Article

Testing for A Set of Linear Restrictions in VARMA Models Using Autoregressive Metric: An Application to Granger Causality Test

Francesca Di Iorio [1],* and Umberto Triacca [2]

[1] Department of Political Science, University of Naples Federico II, via L. Rodinò 22 ,
  I-80138 Naples, Italy

[2] Department of Computer Engineering, Computer Science and Mathematics, University of L'Aquila,
  via Vetoio, I-67010 Coppito, L'Aquila, Italy; E-Mail: umberto.triacca@ec.univaq.it

* Author to whom correspondence should be addressed; E-Mail: fdiiorio@unina.it;
  Tel.: +39-06-4456100.

External Editor: Kerry Patterson

Abstract: In this paper we propose a test for a set of linear restrictions in a Vector Autoregressive Moving Average (VARMA) model. This test is based on the autoregressive metric, a notion of distance between two univariate ARMA models, M_0 and M_1, introduced by Piccolo in 1990. In particular, we show that this set of linear restrictions is equivalent to a null distance $d(M_0, M_1)$ between two given ARMA models. This result provides the logical basis for using $d(M_0, M_1) = 0$ as a null hypothesis in our test. Some Monte Carlo evidence about the finite sample behavior of our testing procedure is provided and two empirical examples are presented.

---

1. Introduction

In this paper, we investigate the relationship between a set of linear restrictions on the parameters of a Vector Autoregressive Moving Average (VARMA) model (see [1]) and the autoregressive metric

(AR-metric hereafter), a notion of the distance between two univariate ARMA models introduced by Piccolo [2]. In particular, we show that these linear restrictions are satisfied if and only if the distance d between the two given ARMA models (say M_0 and M_1) is zero. This result provides the logical basis for using $d(M_0, M_1) = 0$ as null hypothesis for testing this set of restrictions. Moreover, we show that the set of linear restrictions considered is sufficient for the condition of Granger noncausality ([3]), while in the VAR framework it becomes also a necessary condition (see [4]). This theoretical result allows the implementation of an inferential procedure and a bootstrap algorithm. Our procedure is verified by some Monte Carlo experiments also in a quite small sample. The paper is organized as follows. Section 2 introduces the notion of the distance between ARMA models and specifies the relationship between the AR metric and the set of linear restrictions considered for a VARMA model. Section 3 presents the inferential implication. Section 4 provides some Monte Carlo evidence about the finite sample behavior of our testing procedure. Section 5 contains two empirical illustrations. Section 6 gives some concluding remarks.

2. Linear Restrictions in a VARMA Model and AR-Metric

Let z_t be a zero mean invertible ARMA model defined as

$$\phi(L)z_t = \theta(L)\epsilon_t$$

where $\phi(L)$ and $\theta(L)$ are polynomials in the lag operator L, with no common factors, and ϵ_t is a white noise process with constant variance σ^2. It is well-known that this process admits the following representation:

$$\pi(L)z_t = \epsilon_t$$

where the AR(∞) operator is defined by

$$\pi(L) = \phi(L)\theta(L)^{-1} = 1 - \sum_{i=1}^{\infty} \pi_i L^i$$

with $\sum_{i=1}^{\infty} |\pi_i| < \infty$.

Let ℓ be the class of ARMA invertible models. If $X \in \ell$ and $Y \in \ell$, following Piccolo [2], the AR-metric is defined as the Euclidean distance between the corresponding π-weights sequence, $\{\pi_j\}$,

$$d(X, Y) = \left[\sum_{i=1}^{\infty} (\pi_{xi} - \pi_{yi})^2 \right]^{\frac{1}{2}} \tag{1}$$

The AR-metric d has been widely used in time series analysis (see, e.g., [5–10]). We observe that Equation (1) is a well-defined measure because of the absolute convergence of the π-weights sequences.

Now, we consider the following VARMA model of order p, q, for an $n \times 1$ vector time series $\{w_t; \ t \in \mathbb{Z}\}$:

$$A(L)w_t = B(L)\epsilon_t \tag{2}$$

where $A(L) = I_n - A_1 L - A_2 L^2 - \cdots - A_p L^p$ and $B(L) = I_n - B_1 L - B_2 L^2 - \cdots - B_q L^q$ are two $n \times n$ matrices of polynomials in the lag operator L, and ϵ_t is an $n \times 1$ vector white noise process with

positive definite covariance matrix Σ. We assume that $\det(A(z)) \neq 0$ for $|z| < 1$. This condition allows non-stationarity for the series, in the sense that the characteristic polynomial of the VARMA model described by the equation $\det(A(z)) = 0$ may have roots on the unit circle. Condition $\det(A(z)) \neq 0$ for $|z| < 1$, however, excludes explicitly explosive processes from our consideration. We further assume that the model Equation (2) satisfies the usual identifiability conditions. If $B(L) = I$, we obtain a pure vector autoregressive (VAR) model of order p. If $A(L) = I$, we obtain a pure vector moving average (VMA) model of order q. Consider the partition $w_t = (x_t, y_t')'$ where x_t is a scalar time series and y_t is an $(n-1) \times 1$ vector of time series. Accordingly, the model Equation (2) for the partition of w_t can be rewritten as:

$$\begin{bmatrix} 1 - A_{11}(L) & A_{12}(L) \\ A_{21}(L) & I - A_{22}(L) \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1 - B_{11}(L) & B_{12}(L) \\ B_{21}(L) & I - B_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{x_t} \\ \epsilon_{y_t} \end{bmatrix} \tag{3}$$

$$E\left(\begin{bmatrix} \epsilon_{x_t} \\ \epsilon_{y_t} \end{bmatrix} \begin{bmatrix} \epsilon_{x_s} & \epsilon_{y_s} \end{bmatrix} \right) = \begin{cases} \Sigma & t = s \\ \mathbf{0} & t \neq s \end{cases}$$

where $A_{ij}(L) = \sum_{h=1}^{p} A_{ij}^{(h)} L^h$ and $B_{ij}(L) = \sum_{h=1}^{q} B_{ij}^{(h)} L^h$ $i, j = 1, 2$ are matrix polynomials in the lag operator L, with $\det(A_{22}(L)) \neq 0$. In this framework it is well-known (see, for example, [11]) that y_t does not Granger-cause x_t if and only if

$$B_{12}(L) - A_{12}(L)A_{22}(L)^{-1}B_{22}(L) = \mathbf{0} \tag{4}$$

and that a sufficient condition for Equation (4) to hold is

$$A_{12}(L) = B_{12}(L) = \mathbf{0} \tag{5}$$

We note that if the condition Equation (5) holds then x_t follows a univariate ARMA model given by:

$$[1 - A_{11}(L)] x_t = [1 - B_{11}(L)] \epsilon_{x_t} \tag{6}$$

The main aim of this paper is to establish the implications of the set of linear restrictions Equation (5), using the notion of the distance between ARMA models measured by Equation (1). In particular, we will consider the distance between the ARMA(p, q) model Equation (6) (denoted M_0) and the ARMA model for the subprocess $\{x_t; t \in \mathbb{Z}\}$ implied by the VARMA(p, q) model Equation (2) (denoted M_1).

Following Lütkepohl [1], the implied ARMA model M_1 can be obtained as follows. Premultiplying both sides of Equation (2) by the adjoint of $A(L)$, denoted as $\text{Adj}(A(L))$, we obtain

$$\det(A(L)) w_t = \text{Adj}(A(L)) B(L)\epsilon_t \tag{7}$$

We note that each component of $\text{Adj}(A(L)) B(L)\epsilon_t$ is a sum of finite order MA processes, thus it is a finite order MA process (see Proposition 11.1 in [1]). Hence, the subprocess $\{x_t; t \in Z\}$ follows an ARMA model given by:

$$\det(A(L)) x_t = \delta(L)u_t \tag{8}$$

where u_t is univariate white noise and $\delta(L)$ is an invertible polynomial in the lag operator L. More precisely, $\delta(L)$ and u_t are such that

$$\delta(L)u_t = C_1(L)\epsilon_t$$

where $C_1(L)$ denotes the first row of the matrix $C(L) = \text{Adj}\,(A(L))\,B(L)$. Finally, we observe that x_t has also the following autoregressive representation of infinite order:

$$\varphi(L)x_t = u_t$$

where

$$\varphi(L) = \frac{\det\,[A(L)]}{\delta(L)} = 1 + \varphi_1 L + \varphi_2 L^2 + \dots$$

2.1. Theoretical Results

We consider the distance according to Equation (1) between the model Equations (6) and (8) M_0 and M_1:

$$d(M_0, M_1) = \left[\sum_{i=1}^{\infty} (\lambda_i - \varphi_i)^2\right]^{\frac{1}{2}}$$

where

$$\lambda(L) = \frac{1 - A_{11}(L)}{1 - B_{11}(L)} = 1 + \lambda_1 L + \lambda_2 L^2 + \dots$$

The following proposition provides a necessary and sufficient condition for the set of linear restrictions Equation (5) in terms of the distance $d(M_0, M_1)$.

Proposition 1. $A_{12}(L) = B_{12}(L) = \boldsymbol{0}$ *if and only if* $d(M_0, M_1) = 0$.

Proof of Proposition 1. $(\Rightarrow)$ We have

$$\det\,[A\,(L)] = (1 - A_{11}(L))\det\,\left[I - A_{22}(L) - A_{21}(L)\,(1 - A_{11}(L))^{-1}\,A_{12}(L)\right]$$

and the first row the matrix $C(L)$ is such that $C_1(L) = [C_{11}(L), C_{12}(L)]$ where

$$C_{11}(L) = [\det\,(A\,(L))\,D(L)\,(1 - B_{11}(L)) - \det\,(A\,(L))\,D(L)A_{12}(L)\,(I - A_{22}(L))^{-1}\,B_{21}(L)]$$

and

$$C_{12}(L) = [\det\,(A\,(L))\,D(L)B_{12}(L) - \det\,(A\,(L))\,D(L)A_{12}(L)\,(I - A_{22}(L))^{-1}\,(I - B_{22}(L))]$$

with $D(L) = [1 - A_{11}(L) - A_{12}(L)\,(I - A_{22}(L))^{-1}\,A_{21}(L)]^{-1}$
   If $A_{12}(L) = B_{12}(L) = \boldsymbol{0}$, then

$$\det\,(A(L)) = (1 - A_{11}(L))\det\,(I - A_{22}(L))$$

and

$$C_1(L) = [\det\,(I - A_{22}(L))\,(1 - B_{11}(L)),\ 0]$$

Thus we have that $u_t = \epsilon_{xt}$ (where this equality between random variables means equality with probability 1) and $\delta(L) = \det\,(I - A_{22}(L))\,(1 - B_{11}(L))$. It follows that

$$\varphi(L) = \frac{\det\,(A(L))}{\delta(L)} = \frac{1 - A_{11}(L)}{1 - B_{11}(L)}$$

and hence $d(M_0, M_1) = 0$.

($\Leftarrow$) We have to show that if $d(M_0, M_1) = 0$, then $A_{12}(L) = B_{12}(L) = \mathbf{0}$. We may have two cases: $A_{21}(L) \neq \mathbf{0}$ or $A_{21}(L) = \mathbf{0}$.

First case: $A_{21}(L) \neq \mathbf{0}$.

If $d(M_0, M_1) = 0$, then

$$\varphi(L) = \frac{1 - A_{11}(L)}{1 - B_{11}(L)}$$

On the other hand, we have

$$\varphi(L) = \frac{\det(A(L))}{\delta(L)}$$

and hence

$$\frac{1 - A_{11}(L)}{1 - B_{11}(L)} = \frac{(1 - A_{11}(L)) \det\left(I - A_{22}(L) - A_{21}(L) (1 - A_{11}(L))^{-1} A_{12}(L)\right)}{\delta(L)}$$

Using the Schur's formula, we get

$$\frac{1 - A_{11}(L)}{1 - B_{11}(L)} = \frac{\det(I - A_{22}(L)) \left(1 - A_{11}(L) - A_{12}(L) (I - A_{22}(L))^{-1} A_{21}(L)\right)}{\delta(L)}$$

Thus $\delta(L)$ assume the following expression

$$\delta(L) = \det(C) (1 - B_{11}(L)) - (1 - A_{11}(L))^{-1} \det(C) A_{12}(L) (I - A_{22}(L))^{-1} A_{21}(L) (1 - B_{11}(L)) \tag{9}$$

where $C = I - A_{22}(L)$.

Since the degree of polynomial $\delta(L)$ is finite

$$\deg(\delta(L)) < \infty$$

Equation (9) implies that

$$\deg\left((1 - A_{11}(L))^{-1} \det(I - A_{22}(L)) A_{12}(L) (I - A_{22}(L))^{-1} A_{21}(L) (1 - B_{11}(L))\right) < \infty \tag{10}$$

Since

$$\deg\left((1 - A_{11}(L))^{-1}\right) = \infty$$

it follows for Equation (10) that it must be

$$A_{12}(L) (I - A_{22}(L))^{-1} A_{21}(L) = 0$$

Since by hypothesis $A_{21}(L) \neq 0$, it follows that $A_{12}(L) = 0$ and this in turn implies that

$$C_1(L) = [\det(I - A_{22}(L)) (1 - B_{11}(L)), \det(I - A_{22}(L)) B_{12}(L)]$$

and

$$\delta(L) = \det(I - A_{22}(L)) (1 - B_{11}(L))$$

On the other hand $\delta(L)$ is such that

$$\delta(L)u_t = \det\left(I - A_{22}(L)\right)\left(1 - B_{11}(L)\right)\epsilon_{x_t} + \det\left(I - A_{22}(L)\right)B_{12}(L)\epsilon_{y_t}$$

and hence

$$u_t = \epsilon_{x_t} + \frac{B_{12}(L)}{\left(1 - B_{11}(L)\right)}\epsilon_{y_t} \tag{11}$$

where this equality is with probability 1. Since u_t is a white noise, Equation (11) implies that $B_{12}(L) = 0$.

Second case: $A_{21}(L) = \mathbf{0}$.

By hypothesis $A_{21}(L) = \mathbf{0}$, this implies that

$$\det\left(A(L)\right) = \left(1 - A_{11}(L)\right)\det\left(I - A_{22}(L)\right)$$

and the first row of the matrix $C(L)$ is given by $C_1(L) = [C_{11}(L), C_{12}(L)]$ where

$$C_{11}(L) = [\det\left(I - A_{22}(L)\right)\left(1 - B_{11}(L)\right) - \det\left(I - A_{22}(L)\right)A_{12}(L)\left(I - A_{22}(L)\right)^{-1}B_{21}(L)]$$

$$C_{12}(L) = [\det\left(I - A_{22}(L)\right)B_{12}(L) - \det\left(I - A_{22}(L)\right)A_{12}(L)\left(I - A_{22}(L)\right)^{-1}\left(I - B_{22}(L)\right)]$$

If $d(M_0, M_1) = 0$, then

$$\frac{1 - A_{11}(L)}{1 - B_{11}(L)} = \frac{\left(1 - A_{11}(L)\right)\det\left(I - A_{22}(L)\right)}{\delta(L)}$$

and hence

$$\delta(L) = \left(1 - B_{11}(L)\right)\det\left(I - A_{22}(L)\right)$$

The following equality then occurs with probability 1:

$$u_t = \epsilon_{x_t} - \frac{A_{12}(L)\left(I - A_{22}(L)\right)^{-1}B_{21}(L)}{1 - B_{11}(L)}\epsilon_{x_t} + \frac{B_{12}(L)}{1 - B_{11}(L)}\epsilon_{y_t}$$

$$- \frac{A_{12}(L)\left(I - A_{22}(L)\right)^{-1}\left(1 - B_{22}(L)\right)}{1 - B_{11}(L)}\epsilon_{y_t}$$

Since u_t is a white noise, this implies that $A_{12}(L) = 0$ and $B_{12}(L) = 0$.
$\square$

We have also the following corollaries.

Corollary 1. *Let $w_t = (x_t, y_t')'$ be a pure VAR(p) process. y does not Granger-cause x if and only if $d(M_0, M_1) = 0$.*

Proof of Corollary 1. ($\Rightarrow$) If y does not Granger-cause x, then $A_{12}(L) = 0$. By hypothesis, $B_{12}(L) = \mathbf{0}$. Hence we have $A_{12}(L) = B_{12}(L) = \mathbf{0}$. It follows from Proposition 1 that $d(M_0, M_1) = 0$.

($\Leftarrow$) If $d(M_0, M_1) = 0$, by Proposition 1, it follows that $A_{12}(L) = 0$ and this, in a VAR framework, implies that y does not Granger-cause x. $\square$

Corollary 2. *Let $w_t = (x_t, y_t')'$ be a pure VMA(q) process. y does not Granger-cause x if and only if $d(M_0, M_1) = 0$.*

Proof of Corollary 2. It is similar to the proof of Corollary 1. $\square$

$\square$

3. Inferential Implications

Proposition 1 allows us to test the set of linear restrictions Equation (5) considering the null hypothesis $H_0 : d(M_0, M_1) = 0$. Further, we observe that if the process $\{w_t; t \in \mathbb{Z}\}$ follows a VAR model, Corollary 1 establishes that the Granger noncausality from y_t to x_t is equivalent to the condition $d(M_0, M_1) = 0$. Thus, in a VAR framework, we can test for Granger noncausality from y_t to x_t using the null hypothesis $d(M_0, M_1) = 0$ without considering the nature of the involved variables. In fact, it is well-known that the use of non-stationary data in causality tests can yield spurious causality results (see, e.g., [12]). Thus, before testing for Granger causality, it is important to establish the properties of the time series involved because different model strategies must be adopted when: the series are I(0), the series are partly I(0) and partly I(1), the series are determined I(1) but not cointegrated, or the series are cointegrated. Of course, the weakness of this strategy is that incorrect conclusions drawn from preliminary analysis might be carried over into the causality tests. In the VAR framework an alternative method is the so-called *lag-augmented Wald test* (see [13,14]), which is a modified Wald test that requires the knowledge of the maximum order of integration of the involved variables. In this way, the proposed test based on the AR-metric can be a valid alternative for a Granger noncausality test (see [4]), since it does not require the exact knowledge of the series properties or the knowledge of the maximum order of integration.

To conduct inference on the basis of Proposition 1, we need an asymptotic distribution for $d(M_0, M_1)$. In the class of ARMA processes, the asymptotic distribution of the maximum likelihood estimator $\hat{d}^2$ has been studied, among others, in [5,15]. In this case, for two independent ARMA(p, q) processes X and Y, under the null hypothesis $d(X, Y) = 0$, the maximum likelihood estimator $\hat{d}^2$ has the following asymptotic distribution:

$$\hat{d}^2 \sim 2 \sum_{j=1}^K \lambda_j \chi^2_{g_j}$$

where χ^2_{gj} are independent χ^2-distributions with g_j degrees of freedom, λ_j are the eigenvalues of the covariance matrix of $(\hat{\varphi}_{xi} - \hat{\varphi}_{yi})$ and $K < p + q$. The evaluation of this distribution can be cumbersome; hence approximations, as well as evaluation algorithms, have been proposed (see [15]). Anyhow, in our framework, the ARMA models implied by Equation (6) and by the VARMA model Equation (8) under the null hypothesis $A_{12}(L) = B_{12}(L) = \mathbf{0}$ are equal, so they cannot be considered independent. Then, to conduct the inferential procedures, we suggest the bootstrap algorithm proposed by Di Iorio and Triacca [4], which is described in the next section.

3.1. The Bootstrap Test Procedure

For an easy illustration of our bootstrap procedure, let us consider a bivariate VARMA(p, q) model simply denoted as $\mathbf{A}(L)\mathbf{w}_t = \mathbf{B}(L)\boldsymbol{\epsilon}_t$ where $\mathbf{w}_t = (x_t, y_t)'$, $\boldsymbol{\epsilon}_t = (\epsilon_{xt}, \epsilon_{yt})'$ with covariance matrix $\boldsymbol{\Sigma}$ and, based on Proposition 1, we want to test the null hypothesis $H_0 : A_{12}(L) = B_{12}(L) = \mathbf{0}$ using $H_0 : d(M_0, M_1) = 0$

1. Estimate on the observed data the VARMA(p, q) and obtain $\hat{\mathbf{A}}(L)$, $\hat{\mathbf{B}}(L)$, $\hat{\boldsymbol{\Sigma}}$ and the residuals $\hat{\boldsymbol{\epsilon}}_t$;
2. using the estimated parameters from Step 1, obtain the univariate ARMA implied by the estimated VARMA for the subprocess x_t;

3. evaluate the AR(∞) representation truncated at some suitable lag p_1 of the ARMA model in Step 2 (model M_1);

4. estimate for x_t, using the observed data, an ARMA(p, q) model under the null hypothesis H_0 : $A_{12}(L) = B_{12}(L) = \mathbf{0}$ and evaluate its AR(∞) representation truncated at some suitable lag p_0 (model M_0);

5. evaluate the distance $\hat{d}(M_0, M_1)$ between the AR(p_0) and the AR(p_1) obtained in Steps 3 and 4;

6. estimate the VARMA(p, q) model under the null hypothesis H_0 : $A_{12}(L) = B_{12}(L) = \mathbf{0}$ to obtain the estimates $\tilde{A}(L)$, $\tilde{B}(L)$ and $\tilde{\Sigma}$;

7. apply bootstrap to the re-centered residuals $\hat{\epsilon}_t$ and obtain the pseudo-residuals ϵ_t^*;

8. generate the pseudo-data $(x_t^*, y_t^*)'$ obeying the null hypothesis using $\tilde{\boldsymbol{A}}(L)(x_t^*, y_t^*)' = \tilde{\boldsymbol{B}}(L)\epsilon_t^*$ with $\tilde{\Sigma}$;

9. using the pseudo-data $(x_t^*, y_t^*)'$, repeat Steps 1–5 to obtain the bootstrap estimate of the distance $d^*(M_0, M_1)$;

10. repeat Steps 7–9 for b times;

11. evaluate the bootstrap p-value as the proportion of the b estimated bootstrap distance d^* that exceeds the same statistic evaluated on the observed data $\hat{d}$, that is, $pval_b = proportion(d^* > \hat{d})$.

When this procedure is applied, two remarks concerning the pseudo-data generation and the modeling of the dependency across the subprocess are in order. Firstly, in a well-specified model framework (as well as during a simulation exercise), the estimated residuals $\hat{\epsilon}_t$ do not show any autocorrelation structure, so we do not need any particular resampling scheme for dependent data to obtain pseudo-error terms ϵ_t^*, and we can then apply a simple resampling procedure. Besides, for empirical studies the pseudo-data can be obtained considering several resampling strategies, as a block bootstrap algorithm (see [16]). Secondly, in order to reproduce the dependency across the subprocess expressed by Σ in the pseudo-data, we simply have to apply the resampling algorithm to the entire $T \times n$ matrix of the estimated residuals $\hat{\epsilon}_t$.

4. Monte Carlo Experiments

The performance of the proposed inferential strategy can be investigated by means of a set of Monte Carlo experiments. In particular, we consider the test for the set of linear restriction associated to a Granger noncausality test for two different DGP: a stable bivariate VARMA($1, 1$) model and a cointegrated bivariate VAR(2) model. Our test will be compared with the performance of a Wald test for the VARMA($1, 1$) and with the lag-augmented Wald test suggested by Dolado *et al.* and Toda *et al.* [13,14] for the cointegrated VAR model.

4.1. Bivariate VARMA($1, 1$) Model

Consider the following stable VARMA($1, 1$) model:

$$\begin{bmatrix} 1 - 0.8L & -\alpha_1 L \\ -0.3L & 1 - 0.5L \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & -\beta_1 L \\ 0.25L & 1 - 0.5L \end{bmatrix} \begin{bmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{bmatrix} \tag{12}$$

with covariance matrix $\Sigma_\epsilon = \begin{bmatrix} 4 & 3 \\ 3 & 6 \end{bmatrix}$.

In our study, the tests of the null hypothesis

$$H_0 : \alpha_1 = \beta_1 = 0$$

were carried out using nominal significance levels of 1%, 5% and 10%. To analyze the power of the test, we consider the two cases below to verify how the test reacts when the parameter values move away from zero:

Power 1. $\alpha_1 = 0.2$, $\beta_1 = -0.7$,

Power 2. $\alpha_1 = 0.5$, $\beta_1 = -0.7$.

It is well-known that a maximum likelihood estimation of a VARMA model can be a challenging task (see, e.g., [1,17]). For this reason we consider sample size $T = 100$ and $T = 200$, which are quite large compared with what is usually found in empirical applications. Taking into account the dimension of our exercise, we perform the maximum likelihood estimation using the Kalman filter procedure implemented in $Gretl$ (ver. 1.9.14) (see [18]). Therefore, due to computational time involved by the maximum likelihood estimation of the VARMA model, the experiments are based on 400 Monte Carlo replications and 400 bootstrap redrawings. We compare our results with the usual Wald test using, for a proper comparison, also the bootstrap p-values obtained by the same bootstrap algorithm described above. Finally, we verify by some preliminary experiments that a suitable value for p_0 and p_1 in Steps 3 and 4 in the bootstrap algorithm is 15. The results are reported in Table 1.

Table 1. VARMA$(1, 1)$ AR-metric and Wald test. Size and Power.

	AR–$Metric$ Boot p-Values			$Wald$ Boot p-Values			$Wald$ Asy p-Values		
nom	Size	Power 1	Power 2	Size	Power 1	Power 2	Size	Power 1	Power 2
					$T = 100$				
0.01	0.01	0.15	0.45	0.01	0.85	0.89	0.01	0.97	0.99
0.05	0.09	0.48	0.90	0.06	0.90	0.90	0.06	0.99	100
0.10	0.14	0.62	0.97	0.12	0.90	0.90	0.11	0.99	100
					$T = 200$				
0.01	0.03	0.49	0.98	0.01	0.97	0.98	0.01	100	100
0.05	0.08	0.70	1.00	0.05	0.99	0.98	0.04	100	100
0.10	0.13	0.79	1.00	0.07	0.99	1.00	0.09	100	100

As we can see from Table 1, the size for the AR-metric test is quite satisfactory, and the power increases with growing sample size and as the true parameter values move away from zero. In any case, as expected, the difficulties of the maximum likelihood estimation for the VARMA model affect the distance more than the Wald test, which shows a better power. In fact, as the bootstrap algorithm underlines, the distance-based test is built on the autocovariances obtained by the estimated values of the parameters. Hence, its performances are heavily dependent on the quality of these estimates.

4.2. Bivariate Cointegrated VAR(2) Model

Most encouraging results are obtained with the second DGP. Consider the following cointegrated bivariate VAR(2) model:

$$\begin{bmatrix} 1 - 1.5L + 0.5L^2 & -\alpha_1 L - \alpha_2 L^2 \\ -0.8L + 0.3L^2 & 1 - L + 0.5L^2 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{bmatrix} \tag{13}$$

with covariance matrix $\Sigma_\epsilon = \begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$.

As before, the tests of the null hypothesis

$$H_0 : \alpha_1 = \alpha_2 = 0$$

were carried out using nominal significance levels of 1%, 5% and 10%. To analyze the power of the test, we consider again the two cases below:

Power 1. $\alpha_1 = -\alpha_2 = 0.3$

Power 2. $\alpha_1 = -\alpha_2 = 0.6$

In this case the parameter estimation is easier. To make our Monte Carlo experiment more relevant for actual empirical applications, we consider sample size $T = 50$, a medium size in terms of annual data but small size for a quarterly frequency, and $T = 100$, which is a time span large in terms of annual data but pretty common for quarterly data. Now we compare the performances for our test with the lag-augmented Wald test proposed by Dolado *et al.* and Toda *et al.* [13,14] in this framework. The lag-augmented Wald test has an asymptotic χ^2-distribution with p degrees of freedom when a VAR($p + d_{max}$) is estimated, where d_{max} is the maximal order of integration for the series in the system. However, it is well-known that the lag-augmented Wald test based on asymptotic critical values may suffer from size distortion and low power especially for small samples [19,20]. Thus, to overcome this problem, we apply the same bootstrap algorithm described above using the Wald test from an augmented VAR($2 + a_{max}$), with augmentation order $a_{max} = 1$, and we evaluate the bootstrap p-values.

For this DGP the experiment is based on 1000 Monte Carlo replications and 1000 Bootstrap redrawings, and, as before, in Step 3 we set $p_0 = 15$. The results are collected in Table 2. We note that, for a nominal significance level of 5%, our results are rather similar to those of the second part of Table 3 reported in Shukur and Mantalos [21]. The comparison of the power estimates for our test and the lag-augmented Wald test of Toda *et al.* [14] shows that our test has relatively high power properties in all situations, while the size is very close to the nominal values for both tests.

Table 2. VAR(2) AR-metric and lag-augmented Wald test. Size and Power Bootstrap *p*-values. Asy *p*-values.

	AR-Metric Boot *p*-Values			*Aug-Wald* Boot *p*-Values			*Aug-Wald* Asy *p*-Values		
nom	**Size**	**Power 1**	**Power 2**	**Size**	**Power 1**	**Power 2**	**Size**	**Power 1**	**Power 2**
					$T = 50$				
0.01	0.02	0.22	0.64	0.01	0.05	0.35	0.01	0.06	0.36
0.05	0.07	0.42	0.82	0.04	0.18	0.62	0.04	0.17	0.61
0.10	0.12	0.56	0.89	0.08	0.27	0.73	0.08	0.26	0.72
					$T = 100$				
0.01	0.01	0.54	0.98	0.01	0.18	0.78	0.01	0.18	0.80
0.05	0.04	0.78	1.00	0.04	0.38	0.92	0.04	0.38	0.91
0.10	0.11	0.85	1.00	0.09	0.50	0.95	0.08	0.50	0.95

5. Empirical Applications

In this section we present two empirical examples to illustrate the application of the test suggested in the paper. First, we consider a VAR model and in particular we examine the causal relationship between the log of real per capita income and the inflation. Then, we consider a VARMA example based on the *SCC* dataset discussed in [22].

To take into account any possible dependence structure in the residuals of the estimated models, we use the Stationary Bootstrap ([23]) as resampling algorithm. The Stationary Bootstrap is a block bootstrap scheme where the resampled pseudo-series are stationary; this scheme chains blocks of observations of the original series starting at random locations, and the length of each block is randomly chosen from a geometric distribution. Following Palm *et al.* [24], the mean block length can be computed as a function of the length of the time sample; by some exploratory simulations we verify the robustness of the tests to different block sizes, so we report results for blocks $1.75\sqrt{T}^3$.

To discuss the possible causal relationship between the log of real per capita income (y) and inflation (Δp) we re-examined the dataset used by Ericsson *et al.* [25]. The dataset refers to United States over the period 1953–1992 and can be downloaded from the *Journal of Applied Econometrics* Data Archive. The VAR order selection is based on Bayesian Information Criterion and the following model is estimated.

$$y_t = \underset{(0.21)}{0.03} + \underset{(0.15)}{0.93}y_{t-1} + \underset{(0.16)}{0.93}y_{t-2} - \underset{(0.24)}{0.82}\Delta p_{t-1} + \underset{(0.23)}{0.53}\Delta p_{t-2} + \epsilon_{1_t}$$

$$\Delta p_t = \underset{(0.12)}{-0.35} + \underset{(0.09)}{0.34}y_{t-1} - \underset{(0.09)}{0.33}y_{t-2} + \underset{(0.13)}{1.15}\Delta p_{t-1} - \underset{(0.13)}{0.33}\Delta p_{t-2} + \epsilon_{2_t}$$

The computed $\hat{d}(M_0, M_1)$-statistic is equal to 0.35 with a bootstrap *p*-value 0. This result indicates the presence of Granger causality from output to inflation. This finding is in accordance with the results of Ericsson *et al.* [25]. The same result is obtained using the lag-augmented Wald test.

The *SCC* dataset discussed by Tiao and Box [22] considers the quarterly time series of the U.K. Financial Time Ordinary Share Index, the U.K. Car Production and the U.K. Financial Time Commodity Price from the III Quarter 1952 to the IV Quarter 1967. The goal is verify the possibility of predicting the first variable from the lagged values of the last two. According to Tiao and Box [22], a VARMA$(1, 1)$ is the best model for this data, then a null hypothesis following Equation (5) will be the inferential base to test just a sufficient condition on the predictability hypothesis. The VARMA$(1, 1)$ maximum likelihood parameter estimates using the Kalman filter procedure implemented in Gretl (ver. 1.9.9) are the following (standard errors in bracket):

$$
\mu = \begin{pmatrix} 1.157 \\ \scriptstyle (1.178) \\ 0.774 \\ \scriptstyle (0.879) \\ 2.444 \\ \scriptstyle (1.941) \end{pmatrix} \quad
A(L) = \begin{pmatrix} 0.853 & 0.117 & -0.057 \\ \scriptstyle (0.132) & \scriptstyle (0.147) & \scriptstyle (0.07) \\ -0.034 & 0.960 & -0.030 \\ \scriptstyle (0.110) & \scriptstyle (0.108) & \scriptstyle (0.051) \\ -0.091 & 0.153 & 0.843 \\ \scriptstyle (0.308) & \scriptstyle (0.308) & \scriptstyle (0.112) \end{pmatrix} \quad
B(L) = \begin{pmatrix} 0.386 & -0.488 & -0.223 \\ \scriptstyle (0.134) & \scriptstyle (0.179) & \scriptstyle (0.097) \\ 0.752 & -0.782 & -0.175 \\ \scriptstyle (0.164) & \scriptstyle (0.181) & \scriptstyle (0.117) \\ 1.373 & -0.820 & -0.017 \\ \scriptstyle (0.353) & \scriptstyle (0.372) & \scriptstyle (0.185) \end{pmatrix}
$$

The estimates are quite similar to the values reported as "full model" in the Table 10 in [22], taking into account the difference in the estimation algorithm and software. The computed $\hat{d}(M0, M_1)$-statistic is equal to 4.58 with a bootstrap p-value 0.225, evaluated on 500 bootstrap replications, and this finding is in accordance with the results of "final model" in the Table 10 in Tiao and Box [22]. We perform also a Wald test on the same null hypothesis, the value is 36.684, which asymptotically rejects the null, but with a bootstrap p-value 0.146 that sustains the results of our test.

6. Conclusions

In this paper we characterized a set of linear restrictions in a Vector Autoregressive Moving Average (VARMA) model in terms of the notion of distance between ARMA models and we have derived a new inferential procedure. In particular, this new procedure can be useful for a new Granger noncausality test in a VAR framework. The advantage of this test is that it can be carried out irrespective of whether the variables involved are stationary or not and regardless of the existence of a cointegrating relationship among them. Our inferential procedure has been validated by a set of Monte Carlo experiments. In a VARMA framework this procedure shows encouraging results even if a deeper investigation, made complex by the computational time, is needed. In a cointegrated VAR framework our method for detecting causality has provided better results, as the conducted simulation study has shown that our test exhibits good performance in terms of size and power properties, even in small samples. Finally, we have shown that this test can be usefully applied in practical situations to test causality between economic time series.

Author Contributions

The authors contributed jointly to the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*; Springer: New York, NY, USA, 2005.
2. Piccolo, D. A distance measure for classifying ARIMA models. *J. Time Ser. Anal.* **1990**, *11*, 153–164.
3. Granger, C.W.J. Investigating causal relations by econometric methods and cross-spectral methods. *Econometrica* **1969**, *37*, 424–438.
4. Di Iorio, F.; Triacca, U. Testing for Granger non-causality using the autoregressive metric. *Econ. Model.* **2013**, *33*, 120–125.
5. Corduas, M.; Piccolo, D. Time series clustering and classification by the autoregressive metric. *Comput. Stat. Data Anal.* **2008**, *52*, 1860–1862.
6. Gonzalo, J.; Lee, T.H. Relative power of t type tests for stationary and unit root processes. *J. Time Ser. Anal.* **1996**, *17*, 37–47.
7. Grimaldi, S. Linear parametric models applied on daily hydrological series. *J. Hydrol. Eng.* **2004** *9*, 383–391.
8. Maharaj, E.A. A significance test for classifying ARMA models. *J. Stat. Comput. Simul.* **1996**, *54*, 305–331.
9. Otranto, E. Clustering heteroskedastic time series by model-based procedures. *Comput. Stat. Data Anal.* **2008**, *52*, 4685–4698.
10. Otranto, E. Identifying financial time series with similar dynamic conditional correlation. *Comput. Stat. Data Anal.* **2010**, *54*, 1–15.
11. Boudjellaba, H.; Dufour, J.-M.; Roy, R. Testing causality between two vectors in multivariate ARMA models. *J. Am. Stat. Assoc.* **1992**, *87*, 1082–1090.
12. Sims, C.A.; Stock J.H.; Watson M.W. Inference in linear time series models with some unit roots. *Econometrica* **1990**, *58*, 113–144.
13. Dolado, J.; Lütkepohl, H. Making Wald Tests work for cointegrated VAR systems. *Econ. Rev.* **1996**, *15*, 369–386.
14. Toda, H.Y.; Yamamoto, T. Statistical inferences in vector autoregressions with possibly integrated processes. *J. Econ.* **1995**, *66*, 225–50.
15. Corduas, M. La metrica Autoregressiva tra modelli ARIMA: Una procedura operativa in linguaggio GAUSS. *Quad. Stat.* **2000**, *2*, 1–37.
16. MacKinnon, J. Bootstrap inference in econometrics. *Can. J. Econ.* **2002**, *35*, 615–645.
17. Metaxoglou, K.; Smith, A. Estimating VARMA models using EM algorithm. *J. Time Ser. Anal.* **2007**, *28*, 666–685.

18. GRETL Gnu Regression, Econometrics and Time-series Library. Available online: http://gretl.sourceforge.net/ (acessed on 16 June 2014).

19. Giles, D.E.A. Causality between the measured and underground economies in New Zealand. *Appl. Econ. Lett.* **1997**, *4*, 63–67.

20. Mavrotas, G.; Kelly, R. Old wine in new bottles: Testing causality between savings and growth. *Manch. Sch.* **2001**, *69*, 97–105.

21. Shukur, G.; Mantalos, P. A simple investigation of the Granger causality test in integrated-cointegrated VAR systems. *J. Appl. Stat.* **2000**, *27*, 1021–1031.

22. Tiao, G.C.; Box, G.E.P. Modeling multiple times series with applications. *J. Am. Stat. Assoc.* **1981**, *76*, 802–816.

23. Politis, D.N.; Romano, J.P. The stationary bootstrap. *J. Am. Stat. Assoc.* **1994**, *89*, 1303–1313.

24. Palm, F.C.; Smeekes, S.; Urbain, J.P. Cross-sectional dependence robust block bootstrap panel unit root tests. *J. Econ.* **2011**, *163*, 85–104.

25. Ericsson, N.R.; Irons, J.S.; Tryon, R.W. Output and inflation in the long run. *J. Appl. Econ.* **2001**, *16*, 241–253.