TESTING FOR A SET OF LINEAR RESTRICTIONS IN VARMA MODELS USING AR METRIC

Francesca Di Iorio
Dipartimento TEOMESUS,
Università di Napoli Federico II, *fdiiorio@unina.it*

Umberto Triacca
Dipartimento di Ingegneria Informatica, Informatica e Matematica
Università dell'Aquila, *umberto.triacca@ec.univaq.it*

Abstract[1]

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 notion of distance between two univariate ARMA models, the autoregressive metric introduced by Piccolo (1990). In particular, we show that this set of linear restrictions is equivalent to a null distance d between two given ARMA models. This result provides the logical basis for using $d = 0$ 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.

Keywords VARMA, linear restriction, AR metric, bootstrap

1 Motivation

In this paper we investigate the relationship between a set of linear restrictions concerning the parameters of a vector autoregressive moving average (VARMA) model and the notion of the distance between two univariate ARMA models, the autoregressive metric (AR) introduced by Piccolo (1990). In particular, we show that these linear restrictions are satisfied if and only if the distance d between two given ARMA models is zero. This result provides the logical basis for using $d = 0$ as a null hypothesis for testing this set of restrictions. It is important to underline that the set of linear restrictions considered is sufficient for the condition of Granger noncausality, while in the VAR framework, it becomes also a necessary condition. This theoretical result allows the implementation of an inferential procedure and a bootstrap algorithm. Our procedure is validated by some Monte Carlo experiments also in small sample. The paper is organized as follows. Section 2 introduces the notion the distance between ARMA models and specifies the relationship between 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

[1]Comments and suggestions from Giorgio Calzolari, Riccardo "Jack" Lucchetti and participants to the Conference ICEEE2011 (Pisa, Italy) are gratefully acknowledged; the usual disclaimers apply.

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 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 admit 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 the class of ARMA invertible models. If $x_t \in \ell$ and $y_t \in \ell$, following Piccolo (1990), the AR metric is defined as the Euclidean distance between the corresponding π-weights sequence, $\{\pi_j\}$,

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

The AR metric d has been widely used in time series analysis (see e.g. Maharaj(1996), Gonzalo and Lee (1996), Grimaldi (2004), Corduas and Piccolo (2008) and Otranto (2008, 2010)). We observe that (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 a $n \times 1$ 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 a $n \times 1$ 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 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 model (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 a $(n-1) \times 1$ vector of time series. Model (2) accordingly to 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, Boudjellaba, Dufour and Roy (1992)) that y_t does not 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 (4) to hold is

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

We note that, if the condition (5) holds then x_t follows an 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 (5), using the notion the distance between ARMA models measured by (1). In particular, we will consider the distance between the ARMA(p,q) model (6) and the ARMA model for the subprocess $\{x_t; t \in Z\}$ implied by the VARMA(p,q) model (2).

The implied ARMA model can be obtained as follows. Premultiplying both sides of (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 $\text{Adj}\,(A(L))\,\epsilon_t$ is a sum of finite order MA processes, thus it is a finite order MA process (see Lütkepohl, 2005, Proposition 11.1). 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

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

2.1 Theoretical results

We consider the distance according to (1) between the models (8) and (6):

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

where

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

The following proposition provide a necessary and sufficient condition for the set of linear restrictions (5) in terms of the distance d.

Proposition 1. $A_{12}(L) = B_{12}(L) = \mathbf{0}$ if and only if $d = 0$.

Proof. ($\Rightarrow$) 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

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

whit $D(L) = [1 - A_{11}(L) - A_{12}(L) (I - A_{22}(L))^{-1} A_{21}(L)]^{-1}$.

If $A_{12}(L) = B_{12}(L) = \mathbf{0}$, 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 $u_t = \epsilon_{xt}$ 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 $d = 0$.

($\Leftarrow$) We have to prove that $A_{12}(L) = B_{12}(L) = \mathbf{0}$. 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 $d = 0$, 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

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

Using the Schur's formula, we get

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

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

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

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

$$\deg \left(\delta(L)\right) < \infty,$$

(9) implies that

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

Since

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

it follows for (10) that it must be

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

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

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

and

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

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

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

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

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

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

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

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

If $d = 0$, then

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

and hence

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

It follows that

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

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

□

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 $d = 0$.

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

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

□

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 $d = 0$.

Proof. It is similar to the proof of Corollary 1

□

3 Inferential implications

We verify that Proposition 1 allows us to test for the set of linear restrictions (5) considering the null hypothesis $H_0 :\ d = 0$. Further, we observe that if the process $\{w_t;\ t \in \mathbb{Z}\}$ follows a VAR model, Corollary 1 establishes that the non-causality from y_t to x_t is equivalent to the condition $d = 0$. Thus, in a VAR framework, we can test for Granger non-causality from y_t to x_t using the null hypothesis $d = 0$ without considering the nature of the involved variables. Infact, it well known that the use of non-stationary data in causality tests can yield spurious causality results (see *e.g.* Sims et al. (1990)). Thus, before testing for Granger causality, it is important to establish the properties of the time series involved because different model strategies must be adopt when the series are I(0), or when part of the series are I(0) and part I(1), when the series are determined I(1) but not cointegrated or when 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. An alternative method is the so called *lag-augmented Wald test* (see Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996)), that is a modified Wald test which requires the knowledge of the maximum order of integration of the involved variables. In this way, the AR-metric based test proposed, since it not require the exact knowledge of the series properties or the knowledge of the maximum order of integration, can be a valid alternative for a Granger non-causality tests.

To conduct inference on the basis of Proposition 1 we need an asymptotic distribution for d. In the class of ARMA processes, the asymptotic distribution of the maximum likelihood estimator $\hat{d}^2$ has been studied, among others, in Corduas and Piccolo (2008). In this case, for two independent ARMA(p,q) process x_t and y_t, under the null hypothesis $d(x_t, y_t) = 0$, 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 chi-squared distributions whit g_j degree of freedom, λ_j are the eingvalues of the covariance matrix of $(\hat{\varphi}_{xi} - \hat{\varphi}_{yi})$ and $K < p + q$. The evaluation of this distribution can be cumbersome, then approximations, as well as evaluation algorithms, have been proposed (see Corduas 2000). Anyhow, in our framework the ARMA models implied by (6) and by the VARMA model (8) under the null hypothesis $A_{12}(L) = B_{12}(L) = \mathbf{0}$ are equal, so they can not be consider indipendent. Then, to conduct the inferential procedures, we propose to adopt the bootstrap algorithm described in the next section.

3.1 The bootstrap test procedure

For an easy illustration of our bootstrap procedure, let us consider a bivariate VAR(1) model simply denoted as $\boldsymbol{A}\boldsymbol{w}_t = \boldsymbol{\epsilon}_t$ where $\boldsymbol{w}_t = (x_t, y_t)'$, $\boldsymbol{\epsilon}_t = (\epsilon_{xt}, \epsilon_{yt})'$ with covariance matrix $\boldsymbol{\Sigma}$ and, based on Corollary 2, we want to test the null $H_0 : y_t \not\Rightarrow x_t$

1. Estimate on the observed data the VAR(p) and obtain $\hat{\boldsymbol{A}}(L)$, $\hat{\boldsymbol{\Sigma}}$ and the residuals $\hat{\boldsymbol{\epsilon}}_t$;

2. using the estimated parameters from step 1, obtain the univariate ARMA implied by the estimated VAR for the sub-process x_t;

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

4. estimate for x_t, using the observed data, an AR(p) model under null hypothesis of non causality $H_0 : y_t \not\Rightarrow x_t$;

5. evaluate the distance $\hat{d}$ between the AR(p_1) and the AR(p) model obtained in step 3 and 4;

6. estimate the VAR(p) model under the null hypothesis $H_0 : y_t \not\Rightarrow x_t$ obtaining the estimates $\tilde{A}(L)$ and $\tilde{\Sigma}$;

7. apply Bootstrap on $\hat{\epsilon}_t$ and obtain the pseudo-residuals $\boldsymbol{\epsilon}^*_t$;

8. generate the pseudo-data $(x^*_t, y^*_t)'$ obeying to the null of Granger non-causality using $\tilde{\boldsymbol{A}}(L)(x^*_t, y^*_t)' = \boldsymbol{\epsilon}^*_t$ with $\tilde{\boldsymbol{\Sigma}}$;

9. using the pseudo data $(x_t^*, y_t^*)'$, repeat steps from 1 to 5 obtaining the bootstrap estimate of the distance d^*

10. repeat steps from 7 to 9 for B times

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

Two remarks are in order: i) in this framework the estimated residuals $\hat{\epsilon}_t$ do not show any autocorrelation structure, so we don't need any particular resampling scheme for dependent data, then we can apply a simple resampling procedure (MacKinnon, 2002); ii) an essential feature to be taken into account is the dependency across the sub-process espressed by $\mathbf{\Sigma}$. In order to reproduce it 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 the means of a set of Monte Carlo experiment. In particular we consider the test for the set of linear restriction associated to a Granger non-causality 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 stardand Wald test for the VARMA(1,1) and with the lag-augmented Wald test suggested by Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) 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 considered the two cases below, to verify how the test react 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 af a VARMA model can be a challenging task (see *e.g.* Lütkepohl, 2005, or Metaxoglou and Smith, 2007). Taking in to account the dimension of our exercise,

Table 1: VARMA(1,1) AR-metric and Wald test. Size and Power - Bootstrap p-values

	$AR - metric$			$Wald$		
nom	Size	Power1	Power2	Size	Power1	Power2
			T=100			
0.01	0.01	0.15	0.45	0.01	0.85	0.89
0.05	0.09	0.48	0.90	0.06	0.90	0.90
0.10	0.14	0.62	0.97	0.12	0.90	0.90
			T=200			
0.01	0.03	0.49	0.98	0.01	0.97	0.98
0.05	0.08	0.70	1.00	0.05	0.99	0.98
0.10	0.13	0.79	1.00	0.07	0.99	1.00

we perform the ML estimation using the Kalman filter procedure implemented in Gretl (ver. 1.9.9). For these reason we consider as sample size T=100 and T=200, that are quite large compared to that usually is found in empirical applications. 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, and to a better comparison we consider for the Wald test the bootstrap p-values obtained by the same bootstrap algorithm described above. Finally, we verify that a suitable value for p_1 in step 3 in the bootstrap algorithm is $p_1 = 15$. The results are reported in table 1.
As we can see form table 1 the size for the AR metric test is quite satisfactory, and the power grows up with the sample size and as the true parameter values move away from zero. In any cases, as expected, the difficulties of the Maximum Likelihood evaluation for the VARMA model affect more the distance than the Wald test, that show a better power. Infact, as the bootstrap algorithm underlines, the distance based test is build 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 considered 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, then, to make our Monte Carlo experiment more relevant for actual empirical applications, we consider as sample size $T = 50$, a size medium in terms of annual data

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

	$AR-metric$			$Aug-Wald$		
nom	Size	Power1	Power2	Size	Power1	Power2
T=50						
0.01	0.02	0.22	0.64	0.01	0.05	0.35
0.05	0.07	0.42	0.82	0.04	0.18	0.62
0.10	0.12	0.56	0.89	0.08	0.27	0.73
T=100						
0.01	0.01	0.54	0.98	0.01	0.18	0.78
0.05	0.04	0.78	1.00	0.04	0.38	0.92
0.10	0.11	0.85	1.00	0.09	0.50	0.95

but small for a quarterly frequency, and $T = 100$, that 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 Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996). 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 (Giles (1997) and Mavrotas and Kelly (2001)). Thus, to overcome this problem, we apply the same bootstrap algorithm described above using the Wald test from an augmented VAR(2+d_{max}), with $d_{max} = 1$ 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 $p_1 = 15$. 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 (2000). The comparison of the power estimates for our test and the lag-augmented Wald test of Toda and Yamamoto 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.

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 the relationship between income and CO_2 emissions, then examine the causal relationship between the log of real per capita income.

It is well known that the conjecture of the Environmental Kuznets Curve (EKC) hypothesis (Coondoo and Dinda, 2002) is such that, initially as per capita income rises, environmental degradation intensifies, but in later levels of economic growth it tends to subside. Thus, it is presumed that income Granger-causes CO_2 emissions. Hence, we investigate the causal relationship from CO_2 emissions to income by using our test. To establishes if the CO_2 emissions Granger cause or not the GDP may be useful for policy implication.

For example, if for a given country the CO_2 emissions does not Granger-cause the GDP, then any effort to reduce CO_2 emissions does not restrain the development of the economy. If, on the other hand, the causality runs from CO_2 emissions to income, reducing energy consumption (by a carbon tax policy, say) may lead to fall in income.

We use annual data on per capita Real Gross Domestic Product (y) and per capita of Carbon Dioxide Emissions (c) in United States, for the period 1960-2006. All data are from World Development Indicators and are in natural logarithms.

Based on Bayesian Information Criterion, a VAR model of order 1 was selected. The estimated model is given by:

$$y_t = \underset{(0.10)}{0.18} + \underset{(0.01)}{0.99}\, y_{t-1} - \underset{(0.03)}{0.05}\, c_{t-1} + \epsilon_{1_t}$$

$$c_t = \underset{(0.17)}{0.43} - \underset{(0.01)}{0.02}\, y_{t-1} + \underset{(0.06)}{0.88}\, c_{t-1} + \epsilon_{2_t}$$

The estimated distance is $\hat{d} = 0.0073$ and the bootstrap p-value is 0.58. Thus we can conclude that there is no evidence of Granger causality from CO_2 emissions to output.

We now consider the causal relationship between the log of real per capita income (y) and inflation (Δp) in the United States over the period 1953-1992. In particular, we have re-examined the data set used by Ericsson *et al.* (2001). We downloaded the annual time series data from the *Journal of Applied Econometrics* Data Archive. The following bivariate VAR model is estimated.

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

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

The order of the VAR has been chosen using the Bayesian Information Criterion. The computed $\hat{d}$-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.* (2001). The same result is obtained using the lag-augmented Wald test.

6 Conclusions

In this paper we characterized a set of linear restrictions in a vector autoregressive-moving average (VARMA) model in term 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 non-causality test in a VAR framework. The advantage of this test is that it can be can be carried out irrespectively 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 a 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.

7 References

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1992). Testing causality between two vectors in multivariate ARMA models. *Journal of the American Statistical Association* 87, 1082–1090.

Coondoo, D., Dinda, S., (2002). Causality between income and emission: a country group-specific econometric analysis. *Ecological Economics*, 40, 351-367.

Corduas, M., (2000). La metrica Autoregressiva tra modelli ARIMA: una procedura operativa in linguaggio GAUSS. *Quaderni di Statistica*, 2, 1-37.

Corduas, M., Piccolo, D. (2008). Time series clustering and classification by the autoregressive metric. *Computational Statistics & Data Analysis*, 52(4), 1860-1862.

Dolado, J., Lütkepohl H. (1996). Making Wald Tests Work for Cointegrated VAR Systems. *Econometric Review*, 15, 369-386.

Ericsson, N.R., Irons, J.S., Tryon, R.W., 2001. Output and Inflation in the Long Run. Journal of Applied Econometrics, 16, 241-253.

Giles, D.E.A., 1997. Causality between the measured and underground economies in New Zealand. Applied Economics Letters, 4, 63-67.

Gonzalo, J., Lee, T.H. (1996). Relative power of t type tests for stationary and unit root processes, *Journal of Time Series Analysis*, 17, 37-4.

Granger C.W.J. (1969). Investigating Causal Relations by Econometric Methods and Cross-Spectral Methods. *Econometrica* 37: 424-438

Grimaldi, S. (2004). Linear parametric models applied on daily hydrological series, *J. Hydrol. Eng.* 9, 383-39.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer: Berlin.

Maharaj, E.A. (1996). A significance test for classifying ARMA models. *Journal of Statistical Computation and Simulation*, 54, 305-331.

Mavrotas, G., Kelly, R. (2001). Old Wine in New Bottles: Testing Causality between Savings and Growth. *The Manchester School*, 69, 97-105.

MacKinnon J. (2002). Bootstrap inference in econometrics, *Canadian Journal of Economics*, 35, 615-645.

Metaxoglou, K., Smith, A. (2007). Estimating VARMA models using EM algorithm. *Journal of time Series Analysis*, 28, 666-685.

Otranto, E. (2008). Clustering heteroskedastic time series by model-based procedures. *Computational Statistics & Data Analysis*, 52(10), 4685-4698.

Otranto, E. (2010). Identifying financial time series with similar dynamic conditional correlation. *Computational Statistics & Data Analysis*, 54(1), 1-15.

Piccolo, D. (1990). A distance measure for classifying ARIMA models. *Journal of time Series Analysis*, 11, 153-164.

Sims C.A., Stock J.H., Watson M.W. (1990). Inference in Linear Time Series Models with some Unit Roots, *Econometrica*, vol. 58, 113-144.

Shukur, G., Mantalos, P. (2000). A simple investigation of the Granger causality test in integrated-cointegrated VAR systems, *Journal of Applied Statistics*, Vol. 27(8), 1021-1031.

Toda, H.Y., Yamamoto, T. (1995). Statistical inferences in vector autore- gressions with possibly integrated processes, *Journal of Econometrics*, 66, 225-50