10.3**	**10.1**	**6.9**	**7.3**
W_{ms}	**46.3**	**49.0**	**12.2**	**11.2**	**7.3**	**7.4**
$\widetilde{LM}_{ms}$	**9.3**	**12.1**	**8.0**	**8.6**	6.3	**6.6**
$\widetilde{W}_{ms}$	**13.8**	**16.3**	**9.4**	**9.7**	**6.9**	**7.0**
LM_{md}	**30.3**	**29.3**	5.3	6.2	4.8	5.7
LR_{md}	**32.5**	**32.0**	6.2	**8.2**	5.5	5.9
W_{md}	**32.6**	**32.3**	**6.6**	**8.0**	5.5	5.9
$\widetilde{LM}_{md}$	3.9	5.1	5.8	6.2	5.6	5.3
$\widetilde{LR}_{md}$	**7.4**	**9.4**	**6.6**	**8.0**	6.1	5.9
$\widetilde{W}_{md}$	**7.8**	**9.6**	**6.9**	**8.2**	6.1	6.2

TABLE 7

Empirical size (in %) of the tests under comparison for testing linear non causality in mean when the best predictor is not linear. The error process is given in (5.2)

	$T = 100$	$T = 300$	$T = 1000$	$T = 2000$
LM_s	**17.5**	**17.7**	**18.9**	**17.9**
LR_s	**18.3**	**17.9**	**19.0**	**17.9**
W_s	**18.6**	**17.8**	**19.0**	**17.9**
LM_{ms}	**8.7**	**6.6**	4.8	5.4
W_{ms}	**9.8**	**7.1**	4.8	5.4
$\widetilde{LM}_{ms}$	**9.0**	**8.9**	**7.2**	**8.7**
$\widetilde{W}_{ms}$	**10.6**	**9.7**	**7.3**	**8.7**
LM_{md}	**8.7**	**6.6**	4.8	5.4
LR_{md}	**9.9**	**7.1**	4.8	5.4
W_{md}	**9.8**	**7.1**	4.8	5.4
$\widetilde{LM}_{md}$	**9.0**	**8.9**	**7.2**	**8.7**
$\widetilde{LR}_{md}$	**10.2**	**9.5**	**7.3**	**8.7**
$\widetilde{W}_{md}$	**10.6**	**9.7**	**7.3**	**8.7**

is relatively small when compared to the autoregressive order. Therefore we can see that the unconditional noise variance does not play a major role in the behaviour of the modified and standard tests.

In the previous experiments the weak error process given in (5.1) is such that the simplification (3.7) of the matrix Ξ holds. This could explain the relative efficiency of the tests based on the White estimation method to control the error of first kind. We study the case where the best predictor is not linear by considering bivariate AR(1) processes generated using parameter (c) with error processes following the DGP (5.2). We test the null hypothesis that (X_{2t}) does not linearly cause (X_{1t}) in mean. The results are given in Table 7. When the samples are small we remark that the tests are oversized. This can be explained

Table 8

Empirical power (in %) of the different tests with $p_0 = 1$, $d = 2$. The innovations are iid on the left and follow an ARCH model on the right

a	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4	0.5
LM_s	20.1	58.7	88.0	97.8	14.0	41.7	70.3	89.8	96.4
LR_s	20.2	58.4	87.9	97.8	13.6	41.9	70.3	89.6	96.3
W_s	20.0	58.2	87.8	97.7	13.9	41.8	70.3	89.6	96.3
LM_{ms}	20.4	57.6	85.6	97.2	15.6	39.8	67.5	86.1	94.3
W_{ms}	20.5	58.0	85.9	97.4	15.8	40.3	68.9	86.7	95.0
$\widetilde{LM}_{ms}$	19.8	57.0	85.6	97.4	13.5	39.7	68.2	86.8	94.9
$\widetilde{W}_{ms}$	19.5	56.8	85.8	97.4	13.0	39.4	67.9	87.2	94.9
LM_{md}	20.1	58.7	88.0	97.8	14.0	41.7	70.3	89.8	96.4
LR_{md}	20.2	58.4	87.9	97.8	13.6	41.9	70.3	89.6	96.3
W_{md}	20.0	58.2	87.8	97.7	13.9	41.8	70.3	89.6	96.3
$\widetilde{LM}_{md}$	20.1	58.7	88.0	97.8	14.0	41.7	70.3	89.8	96.4
$\widetilde{LR}_{md}$	20.2	58.4	87.9	97.8	13.6	41.9	70.3	89.6	96.3
$\widetilde{W}_{md}$	20.0	58.2	87.8	97.7	13.9	41.8	70.3	89.6	96.3

by the fact that the standard tests and the tests based on the White estimation method are not intended to take into account this kind of situations. When the samples are small ($T = 100$) the more sophisticated tests based on the HAC estimation are likely to not control well the error of first kind. However we note that when the samples increase, the relative rejections of the tests based on the White estimation does not converge to the asymptotic nominal level. The relative rejection frequencies of the tests based on the HAC estimation are close to 5% for large samples. This result is not surprising since from (3.6) the more sophisticated HAC estimation method provides a valid theoretical framework for the case of errors which are uncorrelated but not necessarily a martingale difference. In this case the White estimation method have no sound theoretical basis and is likely to lead to substantially distorted tests. As expected in Table 7 the standard tests are oversized even when the samples are large.

A further set of Monte Carlo experiments has been conducted to study the ability of the tests under comparison to detect linear causality in mean in different situations. To this aim we compare size-adjusted powers of the tests using only processes of length $T = 100$. Results not reported here show that when the samples are large the different tests have the same power. The errors are iid Gaussian with variance I_d in the standard case, and follow an ARCH model given by (5.1) with parameters (k) and (m) in the weak case. In Table 8 we first consider the case of bivariate AR(1) processes generated using parameter (a). We see that the different tests have the same power in this simple case from Table 8. In order to illustrate the case where p_0 is large we considered bivariate AR(3) processes generated by parameter (b) in Table 9. We also study the power of the different tests when the dimension of the observed process is high by considering four dimensional AR(1) processes obtained using parameter (h) in Table 10. In this case the null hypothesis that $(X_{2t}, X_{3t}, X_{4t})'$ is not linearly causal in mean for (X_{1t}) is tested. From Tables 9 and 10 it appears that the tests with modified statistics built using HAC estimators LM_{ms} and W_{ms} are

Table 9

Empirical power (in %) of the different tests with $p_0 = 3$, $d = 2$. The innovations are iid on the left and follow an ARCH model on the right

a	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4	0.5
LM_s	14.1	47.1	80.2	95.7	10.7	33.3	63.4	84.8	94.8
LR_s	14.0	47.1	79.7	95.5	11.2	33.2	63.2	84.6	94.9
W_s	14.0	47.0	79.1	95.2	11.3	32.5	63.0	84.2	94.8
LM_{ms}	8.9	27.2	60.6	84.1	8.1	16.0	25.2	33.2	45.4
W_{ms}	10.5	33.2	67.4	89.5	8.7	18.3	34.6	54.6	73.1
$\widetilde{LM}_{ms}$	12.6	45.8	77.1	94.8	10.7	30.5	60.1	81.7	92.9
$\widetilde{W}_{ms}$	13.0	46.4	77.8	94.7	10.5	31.0	61.3	82.6	94.0
LM_{md}	14.1	47.1	80.2	95.7	10.7	33.3	63.4	84.8	94.8
LR_{md}	14.0	47.1	79.7	95.5	11.2	33.2	63.2	84.6	94.9
W_{md}	14.0	47.0	79.1	95.2	11.3	32.5	63.0	84.2	94.8
$\widetilde{LM}_{md}$	14.1	47.1	80.2	95.7	10.7	33.3	63.4	84.8	94.8
$\widetilde{LR}_{md}$	14.0	47.1	79.7	95.5	11.2	33.2	63.2	84.6	94.9
$\widetilde{W}_{md}$	14.0	47.0	79.1	95.2	11.3	32.5	63.0	84.2	94.8

Table 10

Empirical power (in %) of the different tests with $p_0 = 1$, $d = 4$ and iid innovations. The innovations are iid on the left and follow an ARCH model on the right

a	0.1	0.15	0.2	0.3	0.1	0.15	0.2	0.3
LM_s	35.5	69.1	90.2	99.9	28.5	61.2	84.0	98.5
LR_s	35.2	69.5	90.2	99.9	28.7	61.6	83.8	98.4
W_s	35.6	70.2	91.0	99.9	28.5	61.8	83.9	98.3
LM_{ms}	17.4	32.8	44.0	60.3	16.3	26.0	38.4	53.6
W_{ms}	20.2	38.6	56.6	87.5	18.2	32.1	50.4	78.7
$\widetilde{LM}_{ms}$	29.3	60.3	87.3	99.5	26.5	53.1	79.0	97.7
$\widetilde{W}_{ms}$	30.2	61.7	87.9	99.6	27.3	55.2	80.4	98.2
LM_{md}	35.5	69.1	90.2	99.9	28.5	61.2	84.0	98.5
LR_{md}	35.2	69.5	90.2	99.9	28.7	61.6	83.8	98.4
W_{md}	35.6	70.2	91.0	99.9	28.5	61.8	83.9	98.3
$\widetilde{LM}_{md}$	35.5	69.1	90.2	99.9	28.5	61.2	84.0	98.5
$\widetilde{LR}_{md}$	35.2	69.5	90.2	99.9	28.7	61.6	83.8	98.4
$\widetilde{W}_{md}$	35.6	70.2	91.0	99.9	28.5	61.8	83.9	98.3

less powerful than the other tests. Finally we also note that in general the use of the LM_{md} and $\widetilde{LM}_{md}$ tests does not lead to a loss of power, when compared to the other tests build in a similar way.

Some additional remarks on the results of the Monte Carlo experiments have to be made. The Wald type tests are widely used to test the linear causality in mean. However it emerges that the Lagrange Multiplier tests have a better control of the error of first kind under the null, and do not suffer of loss of power under the alternative when compared to the Wald tests. The better efficiency of the LM tests is particulary marked when the number of parameters to be estimated is high and can be explained by the fact that the constrained estimators are used in the test statistic. Note also that we found similar results (not reported here) to those given in Table 3 and 4 when the autoregressive order used for the estimation of the VAR model is larger than the true autoregressive order p_0. The choice of the correct lag length is important for the study of linear Granger causality in mean relationship between variables, as pointed out by [35] and [23]. Therefore if the errors are suspected to be weak, one should use adapted tools for the choice of the autoregressive order, as for instance the modified portmanteau tests proposed in [15], for a gain of efficiency when testing the linear causality in mean. Nevertheless [15] pointed out that the use of the standard portmanteau tests may lead to an overparameterization of the weak VAR model. Thus if one use the standard tools for checking the autoregressive order in the weak VAR framework, it emerges from our simulation results that it is preferable to use the $\widetilde{LM}_{md}$ test when the sample is small and the LM_{md} when the sample is large.

6. Illustrative example

We consider an application to the daily log returns USD/BP and USD/NZD from January 2, 1998 to September 4, 2008 to illustrate the robustness of the different tests to the presence of nonlinearities. The length of the series is $T = 2688$. The analyzed data are plotted in Figure 2. Note that the linear causality in mean concept is much used to analyze this kind of data. For instance [24] test linear causality in mean of exchange rates and stock prices to investigate the sources of the 1997 asian financial crisis. We test the null hypothesis that USD/BP does not linearly cause in mean USD/NZD. For this purpose a VAR(1) model is adjusted to the series. We get the following estimates of the parameters

$$\begin{pmatrix} nz_t \\ bp_t \end{pmatrix} = \begin{pmatrix} -0.028_{[0.06]} & 0.073_{[0.06]} \\ -0.001_{[0.01]} & 0.016_{[0.02]} \end{pmatrix} \begin{pmatrix} nz_{t-1} \\ bp_{t-1} \end{pmatrix} + \begin{pmatrix} \hat{\epsilon}_{1t} \\ \hat{\epsilon}_{2t} \end{pmatrix} \qquad (6.1)$$

where the USD/NZD and USD/BP at the date t are respectively denoted by nz_t and bp_t. The standard deviations of the estimates computed using the result (3.4) are into brackets. Here we can note that the estimated parameters are not significantly different from zero.

We use standard Box-Pierce and Ljung-Box portmanteau tests to check the adequacy of the VAR(1) model (6.1). Modified Box-Pierce and Ljung-Box port-

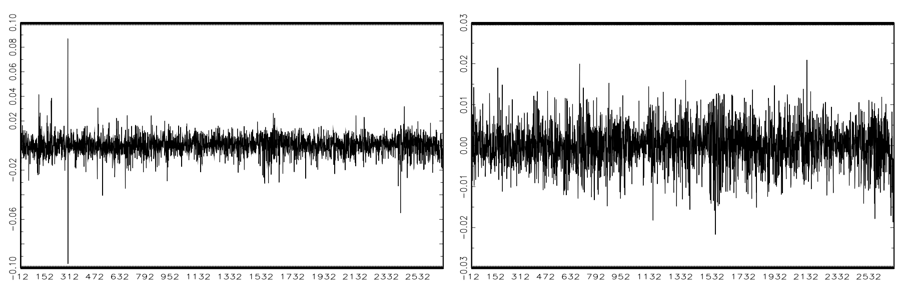

Fig 2. *The daily log returns of the exchange rates of USD to one NZD on the left, and USD to one BP on the right from 01/02/1998 to 09/04/2008 (T=2688). Data source: The research division of the federal reserve bank of St. Louis, www.research.stlouisfed.org.*

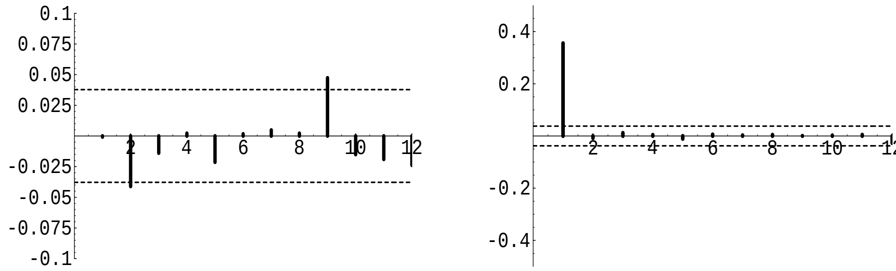

FIG 3. *Autocorrelations of the first component of the residuals of the VAR(1) model adjusted to the analyzed data. The left graphic represent the autocorrelations of the residuals and the right the autocorrelations of the squared residuals. The horizontal lines about zero represent the approximate 5% significance limits for the sample autocorrelations (that is* $\pm 1.96/\sqrt{T}$ *with* $T = 2688$*).*

TABLE 11

The p-values (in%) of the modified and standard causality tests obtained for the test of the linear causality in mean from the returns of daily the exchange rates of USD to one BP to the returns of the daily exchange rates of USD to one NZD

LR_s	2.05
LR_{md}	18.79
$\widetilde{LR}_{md}$	16.50
W_s	3.21
W_{ms}	22.32
W_{md}	22.32
$\widetilde{W}_{ms}$	19.91
$\widetilde{W}_{md}$	19.91
LM_s	1.23
LM_{ms}	15.49
LM_{md}	15.49
$\widetilde{LM}_{ms}$	13.37
$\widetilde{LM}_{md}$	13.37

manteau tests developed in the framework of weak VAR models in [15] are also considered. Since the results for the Box-Pierce and Ljung-Box tests are similar, we only reported the results for the Box-Pierce tests. From Table 13 we can remark that the p-values of the different portmanteau tests are far from zero. We also considered the autocorrelations of the residuals. In order to save space, we only plotted the autocorrelations of the first component of the residuals in Figure 3. We can remark that the autocorrelations are inside or not much larger than the 5% significance limits. Then from these results, it appears that the weak VAR(1) model cannot be rejected. However considering the first autocorrelation of the first component of the squared residuals in Figure 3, the standard iid assumption for the error process have to be rejected. This is in accordance with the fact that it is commonly admitted that financial series often exhibit nonlinearities.

Now we turn to the study of the linear causality in mean for the analyzed data. The different causality tests considered in this paper are applied. Table 11

Table 12
The tests statistics obtained using the analyzed data

LR test statistic Q_{LR}	5.37
Standard Wald test statistic Q_{WS}	4.60
Modified Wald test statistic Q_W	1.48
Modified Wald test statistic $\tilde{Q}_W$	1.65
Standard LM test statistic Q_{LMS}	6.27
Modified LM test statistic Q_{LM}	2.02
Modified LM test statistic $\tilde{Q}_{LM}$	2.25

Table 13
The p-values (in %) and test statistics of the modified and standard Box-Pierce portmanteau tests for the VAR(1) model adjusted to the analyzed data. The number of autocorrelations used is m

	$m = 3$	$m = 6$	$m = 12$	$m = 18$
Box-Pierce statistics	13.54	25.39	49.90	75.40
p-value of the standard Box-Pierce test	9.46	18.69	25.04	25.14
p-value of the modified Box-Pierce test	11.97	23.35	29.83	30.37

displays the p-values of the tests. First remark that the p-values of the modified tests are far from zero. According to these results, the hypothesis that USD/BP does not linearly cause in mean USD/ NZD cannot be rejected. Note that the order q given in (3.6) is chosen equal to one using the AIC criterion. This explain that the p-values of the tests based on the HAC estimation are different from those of the tests based on the White estimation. However since the p-values of the standard tests are very small, the null hypothesis is clearly rejected by the LM_s, LR_s and W_s tests. For the W_{ms} and LM_{ms} tests these contradictory results can be explained by the fact that the modified and standard statistics are very different in Table 12. Similar comments can be made for the tests with modified distribution when comparing $\hat{\Lambda}\hat{\Xi}\hat{\Lambda}$, $\hat{\Lambda}\tilde{\Xi}\hat{\Lambda}$ and $\hat{J}^{-1}$ (not reported here). In view of the results of our Monte Carlo experiments it seems likely that the standard tests detected the nonlinear dynamics of the series as linear causality in mean. In addition note that the log returns of such systems of exchange rates should be uncorrelated in theory.

7. Conclusion

In this paper we considered the problem of testing the linear causality in mean in situations where the dimension is high or the autoregressive order is large. In general the practitioners consider the standard Wald test for the analysis of linear causality in mean when the errors are suspected to be iid. Wald tests with corrected statistics are used if one suspect the presence of nonlinearities in the error process. It appears that the standard Wald test poorly perform when the number of estimated parameters is high. Hence we considered standard LM tests and found that such tests improve the analysis of the linear causality in mean. However in accordance to with earlier studies, we also found that the standard tests are unable to distinguish between nonlinear temporal dynamics and linear causality in mean. Concerning the Wald tests with modified statistics, it also emerges that this kind of tests poorly perform when the dimension is high or the autoregressive order is large in small samples. We proposed LM tests with modified statistics which give some improvement when compared to the Wald tests with modified statistics. Nevertheless it appears that the use of corrected statistics does not provide satisfactory results when the number of parameters is high. Therefore Wald, LM and LR tests with standard statistics but modified critical values evaluated using HAC and White variance corrections were introduced. It is found that the tests with modified distribution have better results than the tests with modified statistics when the dimension is high or the autoregressive order is large. For this kind of tests we also found that the LM tests give significant improvements when compared to the Wald and LR tests. This can be explained by the fact that when the dimension is high or the autoregressive order is large, the number of parameters to estimate is relatively high for the Wald and LR tests, while we only use the constrained estimator for the LM test. Note also that the tests with modified distribution are built taking into account for nonlinear dynamics in the error process contrary to the standard tests. Thus when the sample is small we recommend to use the standard LM test in the iid case when the autoregressive order is large or when the dimension is high. If the existence of temporal dynamics is suspected and the sample is small, it is preferable to use the LM test with modified distribution based on the White variance correction due to its relative simplicity. However the more sophisticated tests based on HAC correction have a larger theoretical basis than the standard tests or the tests based on the White estimation method. Then it is preferable to use the LM test based on HAC correction with modified distribution when we have a large sample. More generally note that practitioners are likely to use the Wald test to analyze parameter restrictions in models. Similarly corrections on the test statistics are commonly used when needed. Therefore when applied to other topics, the approaches presented in this paper can potentially give improvements for the analysis of time series.

Appendix

Proof of Theorem 4.1. To prove the result concerning the statistic Q_{LMS}, using a Taylor expansion about θ_0 and from the consistency of $\hat{\Sigma}_\epsilon$, we write

$$0 = T^{-\frac{1}{2}} \frac{\partial \mathcal{L}(\hat{\theta}, \hat{\Sigma}_\epsilon)}{\partial \theta} = T^{-\frac{1}{2}} \frac{\partial \mathcal{L}(\theta_0, \Sigma_\epsilon)}{\partial \theta} - J T^{\frac{1}{2}} (\hat{\theta} - \theta_0) + o_p(1). \tag{A.1}$$

Similarly under the null hypothesis we have

$$T^{-\frac{1}{2}} \frac{\partial \mathcal{L}(\hat{\theta}^c, \hat{\Sigma}_\epsilon^c)}{\partial \theta} = T^{-\frac{1}{2}} \frac{\partial \mathcal{L}(\theta_0, \Sigma_\epsilon)}{\partial \theta} - J T^{\frac{1}{2}} (\hat{\theta}^c - \theta_0) + o_p(1), \tag{A.2}$$

Note that $\hat{\theta}^c$ and $\hat{\Sigma}_\epsilon^c$ maximize the expression

$$\mathcal{L}(\theta, \Sigma) - \lambda' R \theta,$$

where λ is a vector of dimension pd_1d_2 (see [27], p 671). First order condition give

$$R'\hat{\lambda} = \frac{\partial \mathcal{L}(\hat{\theta}^c, \hat{\Sigma}_\epsilon^c)}{\partial \theta} \quad \text{and} \quad R\hat{\theta}^c = 0. \tag{A.3}$$

In addition under the null hypothesis we have $R\hat{\theta}^c = R\theta_0 = 0$, so that from (3.4) we write

$$T^{\frac{1}{2}} R\{\hat{\theta} - \hat{\theta}^c\} = T^{\frac{1}{2}} R\{\hat{\theta} - \theta_0\} \Rightarrow \mathcal{N}(0, RJ^{-1}IJ^{-1}R'). \tag{A.4}$$

Using (A.3) and subtracting (A.2) from (A.1) we obtain

$$RJ^{-1}R'T^{-\frac{1}{2}}\hat{\lambda} = RJ^{-1}T^{-\frac{1}{2}}\frac{\partial \mathcal{L}(\hat{\theta}^c, \hat{\Sigma}_\epsilon^c)}{\partial \theta} = T^{\frac{1}{2}}R(\hat{\theta} - \hat{\theta}^c) + o_p(1). \tag{A.5}$$

From (A.3) we also have $RR'\hat{\lambda} = \hat{\lambda} = R\frac{\partial \mathcal{L}(\hat{\theta}^c, \hat{\Sigma}_\epsilon^c))}{\partial \theta}$. Then using (A.4) and (A.5) it follows that

$$T^{-\frac{1}{2}} R\frac{\partial \mathcal{L}(\hat{\theta}^c, \hat{\Sigma}_\epsilon^c)}{\partial \theta} = T^{-\frac{1}{2}}\hat{\lambda} \Rightarrow \mathcal{N}(0, (RJ^{-1}R')^{-1}(RJ^{-1}IJ^{-1}R')(RJ^{-1}R')^{-1}), \tag{A.6}$$

so that writing

$$T^{-\frac{1}{2}}(RJ^{-1}R')^{\frac{1}{2}}R\mathcal{S} = T^{-\frac{1}{2}}\hat{\lambda} \Rightarrow \mathcal{N}(0, (RJ^{-1}R')^{-\frac{1}{2}}(RJ^{-1}IJ^{-1}R')(RJ^{-1}R')^{-\frac{1}{2}})$$

we obtain the result for the Q_{LMS} statistic. The proof of the assertion concerning the statistic Q_{WS} is a straightforward consequence of (3.4). To prove the result concerning the statistic Q_{LR}, using again a Taylor expansion around θ_0 we write

$$\mathcal{L}(\hat{\theta}, \hat{\Sigma}_\epsilon) = \mathcal{L}(\theta_0, \Sigma_\epsilon) + \frac{\partial \mathcal{L}(\theta_0, \Sigma_\epsilon)}{\partial \theta'}(\hat{\theta} - \theta_0) - \frac{T}{2}(\hat{\theta} - \theta_0)'J(\hat{\theta} - \theta_0) + o_p(1), \tag{A.7}$$

and

$$\mathcal{L}(\hat{\theta}^c, \hat{\Sigma}_\epsilon) = \mathcal{L}(\theta_0, \Sigma_\epsilon) + \frac{\partial \mathcal{L}(\theta_0, \Sigma_\epsilon)}{\partial \theta'}(\hat{\theta}^c - \theta_0) - \frac{T}{2}(\hat{\theta}^c - \theta_0)'J(\hat{\theta}^c - \theta_0) + o_p(1). \tag{A.8}$$

Under the null hypothesis and subtracting (A.8) from (A.7) we obtain

$$\begin{aligned} Q_{LR} &= 2\left\{\mathcal{L}(\hat{\theta}, \hat{\Sigma}_\epsilon) - \mathcal{L}(\hat{\theta}^c, \hat{\Sigma}_\epsilon^c)\right\} = 2\frac{\partial \mathcal{L}(\theta_0, \Sigma_\epsilon)}{\partial \theta'}(\hat{\theta} - \hat{\theta}^c) \\ &- T(\hat{\theta} - \theta_0)'J(\hat{\theta} - \theta_0) + T(\hat{\theta}^c - \theta_0)'J(\hat{\theta}^c - \theta_0) + o_p(1). \end{aligned}$$

Recall that from (A.1) we have

$$T^{-\frac{1}{2}}\frac{\partial \mathcal{L}(\theta_0, \Sigma_\epsilon)}{\partial \theta} = T^{\frac{1}{2}}J(\hat{\theta} - \theta_0) + o_p(1),$$

so that we write

$$\begin{aligned} Q_{LR} &= 2\left\{\mathcal{L}(\hat{\theta}, \hat{\Sigma}_\epsilon) - \mathcal{L}(\hat{\theta}^c, \hat{\Sigma}_\epsilon^c)\right\} = 2T(\hat{\theta} - \theta_0)'J(\hat{\theta} - \hat{\theta}^c) \\ &- T(\hat{\theta} - \theta_0)'J(\hat{\theta} - \theta_0) + T(\hat{\theta}^c - \theta_0)'J(\hat{\theta}^c - \theta_0) + o_p(1). \end{aligned} \tag{A.9}$$

Then writing the last term of (A.9) as follows

$$(\hat{\theta}^c - \theta_0)' J(\hat{\theta}^c - \theta_0) = (\hat{\theta}^c - \hat{\theta} + \hat{\theta} - \theta_0)' J(\hat{\theta}^c - \hat{\theta} + \hat{\theta} - \theta_0)$$

we obtain

$$Q_{LR} = T(\hat{\theta} - \hat{\theta}^c)' J(\hat{\theta} - \hat{\theta}^c).$$

Similarly to (A.5) we can write

$$J^{-1} R' \hat{\lambda} = T(\hat{\theta} - \hat{\theta}^c) + o_p(1).$$

so that we have

$$Q_{LR} = T^{-1} \hat{\lambda}' R J^{-1} R' \hat{\lambda}.$$

Then using (A.6) we obtain the result (4.1). □

Proof of Theorem 4.2. The assertion concerning the statistic Q_W is a straightforward consequence of (3.4). From (A.6) it easy to see that the result concerning the statistic Q_{LM} hold. □

References

[1] Amendola, A. and Francq, C. , (2009). Concepts and tools for nonlinear time series modelling. Handbook of Computational Econometrics, (eds: D. Belsley and E. Kontoghiorghes, Wiley.

[2] Andrews, B., Davis, R.A. and Breidt, F.J., (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis*, 97, 1638-1659. MR2256234

[3] Andrews, D.W.K., and Monahan, J.C., (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60, 953-966. MR1168742

[4] Baek, E., and Brock, W., (1992). A general test for nonlinear Granger causality: bivariate model. Working paper. http://www.ssc.wisc.edu/~wbrock/Baek Brock Granger.pdf.

[5] Bauwens, L., Laurent, S. and Rombouts, J.V.K., (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, 21, 79-109. MR2225523

[6] Bertholon, H., Monfort, A. and Pegoraro, F., (2008). Econometric asset pricing modelling. *Journal of Financial Econometrics*, 6, 407-458.

[7] Breidt, B., Davis, R.A. and Trindade, F.J., (2001). Least absolute deviation estimation for all-pass time series models. *The Annals of Statistics*, 29, 919-946. MR1869234

[8] Bollerslev, T., Chou, R.Y. and Kroner, K.F., (1992). ARCH modeling in finance. *Journal of Econometrics*, 52, 5-59.

[9] Caporale, G.M. and Howells, P., (2001). Money, credit and spending: drawing causal inferences. *Scottish Journal of Political Economy*, 48, 547-557.

[10] Carrasco, M. and Chen, X., (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18, 17-39. MR1885348
[11] Cheung, Y-W. and Ng, L.K., (1996). A causality-in-variance test and its application to financial market prices. *Journal of Econometrics*, 72, 33-48. MR1396899
[12] Chow, W.W. and Fung, M.K., (2008). Volatility of stock price as predicted by patent data: An MGARCH perspective. *Journal of Empirical Finance*, 15, 64-79.
[13] Den Haan, W.J. and Lievin, A., (1997). *A practioner's guide to robust covariance matrix estimation*. Handbook of Statistics 15, Chapter 12, (eds: G.S. Maddala and C.R. Rao) Amsterdam: Elsevier, 299-342.
[14] Feige, E.L. and Pearce, D.K., (1979). The casual causal relationship between money and income: some caveats for time series analysis. *Review of Economics and Statistics*, 61, 21-33.
[15] Francq, C. and Raïssi, H., (2007). Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors. *Journal of Time Series Analysis*, 28, 454-470. MR2355317
[16] Francq, C., Roy, R. and Zakoïan, J-M., (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of American Statistical Association*, 100, 532-544. MR2160557
[17] Francq, C. and Zakoïan, J-M., (2007). HAC estimation and strong linearity testing in weak ARMA models. *Journal of Multivariate Analysis*, 98, 114-144. MR2292920
[18] Granger, C.W.J., (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 12, 424-438.
[19] Granger, C.W.J., Huang, B. and Yang, C., (2000). A bivariate causality between stock prices and exchange rates: evidence from recent asian flu. *Quaterly Review of Economics and Finance*, 40, 337-357.
[20] Hiemstra, C. and Jones, J.D., (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance*, 14, 1639-1664.
[21] Imhof, J. P., (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48, 419-426. MR0137199
[22] Jeantheau, T., (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, 14, 70-86. MR1613694
[23] Jones, J.D., (1989). A comparison of lag-length selection techniques in tests of Granger causality between money growth and inflation: evidence for the US, 1959-86. *Applied Economics*, 21, 809-822.
[24] Khalid, A.M. and Kawai, M. (2003). Was financial market contagion the source of economic crisis in Asia? Evidence using a multivariate VAR model. *Journal of Asian Economics*, 14, 131-156.
[25] Koutmos, G. and Booth, G.G., (1995). Asymmetric volatility transmission in international stock markets. *Journal of International Money and Finance*, 14, 747-762.

[26] Kuonen, D., (1999). Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika*, 86, 929-935. MR1741988

[27] Lütkepohl, H., (2005). *New introduction to multiple time series analysis*. Springer, Berlin. MR2172368

[28] Lütkepohl, H., (1982). Non-causality due to omitted variables. *Journal of Econometrics*, 19, 367-378.

[29] Nakatani, T. and Teräsvirta, T., (2009). Testing for volatility interactions in the constant conditional correlation GARCH model. *Econometrics Journal*, 12, 147-163. MR2500201

[30] Pantelidis, T. and Pittis, N., (2004). Testing for Granger causality in variance in the presence of causality in mean. *Economics Letters*, 85, 201-207. MR2149233

[31] Pham, D.T., (1986). The mixing property of bilinear and generalized random coefficient autoregressive models. *Stochastic Processes and their Applications*, 23, 291-300.

[32] Romano, J. P. and Thombs, L. A., (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association*, 91, 590-600. MR1395728

[33] Sims, C.A., (1972). Money income and causality. *American Economic Review*, 62, 540-542.

[34] Stock, J.H. and Watson, M.W., (1989). Interpreting the evidence on money-income causality. *Journal of Econometrics*, 40, 161-181.

[35] Thornton, D.L. and Batten, D.S., (1985). Lag-length selection and tests of Granger causality between money and income. *Journal of Money, Credit, and Banking*, 17, 164-178.

[36] Vilasuso, J., (2001). Causality tests and conditional heteroskedasticity: Monte Carlo evidence. *Journal of Econometrics*, 101, 25-35. MR1817870

[37] Warne, A., (2000). Causality and regime inference in a markov switching VAR. Working paper No 118, Research Department, Sveriges Riksbank. http://www.riksbank.com/templates/Page.aspx?id=25439.

[38] White, H., (1980). A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817-838. MR0575027