Research Journal of Mathematical and Statistical Sciences ISSN 2320-6047
Vol. **8(1),** 35-38, January **(2020)** *Res. J. Mathematical and Statistical Sci.*

Short Communication

Type I error concerning the asymptotic distribution of Hotelling's T^2 for autoregressive processes of order 1

Athanase Polymenis
Department of Economics, University of Patras, 26504 Rio-Patras, Greece
athanase@upatras.gr

Available online at: www.iscamaths.com , www.isca.in , www.isca.me
Received 4th June 2019, revised 13th November 2019, accepted 4th January 2020

Abstract

It has been shown in existing literature that the probability of Type I error for the asymptotic distribution of a Student's t-type of statistic concerning usual autoregressive processes of order 1 increases with autocorrelation. In the present paper this problem is also investigated in a multidimensional context using a Hotelling's T^2 - type of statistic and the effect of the dimension of the process on the Type I error is discussed.

Keywords: Type I error, Hotelling's statistic, autoregressive process, asymptotic distribution.

Introduction

Let us consider a multivariate model which consists of a system of p equations where each equation is an AR(1) process taking the form $X_{n,k} = aX_{n-1,k} + s_k\varepsilon_{n,k}$, for $k = 1, \dots, p$, $|a|<1$, $s_k > 0$, and $\varepsilon_{n,k} \sim N(0,1)$. Also consider that vector $X_n \in \mathbb{R}^p$ satisfies the equation $X_{n+1} = aX_n + s\varepsilon_{n+1}$ with $X_n = (X_{n,1}, \dots, X_{n,p})'$ and $\varepsilon_n \sim N(\vec{0}, I)$ are serially independent random vectors (where $\vec{0}$ is the vector with all components equal to 0, I the pxp identity matrix, and s is a diagonal pxp matrix with positive entries so that the vector components of $s\varepsilon_n$ are independent random variables). We call this model Multivariate Model 1. It is obvious that, for any $k = 1, \dots, p$, $X_{n,k} \sim N(0, s^2 \frac{(1-a^{2n})}{(1-a^2)})$, and so this model can be regarded as a multidimensional extension of the usual AR(1) univariate model (i.e. $p = 1$), where we consider a pre-determined value for a; the latter AR(1) model has then the same form as Multivariate Model 1, with $X_n \in \mathbb{R}$, $|a|<1$, $s > 0$, and $\varepsilon_n \sim N(0,1)$. Note that AR(1) models are known to be particularly useful in time series and econometrics[1-3]. Since Type I errors concerning the asymptotic distribution of a Student's t-type of statistic (denoted as t^*) based on this univariate AR(1) process appear in existing literature when a takes pre-specified values[4], the aim of the present paper is to also find Type I errors for the asymptotic distribution of a Hotelling's type of statistic (denoted as T^{*2}) for Multivariate Model 1 using the same values for a as those reported in the aforementioned reference, for reasons of comparison. More specifically the results of V. Belle[4], obtained in a univariate context, clearly show that the chance of a Type I error for t^* increases with autocorrelation a under the null hypothesis of an AR(1) process, and so we will examine whether this result can be extended to Multivariate Model 1, and investigate whether dimension p of Multivariate Model 1 also affects the chance of Type I error, keeping in mind that, for $a = 0$, this probability (which is associated with the usual Hotelling's T^2) remains constant as dimension p increases.

Some Asymptotic Theory

Consider a multivariate AR(1) random process $\{X_n\}$ as described in Polymenis[5] or Polymenis[6], that is, satisfying a recurrence equation of the form $X_{n+1} = AX_n + s\varepsilon_{n+1}$, with initial condition $X_0 = \vec{0}$, where A is a pxp matrix with real entries and with all eigenvalues less than 1 in absolute value, s is a pxp matrix with real entries, X_n is a p-dimensional vector random variable, ε_n is an independently normally $N(\vec{0}, I)$ distributed p - dimensional vector random variable, and I the pxp identity matrix. This multivariate model and consecutive properties have been described in Anderson[7] and Hamilton[3]. Results provided in the references by Polymenis[5,6] showed that the asymptotic (as $n \to \infty$) distribution of a Hotelling's type of statistic denoted as $T^{*2} = n\bar{X}'M^{-1}\bar{X}$ under the null hypothesis that process $\{X_n\}$ satisfies the aforementioned equation, (where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the vector sample mean and $M = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'$ is the sample covariance matrix) takes the form $\delta_1 Z_1^2 + \dots + \delta_p Z_p^2$, where Z_i^2 $(i = 1, \dots, p)$ are i.i.d $\chi^2(1)$ variates, δ_i $(i = 1, \dots, p)$ are eigenvalues of $C^{-\frac{1}{2}}(I - A)^{-1}ss'(I - A')^{-1}C^{-\frac{1}{2}}$, and $C = \text{plim}_{n\to\infty}(M) = \sum_{k=0}^{\infty} A^k ss'(A')^k$. In the particular case of $A = [0]$, i.e., where all entries of A are equal to 0, we obtain $C = ss'$, and so, $C^{-\frac{1}{2}}(I - A)^{-1}ss'(I - A')^{-1}C^{-\frac{1}{2}} = I$, implying that all δ_i $(i = 1, \dots, p)$ are equal to 1; as a result $T^{*2} = T^2$, i.e., the well-known Hotelling's statistic which has its usual asymptotic distribution of $\chi^2(p)$[8], Theorem 5.2.3, with $\mu = \mu_0 = \vec{0}$).

We also mention that the one-dimensional version of this process takes the form $X_{n+1} = aX_n + s\varepsilon_{n+1}$, with $|a| < 1$, $s > 0$, and $\varepsilon_n \sim N(0,1)$, implying that $X_n \sim N(0, s^2 \frac{(1-a^{2n})}{(1-a^2)})$. This is the well-known autoregressive process of order 1 (or AR(1)) with autocorrelation a. For this model the asymptotic distribution of a Student's t-type of statistic denoted as t^* was derived and was found to have the same distribution as the random variable $\sqrt{\frac{1+a}{1-a}} Z_1$, where Z_1 is a standardized normal or, equivalently, $T^{*2} = t^{*2}$ takes the form $\delta_1 Z_1^2$, with $\delta_1 = \frac{1+a}{1-a}$, and Z_1^2 is $\chi^2(1)$ distributed (Lettenmaier[9], Millard, Yearsley and Lettenmaier[10]) or more recent one from Polymenis[6].

Asymptotic Distribution of T^{*2} for Multivariate Model 1

In the sequel we will also use the notation $X_{n+1} = AX_n + s\varepsilon_{n+1}$ for Multivariate Model 1, where matrix A is just the product of a with the identity matrix I, that is, A refers to a square matrix with all diagonal elements equal to a and all other elements equal to 0. Furthermore, as aforementioned, s is a diagonal matrix with positive entries. In this case the vector components of X_n depend only on their own past. It results that Multivariate Model 1 is just a special case of the model appearing in the more general recurrence equation presented in the beginning of the previous section, and so the asymptotic results mentioned in this section can be applied.

In Table-1 of V. Belle[4] Type I error probabilities of the t^* -statistic were provided for the case $p = 1$ using an earlier univariate asymptotic result which coincides with univariate theory reported in the previous section. Since Multivariate Model 1 uses the same coefficient a as that used for its one-dimensional counterpart the aim of this section is to derive the asymptotic distribution of T^{*2} for this model as we now describe. Let us first consider the bivariate $(p = 2)$ random process $\{X_n\}$ with $X_{n+1} = aX_n + s\varepsilon_{n+1}$, where conditions mentioned in the introduction hold true.

Then $X_n = \begin{pmatrix} X_n^{(1)} \\ X_n^{(2)} \end{pmatrix}$, $\varepsilon_{n+1} = \begin{pmatrix} \varepsilon_{n+1}^{(1)} \\ \varepsilon_{n+1}^{(2)} \end{pmatrix}$, and s is a $2x2$ diagonal matrix of the form $\begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix}$, with $s_1 > 0$, $s_2 > 0$; the process can then be written as

$$X_{n+1} = \begin{pmatrix} X_{n+1}^{(1)} \\ X_{n+1}^{(2)} \end{pmatrix} = a \begin{pmatrix} X_n^{(1)} \\ X_n^{(2)} \end{pmatrix} + \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \begin{pmatrix} \varepsilon_{n+1}^{(1)} \\ \varepsilon_{n+1}^{(2)} \end{pmatrix}.$$

This process also takes the form

$$\begin{pmatrix} X_{n+1}^{(1)} \\ X_{n+1}^{(2)} \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} X_n^{(1)} \\ X_n^{(2)} \end{pmatrix} + \begin{pmatrix} s_1 \varepsilon_{n+1}^{(1)} \\ s_2 \varepsilon_{n+1}^{(2)} \end{pmatrix},$$

Where: $A = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$. Using results from Polymenis[5,6] we have $C = \sum_{k=0}^{\infty} A^k ss'(A')^k = \lim_{k\to\infty}\left(I + Ass'A' + A^2 ss'(A')^2 + \cdots + A^k ss'(A')^k\right)$, with $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. But since $A = A' = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} = a \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $ss'A' = a^2 ss'$, $A^2 ss'(A')^2 = a^4 ss'$,…, $A^k ss'(A')^k = a^{2k} ss'$, and so $C = \lim_{k\to\infty}(1 + a^2 + a^4 + \cdots + a^{2k})ss' = \lim_{k\to\infty}(\frac{1-a^{2k+2}}{1-a^2})ss' = \frac{1}{(1-a^2)} ss'$, since $|a| < 1$.

On the other hand, since $I - A = \begin{pmatrix} 1-a & 0 \\ 0 & 1-a \end{pmatrix}$, we have $(I - A)^{-1} ss'(I - A')^{-1} =$
$\begin{pmatrix} 1/(1-a) & 0 \\ 0 & 1/(1-a) \end{pmatrix} ss' \begin{pmatrix} 1/(1-a) & 0 \\ 0 & 1/(1-a) \end{pmatrix} = \left(\frac{1}{1-a}\right) Iss' \left(\frac{1}{1-a}\right) I = \frac{1}{(1-a)^2} ss'$.

Thus
$C^{-\frac{1}{2}}(I - A)^{-1} ss'(I - A')^{-1} C^{-\frac{1}{2}} = \frac{(1-a^2)}{(1-a)^2} (ss')^{-\frac{1}{2}}(ss')(ss')^{-\frac{1}{2}}$
$= \frac{(1-a^2)}{(1-a)^2} I = \left(\frac{1+a}{1-a}\right) I$, with eigenvalues both equal to $\frac{1+a}{1-a}$. It results that the asymptotic distribution of T^{*2} is the same as that of $(\frac{1+a}{1-a})X^2$ where X^2 is $\chi^2(2)$ distributed.

The same reasoning also holds for higher dimensions $p = 3, 4, 5, \ldots$ Thus Eigen values δ_i mentioned in the previous section are equal to $\frac{1+a}{1-a}$, for any $i = 1, \ldots, p$, that is, $\delta = \delta_1 = \delta_2 = \cdots = \delta_p = \frac{1+a}{1-a}$, and so and the asymptotic distribution of T^{*2} is $\delta\ \chi^2(1) + \delta\ \chi^2(1) + \ldots + \delta\,\chi^2(1)$. It results that T^{*2} is asymptotically distributed as $\delta\,\chi^2(p) = (\frac{1+a}{1-a})\chi^2(p)$, which is same as the distribution of $(\frac{1+a}{1-a})X^2$, where X^2 is $\chi^2(p)$ distributed. Remark that for $A = [0]$ we have $T^{*2} = T^2$ (as expected) with an asymptotic distribution equal to $\chi^2(p)$.

These results lead to Theorem 1, which we now state.

Theorem-1: Under the null hypothesis of a random process that satisfies Multivariate Model 1, Hotelling's T^{*2} has the same asymptotic distribution as $(\frac{1+a}{1-a})X^2$, where X^2 is $\chi^2(p)$ distributed.

Results and discussion

In this section we provide probabilities of Type I error for $p = 2, 3$ associated with the asymptotic distribution of T^{*2}, for some pre-determined values of a. In order to check for the effect of the autocorrelation matrix $A = aI$ on the chance of making a Type I error we use the values for a which are reported in Table 1 of V. Belle[4]. In the absence of autocorrelation, i. e., for $a = 0$, we have $\left(\frac{1+a}{1-a}\right) I = I$, and so the eigen values $\delta_1, \delta_2, \ldots, \delta_p$ involved in the asymptotic distribution of T^{*2} are all equal to 1. In the latter case the value of T^{*2} is asymptotically $\chi^2(p)$ distributed. For other values of a the asymptotic distribution of T^{*2} is $(\frac{1+a}{1-a})\chi^2(p)$ according to Theorem 1, and thus, the adjusted value of T^2-statistic is $\frac{\chi^2(p)}{(\frac{1+a}{1-a})}$. This value will now be used for obtaining Type I errors.

For reasons of comparison we use the same values for a as in Table-1 of V. Belle[4]. Let us first consider the case $p = 2$. For $a = 0.1$ or, equivalently, $A = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$, we obtain $\left(\frac{1+a}{1-a}\right) I = \begin{pmatrix} 1.22 & 0 \\ 0 & 1.22 \end{pmatrix}$, with eigenvalues $\delta = \delta_1 = \delta_2 = 1.22$. In this case the adjusted value of T^2 is $\frac{\chi^2(2)}{\delta}$, and so $T^2 = 4.9$ (taking as $\chi^2(2) = 5.99$ - the critical value for 5% significance level). This value is smaller than 5.99 and so the Type I error is increased and is approximately equal to 0.09. In the same way, if $A = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}$, we have $\left(\frac{1+a}{1-a}\right) I = \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix}$, with eigenvalues $\delta = \delta_1 = \delta_2 = 1.5$. The adjusted value of T^2 is then equal to 4. This value is smaller than the one obtained for $a = 0.1$ and Type I error increases as expected and is approximately equal to 0.15. In the same way we find adjusted values of T^2 for $a = 0.3, 0.4$, and 0.5, and we approximate corresponding Type I errors, as these appear in Table-1. We also use the same argument for the case $p = 3$. In this case taking a to be equal to 0.1 we obtain all three eigenvalues δ_i equal to 1.22 and so the adjusted value of T^2 is $\frac{\chi^2(3)}{1.22}$. Since $\chi^2(3) = 7.81$ is the critical value for 5% significance level, we obtain that the adjusted value of T^2 is 6.35. The corresponding Type I-error is then approximately 0.1. In the same way, for $a = 0.2$, we obtain this adjusted value to be 5.2 leading to a Type I error of 0.17. Like for the previous case results for $a = 0.3, 0.4, 0.5$ are provided in Table-1. Note that column "Adjusted Value of t^2-Statistic and Type I Error for $p = 1$ "is the same as columns "Adjusted Value of t-statistic and Type I Error" of Table-1 of V. Belle[4], expressed in terms of χ^2 distribution, and it is also reported in Table-1 for reasons of comparison.

Conclusion

In the present study the asymptotic distribution of a Hotelling's T^2 type of statistic concerning a specific form of multivariate AR(1) processes was provided and corresponding Type I errors for the adjusted values of T^2, calculated for pre-determined values of a, were also presented in the two last columns of Table-1. Comparison of these results which were obtained in cases where $p = 1, 2, 3$, leads then to the following conclusions. Like for the univariate case $p = 1$, the Type I error increases with a for both multivariate cases $p = 2, 3$. We also notice that for a fixed value of a the increase in the Type I error gets more important as the vector dimension p increases. Furthermore V. Belle[4] reports the fact that, in case $p = 1$, "even a modest value of a increases the chance of a Type I error substantially". Our results show that this fact is even more emphasized for dimensions 2 and 3; for instance, in case $p = 3$ and $a = 0.2$, the Type I error probability is more than tripled. We conclude that the dimension of the process in presence of autocorrelation affects substantially the Type I error probability, thus leading to an aggravation of the problems reported by V. Belle[4].

Table-1: Comparison of the effect of autocorrelation on Type I error for $p = 1, 2, 3$ – Assumed 0.05 for $a = 0$. Type I error probabilities are in brackets.

a	δ	Adjusted value of t^2-statistic for $p = 1$	Adjusted value of T^2- statistic for $p = 2$	Adjusted value of T^2- statistic for $p = 3$
0	1	3.84 (0.05)	5.99 (0.05)	7.81 (0.05)
0.1	1.23	3.13 (0.08)	4.9 (0.09)	6.35 (0.1)
0.2	1.5	2.6 (0.11)	4 (0.15)	5.2 (0.17)
0.3	1.86	2.07 (0.15)	3.23 (0.20)	4.2 (0.28)
0.4	2.34	1.64 (0.20)	2.57 (0.27)	3.34 (0.35)
0.5	3	1.28 (0.26)	2 (0.38)	2.6 (0.46)

References

1. Davidson R. and MacKinnon J.G. (1993). Estimation and Inference in Econometrics. Oxford University Press, 327-329. ISBN: 0-19-506011-3.
2. Griffiths W.E., Hill R.C. and Judge G.G. (1993). Learning and Practicing Econometrics. Wiley, 517-520. ISBN: 0-471-51364-4.
3. Hamilton J.D. (1994). Time Series Analysis. Princeton University Press, 53-56 and 257-259. ISBN: 0-691-04289-6.
4. Van Belle G. (2002). Statistical Rules of Thumb Wiley Series in Probability and Statistics, 8-10. ISBN: 0-471-40227-3.
5. Polymenis A. (2008). A note on a validity test using the stochastic EM algorithm in order to assess the number of components in a finite mixture model. *Statistics,* 42(3), 261-274.
6. Polymenis A. (2019). A note on the asymptotic distribution of Student's t for AR(1) processes. *Advances and Applications in Statistics,* 56(1), 37-42.
7. Anderson T.W. (1994). The Statistical Analysis of Time Series. Wiley, N.Y., 188-211. ISBN: 0-471-04745-7.
8. Anderson T.W. (1984). An Introduction to Multivariate Statistical Analysis. Wiley, N.Y, 163. ISBN: 0-471-88987-3.
9. Lettenmaier D.P. (1976). Detection of trends in water quality data from data with dependent observations. *Water Resources Research,* 12, 1037-1046.
10. Millard S.P., Yearsley J.R. and Lettenmaier D.P. (1985). Space-time correlation and its effect on methods for detecting aquatic ecological change. *Canadian Journal of Fisheries and Aquatic Science*, 42(8), 1391-1400.