CHECAGEM DO DIAGNÓSTICO PARA MODELOS AR USANDO AS AUTOCORRELAÇÕES DO QUADRADO DOS RESÍDUOS

Maria Emilia Camargo e Anaelena Bragança de Moraes
Departamento de Estatística. Centro de Ciências Naturais e Exatas. UFSM. Santa Maria, RS.
Gilda B. Sortica
Departamento de Matemática. Centro de Ciências Naturais e Exatas. UFSM. Santa Maria, RS.

RESUMO

Neste trabalho, analisa-se o comportamento da distribuição da função de autocorrelação do quadrado dos resíduos utilizada para a checagem do diagnóstico de modelos AR representativos de séries temporais.

SUMMARY

CAMARGO, M.E.; MORAES, A.B. and SORTICA, G.B., 1986. Diagnostic Checking AR Time Series Models Using Squared-Residual Autocorrelations. Ciência e Natura, 8:19-23.

In this paper, is analised the behaviour of the distribution of squared-residual autocorrelations functions residuals used for diagnostic checking of AR models representative of time series.

INTRODUÇÃO

A necessidade de modelização (ZEITOUN, 1971) do mundo real tem estimulado, de forma intensa, o desenvolvimento de novos métodos, cada vez mais potentes, capazes de descrever com maior grau de adequação as inter-relações entre as variáveis.

Dentre os muitos exemplos de novos métodos quantitativos criados recentemente para simular a realidade e fazer previsões sobre o futuro, destaca-se a metodologia que os Professores George E. P. Box e Gwilym M. Jenkins desenvolveram para analisar o comportamento de variáveis através de séries de tempo.

Essa metodologia foi desenvolvida no decorrer dos anos sessenta e divulgada no ano de 1970, na obra intitulada "Time Series Analysis: Forecasting and Control" (BOX & JENKINS, 1976).

O fundamento do trabalho elaborado por esses dois autores está assentado na possibilidade de alguns modelos lineares que se apresentam potencialmente capazes de descrever, com relativa precisão e de forma parcimoniosa, o comportamento do processo estocástico gerador da série temporal em análise, proporcionando, por conseguinte, previsões de valores futuros.

Nessa metodologia a estratégia para a construção de modelos é baseada em um ciclo interativo, no qual a escolha da estrutura do modelo é baseada nos próprios dados, através da análise das funções de autocorrelação (BOX & JENKINS, 1976).

GRANGER & ANDERSON (1978) introduziram uma nova proposta para o estudo da última etapa, do ciclo iterativo de Box-Jenkins, checagem do diagnóstico de modelos de previsão, baseada no uso da função de autocorrelação dos quadrados dos resíduos, tendo em vista que muitas séries em que "Portmanteu Test" mostrava que os resíduos eram independentes, ao se analisar a série temporal do quadrado dos resíduos, esses apresentavam-se autocorrelacionados.

Neste trabalho, analisou-se várias séries temporais representadas por processos do tipo Auto-Regressivo (AR) em que os resíduos não apresentaram-se autocorrelacionados, mas o quadrado dos resíduos apresentaram-se correlacionados entre si.

MODELO AUTO-REGRESSIVO

Neste trabalho, estudou-se somente os modelos usados para descrever séries temporais estacionárias, chamados por BOX & JENKINS (1976), de modelos Auto-regressivos de ordem p, ou seja, AR (p). O processo auto-regressivo pode ser representado sob a forma explícita da seguinte maneira:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} + a_t \qquad (1)$$

onde:

$$\tilde{Z}_t = Z_t - E(Z_t)$$

sendo $E(Z_t)$ o nível em torno do qual oscila o processo, $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, o operador auto-regressivo de ordem p.

O modelo AR possui p + 2 parâmetros desconhecidos, ou seja, $\phi_1, \dots, \phi_p$; $E(Z_t)$ e $Var(a_t)$, variância da série temporal dos resíduos a_t, que deverão ser estimados a partir dos dados.

A CHECAGEM DO DIAGNÓSTICO

Após o modelo ter sido identificado e seus parâmetros eficientemente estimados, chega-se à última etapa do processo de modelagem, denominada de checagem do diagnóstico, que consiste em averiguar a adequação do modelo utilizado para ajustar os dados observados na vida real.

Essa etapa apresenta-se importante, pois, se a verificação mostrar alguma evidência de que o modelo ajustado é inadequado para descrever os valores reais da série temporal em estudo, é necessário que ela própria (a etapa da verificação) aponte as causas da inadequação e sugira as modificações apropriadas.

Um procedimento bastante utilizado no processo de checagem do diagnóstico está baseado na análise dos resíduos. Esse processo de checagem, em geral, envolve dois estágios. Em primeiro lugar, a função de autocorrelação para a série gerada pelo modelo deve ser comparada com a função de autocorrelação da série de dados realmente observados. Se as duas funções de autocorrelação, se apresentarem bastante distintas entre si, pode-se imaginar que o modelo especificado necessita de uma melhor identificação. Por outro lado, se as duas funções de autocorrelação apresentarem comportamentos semelhantes, então se passa para o segundo estágio, que é uma análise quantitativa dos resíduos gerados pelo modelo identificado.

A análise dos resíduos está baseada no fato de testar-se se realmente os resíduos formam um processo de ruido branco. Fundamentalmente, busca-se testar se os resíduos não estão correlacionados entre si.

Assim, se o modelo está bem especificado, espera-se que os $\hat{a}_t$ sejam não correlacionados entre si, e que a função de autocorrelação dos resíduos

$$\hat{r}_a(k) = \frac{\sum_{t=k+1}^{n} \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^{n} \hat{a}_t^2} \qquad (2)$$

se aproxime de zero para deslocamentos $k \geq 1$.

O teste de independência dos resíduos, sugerido por BOX & JENKINS, é o desenvolvido por BOX & JENKINS (1970) a partir das constatações de ANDERSON (1942) para as funções de autocorrelação. Esse teste comprova que, se o modelo está corretamente especificado, então, para grandes deslocamentos de "k", os coeficientes de autocorrelação $\hat{r}_a(k)$ estão não correlacionados entre si, e estão normalmente distribuidos com média zero e variância $1/n$, onde n é o número de observações da série temporal a_t.

O teste proposto por BOX & JENKINS (1976), conhecido como "Portmanteau test", assume que, se o modelo fixado é correto, então a estatística:

$$Qa = n \sum_{k=1}^{M} \hat{r}_a^2(k) \qquad (3)$$

tem aproximadamente uma distribuição de Qui-quadrado, com (m-p) graus de liberdade, e n=N-d, com a condição de M e n serem suficientemente grande. DAVIES, TRIGGS & NEWBOLD (1977) e LJUNG & BOX (1978) introduziram uma modificação na estatística Q_a

$$Q_a^* = n(n+2) \sum_{k=1}^{M} \hat{r}_a \ (k)/(n-k) \qquad (4)$$

para ser utilizada em pequenas amostras, com distribuição aproximada do Qui-quadrado com (M-p) graus de liberdade.

GRANGER & ANDRESON (1978), ao encontrarem séries temporais em que o quadrado dos resíduos apareciam autocorrelacionados, ainda que os resíduos não sugeriram a utilização da função de autocorrelação do quadrado dos resíduos.

Assim, supondo que N observações $Z_1, \ldots, Z_N$ para uma série temporal são geradas por um AR(p) e que os resíduos a_t são independentes e identicamente distribuidos, o teste de dignificância para a função de autocorrelação do quadrado dos resíduos é dado pela seguinte estatística:

$$Q_{aa}^{**} = n(n+2) \sum_{k=1}^{M} \hat{r}_{aa}^{2}(k)/(n-k) \qquad (5)$$

que segue uma distribuição de Qui-quadrado com (M) graus de liberdade, onde M é o número de lags utilizado.

APLICAÇÃO NA ANÁLISE DE MODELOS AR

É mostrado nesta seção a aplicação da distribuição da função de autocorrelação do quadrado dos resíduos a diversos modelos AR, fazendo-se uma comparação com a distribuição da função de autocorrelação residual.

Na tabela 1, apresenta-se os resultados de Q_a^* e de Q_{aa}^{**} para oito (8) séries econômicas geradas pelo processo AR (1), ou seja:

$$\tilde{Z}_t = \phi\,\tilde{Z}_{t-1} + a_t$$

onde:

$t = 1, \ldots, 50$

M = número de lags utilizado = 20

TABELA 1. RESULTADOS DE Q_a^* E DE Q_{aa}^{**}

ϕ	Q_a^*	Q_{aa}^{**}
-0,15606	28,686	24,431
0,02823	20,938	33,022
0,02412	24,679	32,053
-0,16251	22,741	33,102
0,23341	23,251	32,841
0,25044	27,924	36,403
0,35346	16,351	31,510
0,07937	29,060	34,320

$Q_{95\%,\,19} = 30{,}14$ $\qquad$ $Q_{95\%,\,20} = 31{,}40$

Como a estatística tabelada para um intervalo de confiança de 95% e 19 graus de liberdade é 30,14, pode-se verificar que pela distribuição da função de autocorrelação residual todos os modelos são corretos, enquanto pela distribuição da função de autocorrelação dos resíduos, todos os modelos devem ser melhorados, pois não representam adequadamente os dados observados, para um intervalo de confiança de 95% e 20 graus de liberdade.

CONCLUSÃO

Mostrou-se, neste trabalho, uma aplicação específica da distribuição da função de autocorrelação do quadrado dos resíduos, a modelos de séries temporais gerados por processos AR (1), podendo-se observar que essa distribuição representa mais um método robusto para a checagem do diagnóstico de modelos construídos para representar dados reais observados.

REFERÊNCIAS BIBLIOGRÁFICAS

1. ANDERSON, R.L. Distribution of serial correlation coefficient. *Ann. Math. Statist.*, Baltimore, S.S.Wilks, 13:1-13. Mar. 1942.
2. BOX, G.E.P. & JENKINS, Gwilym M. *Time series analysis:fore casting and control.* San Francisco, Holden-Day, 1976.
3. BOX, G.E.P. & PIERCE, D.A. Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association*, Washington, 65: 1509-26, Dec. 1970.
4. DAVIES, N., TRIGGS, C.M. & NEWBOLD, P. Significance of the Box-Pierce Portmanteau Statistics in Finite Samples. *Biometrika*, 64, 517-22, 1977.
5. GRANGER, C.W. & ANDERSEN, A.P. *An Introduction to Bilinear Time Series Models.* Vandenhoeeck and Ruprecht: Gottingen, 1978.
6. LJUNG, G.M. & BOX, G.E.P. On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65: 297-303.
7. ZEITOUN, Jean. *Mòdeles en urbanisme: une étude critique.* Pacente de Recherche d'Urbanisme, 1971.

Recebido em dezembro, 1986; aceito em dezembro, 1986.