Quaderni di Statistica
Vol. 3, 2001

Further results on the asymptotic distribution of the Euclidean distance between MA models

Emma Sarno
Dipartimento di Economia e Statistica, Università della Calabria
E-mail: sarno@unical.it

Summary: In this paper, we study the asymptotic distribution of the Euclidean distance between time series models proposed by Piccolo (1984, 1990) in the case of Moving Averages models comparisons when least squares estimates are used for the unknown parameters. Piccolo showed that, when purely Autoregressive models are compared, the distribution of the distance estimator is a linear combination of independent Chi-squared random variables. Here, invertible Moving Average models are considered to show that under the same assumptions stated by Berks (1974) and Bhansali (1978) for autoregressive model fitting of stationary processes, a similar result holds. Some descriptive statistics are then used to evaluate the effectiveness of such asymptotic result on finite samples sizes under specific model comparisons.

Key words: AR metric, Autoregression fitting, Moving Average models.

1. Introduction

Since its introduction, the Euclidean distance proposed by Piccolo (1984, 1990) to measure the structural discrepancy between time series modelled through the class of admissible ARIMA models, has been applied successfully to address many different questions such as, for instance, classifying time series (Piccolo, 1984; Maharaj, 1996), detecting influential observations (Corduas, 1990), analyzing demographic changes

(Corduas and Piccolo, 1995) or, more recently, checking the adequacy of seasonal adjustments (Corduas and Piccolo, 1999) and evaluating the efficiency of an environmental monitoring network (Costanzo and Sarno, 2000).

In order to asses whether an estimated discrepancy between time series is significant, one needs to know the sample distribution of the AR metric: for comparisons of purely AR models when coefficients are estimated by maximum likelihood, Piccolo (1990) showed that the asymptotic distribution of the Euclidean distance estimator is a linear combination of independent Chi-squared random variables; Corduas (1996) extended this result to the case of stationary processes and proposed an approximation of such distribution for easier computations.

In this paper, we still consider the case of stationary processes -in particular of Moving Averages models- but we focus our attention on least squares estimates rather than on maximum likelihood ones, and show that under the assumptions stated by Berks (1974) and Bhansali (1978) for autoregressive model fitting, a result similar to Piccolo's holds. Remarkable reasons for using least squares consist of providing linear predictions nonparametrically in the time domain, besides their easy computation.

In order to evaluate the effectiveness at finite sample size of the asymptotic distribution of the AR distance estimator, we carry out a Monte-Carlo experiment: Moving Average models are then compared while their parameters vary on the corresponding invertibility regions, to allow for the sensitivity to such values.

The paper is organized as follows. In Section 2, we introduce the AR metric and the autoregression fitting of stationary models. In Section 3, we show that the asymptotic distribution of the AR metric is a linear combination of independent Chi-squared random variables. In Section 4, we illustrate the behaviour of the AR metric under the hypothesis of comparisons between MA(1) or MA(2) models, which results fully consistent with a previous exploratory work (Sarno, 2000). Finally, a brief discussion follows in Section 5.

2. *The AR metric and the autoregression fitting*

Following Box and Jenkins (1976), let $Z_t \sim ARIMA(p,d,q)(P,D,Q)_s$ be a stochastic process such that $\phi(B)\Phi(B^s)\nabla^d\nabla_s^D Z_t = \theta(B)\Theta(B^s)a_t$, where a_t is a Gaussian White Noise process with zero mean and variance σ^2. When $\theta(B)\Theta(B^s) = 0 \rightarrow |B| > 1$, the process is invertible and Z_t admits the $AR(\infty)$ representation $\pi(B)Z_t = a_t$, with $\pi(B) = \phi(B)\Phi(b^s)\nabla^d\nabla_s^D\theta^{-1}(B)\Theta^{-1}(B^s) = 1 - \pi_1 B - \pi_2 B^2 - \ldots$.

Furthermore, let the class of invertible ARIMA models be denoted by L. Then, given initial values and known orders, any process belonging to L is fully characterized by σ^2 and the AR weights sequence associated to its $AR(\infty)$ formulation.

A measure of structural diversity between two processes, say $X_t \in L$ and $Y_t \in L$, that compares their respective sequences of AR weights $\pi^{(X)} = \left(\pi_1^{(X)}, \pi_2^{(X)}, ...\right)'$ and $\pi^{(Y)} = \left(\pi_1^{(Y)}, \pi_2^{(Y)}, ...\right)'$ by means of an Euclidean distance, is the AR metric ${}_{AR}d(X_t, Y_t) = \sqrt{\sum_{j=1}^{\infty}\left(\pi_j^{(X)} - \pi_j^{(Y)}\right)^2}$ proposed by Piccolo (1984, 1990). It follows that, considering a convenient finite approximation of the above $AR(\infty)$ representations and suitable estimates of the AR coefficients leads to the AR distance estimator[1]

$$ {}_{AR}\widehat{d}_L(X_t, Y_t) = \sqrt{\sum_{j=1}^{L}\left(\widehat{\pi}_j^{(X)} - \widehat{\pi}_j^{(Y)}\right)^2}. \tag{1} $$

In this work, we investigate the asymptotic bahaviour of such estimator when least squares estimates of the AR parameters, say $\widehat{\pi}_L = (\widehat{\pi}_{L,1}, \widehat{\pi}_{L,2}, ..., \widehat{\pi}_{L,L})$, are plugged in (1). This refers to Berks's (1974) and Bhansali's (1978) work on fitting autoregression of order L, while L has to diverge with the sample size, to derive asymptotically efficient estimates of the parameters of purely nondeterministic stationary processes.

In particular, our results rely upon the following theorem where the distribution of the AR coefficients is elicited under common hypotheses:

[1] Detailed discussions on the properties of the AR distance estimator are in Piccolo (1989; 1990) and Corduas (1992; 1996).

Theorem 1 (Bhansali)

Assume that the following condititions given by Berks (1974) are satisfied: (i) $\pi(e^{i\omega})$ *is nonzero,* $-\pi < \omega \leq \pi$*; (ii)* $E(a_t^4) < \infty$*; (iii) The choice of the truncation point* L *in terms of time series length* n *is such that* $L^3/n \to 0$*; (iv) The choice of the truncation point* L *in terms of time series length* n *is such that* $n^{1/2}\left(|a_{L+1}| + |a_{L+2}| + ...\right) \to 0$*. Then, the joint asymptotic distribution of the least square estimators of the AR coefficients, i.e.* $n^{1/2}\left(\widehat{\pi}_{L,i} - \pi_i\right)$ *and* $n^{1/2}\left(\widehat{\pi}_{L,j} - \pi_j\right)$*, for* $i, j = 1, 2, ..., L$*, is bivariate Normal with zero means and asymptotic covariance* $V = \left\{n^{-1}v_{i,j}\right\}$*, with* $v_{i,j} = nCov(\widehat{\pi}_{L,i}, \widehat{\pi}_{L,j}) = \sum_{p=0}^{i-1}\pi_p\pi_{p-i+j}$*, for* $1 \leq i \leq j \leq L$*.*

Proof

(See Bhansali, 1978).

3. Asymptotic results

In this section, we present the extension of Piccolo's (1989) and Corduas's (1996) results by eliciting the distribution of the squared Euclidean distance estimator under least squares estimates of the AR weights.

Theorem 2

Let X_t *and* Y_t *be two independent ARMA processes, and let* $\widehat{\pi}_L^{(X)}$ *and* $\widehat{\pi}_L^{(Y)}$ *be the least square estimates of the coefficients corresponding to their autoregression fitting. Under a null hypothesis* $H_0 : \pi_L^{(X)} = \pi_L^{(Y)}$ *and the same conditions of Theorem 1, the asymptotic distribution of* ${}_{AR}\widehat{d}_L^2(X_t, Y_t)$ *is a linear combination of independent Chi-squared random variables with* 1 *degree of freedom.*

Proof

From Theorem 1, $\widehat{\pi}_L^{(X)} \sim N(\pi_L^{(X)}, V^{(X)})$ and $\widehat{\pi}_L^{(Y)} \sim N(\pi_L^{(Y)}, V^{(Y)})$. Following Piccolo (1989), the quantity ${}_{AR}\widehat{d}_L^2(X_t, Y_t) = \left(\widehat{\pi}_L^{(X)} - \widehat{\pi}_L^{(Y)}\right)' \left(\widehat{\pi}_L^{(X)} - \widehat{\pi}_L^{(Y)}\right) = 2\sum_{k=1}^{r}\lambda_k\chi^2_{g_k}$, where λ_k, $k = 1, 2, ..., r$ and $r \leq L$ are the positive eigenvalues of V and $\chi^2_{g_k}$ are Chi-squared random variables with g_k degrees of freedom, where $g_k \equiv 1, \forall k$, since coincident roots are an almost impossible event for real data.

Furthermore, as a linear combinations of independent random variables the distribution of ${}_{AR}\widehat{d}_L^2(X_t, Y_t)$ has mean, variance and asymmetry (Corduas, 1992) given by:

$$E\left({}_{AR}\widehat{d}_L^2(X_t, Y_t)\right) = 2\sum_{k=1}^{r} \lambda_k E(\chi_1^2) = 2\sum_{k=1}^{r} \lambda_k = 2tr(\mathbf{V}), \qquad (2)$$

$$Var\left({}_{AR}\widehat{d}_L^2(X_t, Y_t)\right) = 4\sum_{k=1}^{r} \lambda_k^2 Var(\chi_1^2) = 8\sum_{k=1}^{r} \lambda_k^2 = 8tr(V^2), \qquad (3)$$

$$Asym\left({}_{AR}\widehat{d}_L^2(X_t, Y_t)\right) = \left[\frac{1}{2}tr(V^2)\right]^{-3/2} tr(V^3). \qquad (4)$$

4. Monte-Carlo results

In order to check the adequacy at finite sample sizes of the asymptotic results given in Theorem 2, we carried out a Monte-Carlo experiment on $N = 5000$ replicates in GAUSS 3.2. In particular, we restricted our attention to the MA(1) and the MA(2) models, with parameters tuned on their corresponding invertibility region. Nevertheless, it must be taken into account that close to the boundaries of such regions, during simulations, it is easy to generate models that are not invertible once estimated. Therefore, an artificial inflation of the upper tails of the empirical distributions of ${}_{AR}\widehat{d}_L^2(X_t, Y_t)$ is expected in our Monte-Carlo experiments. Finally, the asymptotic distributions were compared by means of descriptive statistics as the mean, standard deviation and asymmetry corresponding to (2), $\sqrt{(3)}$ and (4) of Section 3 respectively, to appreciate the fitting to finite Monte-Carlo distributions.

4.1 The MA(1) case

Let $X_t = a_t - \theta a_{t-1}$ or equivalently $X_t \simeq \pi_1 X_{t-1} + \pi_2 X_{t-2} + ... + \pi_L X_{t-L} + a_t$ with $\pi_j = \theta\pi_{j-1} = -\theta^j$ for $j > 0$. This model is invertible for $-1 < \theta < 1$, and for values close to the boundaries of this invertibility region ${}_{AR}d(X_t, Y_t) \to \infty$. Moreover, ${}_{AR}d(X_t, Y_t)$ is symmetric in θ. In Table 1, a few summary statistics referred to the asymptotic distribution of the AR metric (in Times New Roman) and the Monte-Carlo ones (in Italics) are reported for various sets of values of the sample size and the truncation point, in particular$(n, L) = \{(100, 5), (125, 5), (200, 6), (216, 6)\}$. In Figure 1, instead, both distributions are depicted through their mean values for a more immediate visual comparison.

Indeed from Berks and Bhansali we know that L must be set equal to $n^{1/3}$ to ensure efficiency and so it is interesting to analize the role played by the truncation point in our setting. Results show that for $|\theta| < 0.7$ the asymptotic and Monte-Carlo means and standard deviations are quite close but not the asymmetry. It follows that any hypothesis test based on such asymptotic distribution may risk to have a significance level higher than the nominal one for these values. On the contrary, for $|\theta| \geq 0.7$ a similar test would be more conservative and likely less powerful. Finally, we may notice that the closer is the truncation point to $n^{1/3}$, the greater is the overall reduction in the discrepancies between the asymptotic and the finite distributions (smaller error are pointed out by the "$\star$" symbol).

4.2 The MA(2) case

The AR representation of the MA(2) model $X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ has the following $\pi-$weights: $\pi_1 = -\theta_1, \pi_j = \theta_1 \pi_{j-1} + \theta_2 a_{t-2}$ for $j > 1$. This model is invertible if $-1 < \theta_2 < 1$, $\theta_2 - \theta_1 < 1$ and $\theta_2 + \theta_1 < 1$, and ${}_{AR}d(X_t, Y_t)$ is symmetric in θ_1.

A few representative couples of values of θ_1 and θ_2, corresponding to different autocorrelation structures, have been selected to carry out our analyses. Results are summarized in Table 2 for $(n, L) = \{(100, 5), (125, 5), (200, 6), (216, 6)\}$. We notice that at the boundaries of the invertible region, distributions always spread out but the Monte-Carlo ones much more for the same reasons already commented in Section 4. Again, the closer is the truncation point to $n^{1/3}$, the better is the reduction in the discrepancies between the asymptotic and the finite distributions, but less evident patterns due to changes in the values of the parameters θ_1 and θ_2 can be appreciated here.

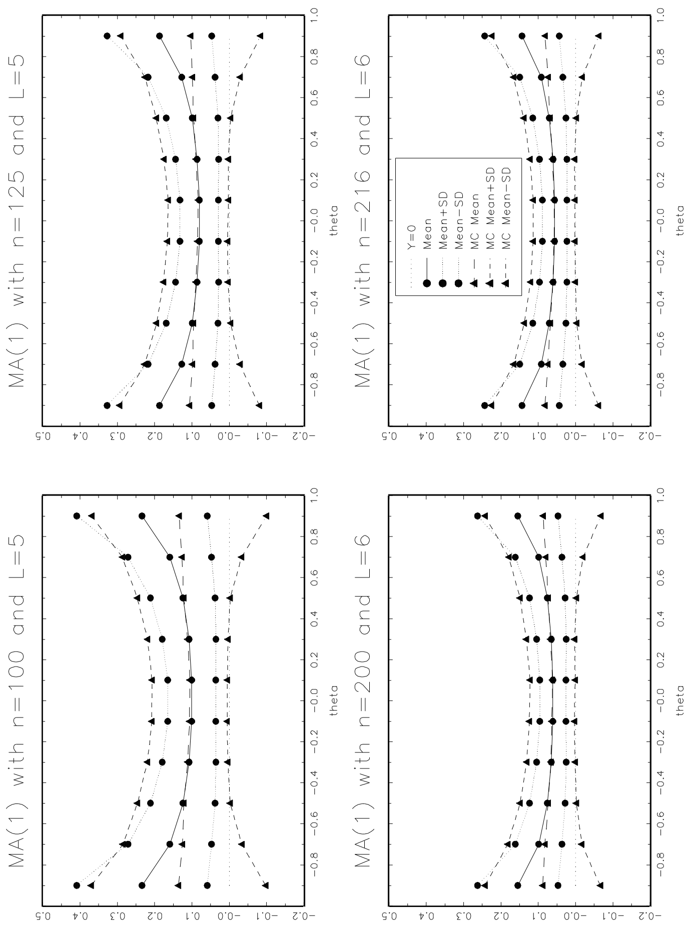

Figure 1. The asymptotic and the Monte-Carlo means of the AR distance estimator between MA(1) models.

Table 1. Asymptotic and finite results in the MA(1) case

θ	n=100 L=5			n=125 L=5		
	Mean	SD	Asym	Mean	SD	Asym
0.1	0.1008	0.0643	⋆1.2934	⋆0.0806	⋆0.0514	1.2934
	0.1061	*0.0685*	*1.3159*	*0.0832*	*0.0538*	*1.4084*
0.3	0.1077	0.0720	⋆1.4256	⋆0.0862	⋆0.0576	1.4256
	0.1114	*0.0758*	*1.5638*	*0.0890*	*0.0604*	*1.5940*
0.5	⋆0.1245	0.0866	⋆1.4644	0.0996	⋆0.0693	1.4644
	0.1221	*0.0958*	*2.0862*	*0.0961*	*0.0747*	*2.1317*
0.7	0.1595	⋆0.1119	1.4138	⋆0.1276	0.0895	⋆1.4138
	0.1266	*0.1079*	*2.2356*	*0.0988*	*0.0846*	*2.1706*
0.9	0.2340	0.1751	1.6839	⋆0.1872	⋆0.1401	⋆1.6839
	0.1344	*0.1223*	*2.4447*	*0.1037*	*0.0918*	*2.0585*

θ	n=200 L=6			n=216 L=6		
	Mean	SD	Asym	Mean	SD	Asym
0.1	⋆0.0605	0.0352	1.1817	0.0560	⋆0.0326	⋆1.1817
	0.0617	*0.0372*	*1.4002*	*0.0579*	*0.0344*	*1.2738*
0.3	0.0648	0.0396	1.3028	⋆0.0600	⋆0.0367	⋆1.3028
	0.0674	*0.0420*	*1.4162*	*0.0621*	*0.0391*	*1.3566*
0.5	0.0756	0.0479	1.3293	⋆0.0700	⋆0.0443	⋆1.3293
	0.0732	*0.0522*	*1.8910*	*0.0684*	*0.0480*	*1.8079*
0.7	0.0991	⋆0.0625	⋆1.2651	⋆0.0917	0.0578	1.2651
	0.0795	*0.0649*	*2.1666*	*0.0741*	*0.0603*	*2.3771*
0.9	0.1548	0.1079	1.6880	⋆0.1433	⋆0.0999	⋆1.6880
	0.0872	*0.0770*	*2.3601*	*0.0813*	*0.0697*	*2.3094*

Table 2. Asymptotic and finite results in the MA(2) case

		n=100 L=5			n=125 L=5		
θ_1	θ_2	Mean	SD	Asym	Mean	SD	Asym
-0.2	0.5	0.1244	0.0878	1.5180	$\star$0.0995	$\star$0.0703	$\star$1.5180
		0.1018	*0.0719*	*1.6878*	*0.0802*	*0.0556*	*1.5864*
0.2	-0.5	0.1181	0.0838	1.5718	$\star$0.0945	$\star$0.0670	$\star$1.5718
		0.1101	*0.0752*	*1.4006*	*0.0884*	*0.0638*	*1.6099*
0.5	-0.6	0.1364	$\star$0.1026	1.6802	$\star$0.1091	0.0821	$\star$1.6802
		0.1232	*0.1026*	*2.1910*	*0.0976*	*0.0805*	*1.9637*
0.8	0.1	0.2096	0.1551	$\star$1.6356	$\star$0.1677	$\star$0.1240	1.6356
		0.1234	*0.1079*	*2.5083*	*0.0974*	*0.0859*	*2.6200*
1.2	-0.7	0.2524	0.1871	1.5935	$\star$0.2019	$\star$0.1497	$\star$1.5935
		0.2478	*0.2985*	*3.3986*	*0.1976*	*0.2318*	*2.9595*
1.5	-0.9	0.4084	0.3085	$\star$1.6528	$\star$0.3267	$\star$0.2468	1.6528
		0.4166	*0.5058*	*2.5536*	*0.3318*	*0.4024*	*2.6166*

		n=200 L=6			n=216 L=6		
θ_1	θ_2	Mean	SD	Asym	Mean	SD	Asym
-0.2	0.5	0.0772	$\star$0.0496	1.3642	$\star$0.0715	0.0460	$\star$1.3642
		0.0641	*0.0444*	*2.1085*	*0.0595*	*0.0400*	*1.6045*
0.2	-0.5	0.0725	0.0469	1.4162	$\star$0.0671	$\star$0.0434	$\star$1.4162
		0.0676	*0.0453*	*1.6449*	*0.0624*	*0.0421*	*1.6021*
0.5	-0.6	0.0849	$\star$0.0582	$\star$1.4986	$\star$0.0786	0.0539	1.4986
		0.0727	*0.0508*	*1.5775*	*0.0684*	*0.0484*	*1.7528*
0.8	0.1	0.1380	$\star$0.0947	$\star$1.6228	$\star$0.1278	0.0877	1.6228
		0.0804	*0.0686*	*2.3429*	*0.0775*	*0.0646*	*2.5384*
1.2	-0.7	0.1616	$\star$0.1109	$\star$1.5037	$\star$0.1497	0.1027	1.5037
		0.1173	*0.1110*	*2.9364*	*0.1069*	*0.1059*	*3.0031*
1.5	-0.9	0.2682	0.1905	$\star$1.6747	$\star$0.2483	$\star$0.1764	1.6747
		0.2504	*0.3233*	*3.3163*	*0.2283*	*0.2844*	*3.2570*

5. *Some concluding remarks*

In this work, we provided the asymptotic distribution of the AR distance estimator when least squares estimators are used for the coefficients of purely nondeterministic stationary and invertible processes comparisons. This can broaden the applicability of testing structural discrepancy between time series if received in a software routine as proposed by Corduas (2000) for maximum likelihood estimators. Finally, we notice that although results are asymptotic, they do not deceive at finite sample sizes even if the respect of all requirements on the truncation point is recommended.

Acknowledgments: This work was supported by UNICAL and MURST funds (PRIN 2000/2001: "Modelli stocastici e metodi di simulazione per l'analisi di dati dipendenti").

References

Berks, K. N. (1974) Consistent autoregressive spectral estimates, *The Annals of Statistics*, 2, 3, 489-502

Bahnsali, R. J. (1978) Linear prediction by autoregressive model fitting in the time domain, *The Annals of Statistics*, 6, 1, 224-231

Box, G. E. P., Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco

Corduas, M. (1990) Detecting influential observations in ARIMA models, *Centro di Specializzazione e Ricerche*, Portici (NA), RS 2/90

Corduas, M. (1992) Misure di distanza tra serie storiche e modelli parametrici, *Quaderni dell'Istituto Economico-Finanziario*, Università di Napoli Federico II, vol. 3

Corduas, M. (1996) Uno studio sulla distribuzione asintotica della metrica autoregressiva, *Statistica*, LVI, 3, 321-332

Corduas, M (2000) La metrica Autoregressiva tra modelli ARIMA: una procedura in linguaggio GAUSS, *Quaderni di Statistica*, 2, 1-38

Corduas, M., Piccolo, D. (1995) Mutamenti strutturali della natalità e metodi statistici, *Atti del Convegno SIS: Continuità e discontinuità dei processi demografici*, Arcavacata di Rende, 315-322

Corduas, M., Piccolo, D. (1999) On the use of the AR metric for seasonal adjustments, *Atti CLADAG*, 1-4

Costanzo, G. D., Sarno, E. (2000) La metrica Autoregressiva per la valutazione e ottimizzazione di sistemi di monitoraggio ambientale, *Quaderni di Statistica*, 2, 205-220

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

Piccolo, D. (1984) Una topologia per i modelli ARIMA, *Statistica*, XLIV, 47-59

Piccolo, D. (1989) On a measure of dissimilarity between ARIMA models, *Proceedings of the Business and Economic Statistics Section of the ASA Meeting*, Washington, 231-236

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

Sarno, E. (2000) The behaviour of the AR metric for MA model comparisons, *Atti della XL Riunione Scientifica SIS*, Firenze, 149-152