

AVALIAÇÃO MONTE CARLO DE TESTES ASSINTÓTICOS E DE *BOOTSTRAP* PARA AUTOCORRELAÇÃO RESIDUAL

Vânia de Fátima Lemes de MIRANDA¹
Daniel Furtado FERREIRA¹

- RESUMO: A diagnose de autocorrelação pode ser feita formalmente por meio do teste de Durbin-Watson (DW). Neste trabalho objetivou-se avaliar uma abordagem *bootstrap* do teste DW, utilizando simulação Monte Carlo e comparar com um teste de *bootstrap* com e sem correção acelerada de viés aplicado diretamente sobre o parâmetro de autocorrelação (ρ) e com os testes t, t de *bootstrap*, normal para ρ , normal para ρ com correção de viés, normal de Young e sua versão *bootstrap*. Adicionalmente, a qualidade dos três estimadores de ρ foi estudada, avaliando-se o viés e o erro quadrático médio. Simulou-se, para isso, uma estrutura de autocorrelação de primeira ordem. Como critério de avaliação, as taxas de erro tipo I e o poder destes testes foram comparados com aqueles do teste de DW. As principais conclusões obtidas são: o viés dos estimadores de autocorrelação aumenta com o aumento do número de covariáveis e de ρ ; a variância dos estimadores autocorrelação não é afetada pelo aumento de covariáveis e de ρ ; os testes DW, Normal para ρ , teste t para ρ , *bootstrap* para ρ e *bootstrap* para ρ com correção de viés são rigorosos e menos poderosos do que seus competidores; o teste DW é o mais rigoroso de todos e apresentou menor poder; os testes DWB e *bootstrap* de Young (BC) são equivalentes; os testes DW *bootstrap*, BC, t *bootstrap* para ρ , normal para ρ com correção de viés e normal de Young (NC) são considerados os melhores por apresentarem tamanhos de teste idênticos ao valor nominal, maiores poderes e por serem robustos; o teste NC foi considerado rigoroso para pequenos valores de α e n. Além disso, o teste normal para ρ com correção de viés é recomendado neste trabalho devido à simplicidade e facilidade para ser aplicado.
- PALAVRAS-CHAVE: Teste de Durbin-Watson; *bootstrap*; Monte Carlo; autocorrelação.

1 Introdução

Muitas são as situações experimentais em que o pesquisador se depara com a necessidade de ajustar modelos, sejam eles lineares ou não-lineares. As metodologias de análises usuais pressupõem que os resíduos, efeitos aleatórios desses modelos, são aditivos, homocedásticos, normais e independentemente distribuídos. Dentre as técnicas utilizadas para avaliar o atendimento a essas pressuposições destacam-se as análises gráficas dos resíduos.

A autocorrelação é definida como sendo a correlação dos valores de uma variável no instante t com os valores, dessa mesma variável, defasados no tempo. Em geral, a autocorrelação possui interesse maior quando aplicada aos resíduos de um modelo linear

¹Departamento de Ciências Exatas, Universidade Federal de Lavras – UFLA, CEP 37200-000, Lavras, MG. E-mail: vaniafatima@bol.com.br / danielff@ulfa.br

ou não-linear. Os resíduos quando são correlacionados no tempo ou no espaço são ditos autocorrelacionados ou correlacionados serialmente. Para observações ordenadas no tempo, a pressuposição de independência é freqüentemente violada. Especificamente, pode-se verificar que os resíduos são geralmente correlacionados positivamente quando se trabalha com observações ordenadas no tempo (Neter, Wasserman e Dutner, 1974). Se a autocorrelação for maior que zero, diz-se que os erros estão positivamente autocorrelacionados, e se for menor que zero, há autocorrelação negativa (Hoffmann e Vieira, 1987). Uma alternativa muito empregada para detectar se os resíduos da regressão são ou não autocorrelacionados é realizar análises gráficas dos resíduos *versus* as observações ordenadas em ordem cronológica (Draper e Smith, 1998).

Uma diagnose mais formal dessa correlação serial é feita por meio do teste de Durbin-Watson (DW). No entanto, segundo Jeong e Chung (2001), este teste apresenta algumas limitações. A primeira delas e a mais crítica reside no fato de a distribuição da estatística do teste não ser matematicamente tratável e, ainda, ser dependente da matriz de delineamento. Devido a estas duas características, vários autores, tal como Hannan & Terrel (1968), apresentaram aproximações numéricas fornecendo valores críticos, considerando a matriz de delineamento, porém estas aproximações apresentam baixo poder.

Os testes para autocorrelação revisados e estudados por De Carlo e Tryon (1993) são simples e relativamente eficientes. Todavia, nenhum relato de comparação desses procedimentos com os métodos de DW e de suas versões de *bootstrap* foi encontrado na literatura. Também não foram encontrados, na literatura, relatos de versões de *bootstrap* dos testes apresentados por De Carlo e Tryon (1993).

A técnica do *bootstrap* foi considerada sistematicamente por Efron (1979), apesar de genericamente o método ter sido usado em circunstâncias particulares antes disso. A essência do *bootstrap* consiste na idéia de que na ausência de qualquer outro conhecimento da população, a distribuição dos valores encontrados em uma amostra aleatória de tamanho n da população é o melhor guia da distribuição da população. Portanto, para aproximar o que acontece quando a população é reamostrada, basta reamostrar a amostra. A amostragem é feita com reposição, o que é a única diferença entre *bootstrap* e aleatorização em muitas aplicações (Manly, 1998).

Em virtude do que foi mencionado, neste trabalho objetivou-se avaliar a abordagem *bootstrap*, utilizando a simulação Monte Carlo aplicada ao teste de Durbin-Watson, teste de *bootstrap* para o parâmetro de autocorrelação ρ com e sem correção acelerada de viés, teste t , teste t de *bootstrap*, teste normal para ρ , teste normal para ρ com correção de viés, teste normal para C de Young e teste *bootstrap* para C . Adicionalmente, a qualidade dos três estimadores de ρ foi estudada avaliando-se o viés e o erro quadrático médio. Simulou-se, para isso, uma estrutura de autocorrelação de primeira ordem. Como critério de avaliação, as taxas de erro tipo I e o poder desses testes foram comparados com aqueles do teste de DW.

2 Metodologia

Para avaliar a *performance* de vários testes para autocorrelação residual de primeira ordem foi realizado esse trabalho. Os testes avaliados consistem daqueles rotineiramente empregados nos testes de hipótese sobre autocorrelação. As suas respectivas adaptações

para *bootstrap* foram implementadas. Para realizar as avaliações utilizou-se simulação Monte Carlo. Para isso foram simulados dados a partir do modelo linear:

$$y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_k X_{k-i} + u_i, \quad (2.1)$$

em que y_i é a variável dependente relativa à i -ésima observação simulada, b_j é o parâmetro associado à j -ésima variável, $j = 1, 2, \dots, (k-1)$ ou o intercepto, $j = 0$; u_i segue um processo de autocorrelação de ordem 1, AR(1), dado pelo seguinte modelo:

$$u_i = r u_{i-1} + e_i, \text{ com } |r| < 1,$$

em que r é o parâmetro de autocorrelação e e_i é supostamente independente de u_{i-1} , para todo i , $i = 1, 2, \dots, n$.

A hipótese a ser testada é dada por: $H_0 : r = 0$ vs $H_1 : r > 0$. Na simulação, os valores de b_j foram fixados em 1, sem perda de generalidade e os valores de X foram gerados aleatoriamente a partir de uma distribuição uniforme (0, 1), conforme procedimentos descrito por Jeong e Chung (2001). Para o ruído branco (e_i) consideraram-se diferentes distribuições, quais sejam: normal, uniforme e exponencial. Todas as distribuições consideradas possuíam média 0 e variância 1. No caso da exponencial e da uniforme, considerou-se um modelo centrado em zero. Com a distribuição normal, procurou-se propiciar um ambiente favorável ao teste original de Durbin-Watson (1950); e com as demais distribuições, ambientes desfavoráveis foram proporcionados.

Os valores paramétricos de ρ foram 0,0; 0,1; 0,3; 0,5; 0,7 e 0,9, e os tamanhos amostrais (n) foram 10; 20; 50; 100 e 200. Os valores de k (número de variáveis regressoras) considerados foram 3; 5; 10 e 20, suas utilizações dependeram do tamanho da amostra considerado. Para cada configuração foram simuladas 2.000 amostras de Monte Carlo; e no caso dos testes que envolveram *bootstrap*, 2.000 amostras com reposição foram retiradas de cada amostra de Monte Carlo. Em cada configuração foram aplicados 8 testes e o número de resultados significativos (erro tipo I e poder) nas 2.000 amostras de Monte Carlo, considerando significância com mais de 1% e 5%, foi computado. Os testes aplicados estão discutidos nas seções subseqüentes.

2.1 Teste de Durbin-Watson (DW)

Reapresentado o modelo (2.1) matricialmente, obtém-se:

$$\underline{y} = x\underline{\beta} + \underline{u} \quad (2.2)$$

em que \underline{y} é o vetor ($n \times 1$) de observações, x é a matriz ($n \times k$) do modelo, $\underline{\beta}$ é o vetor ($k \times 1$) de parâmetros e \underline{u} o vetor ($n \times 1$) de resíduos.

Assim, inicialmente foi obtido o estimador de quadrados mínimos para $\underline{\beta}$ (Drapper e Smith 1998) por:

$$\hat{\underline{\beta}} = (x'x)^{-1}x'y \quad (2.3)$$

O vetor de erros \underline{u} foi estimado por:

$$\underline{u} = \underline{y} - \underline{x}\hat{\underline{\beta}}. \quad (2.4)$$

Assim, o estimador da estatística de Durbin-Watson foi obtido por:

$$d = \frac{\hat{\underline{u}}' A \hat{\underline{u}}}{\hat{\underline{u}}' \hat{\underline{u}}}, \quad (2.5)$$

em quem a matriz $A_{(n \times n)}$ tem estrutura apresentada em Jeong e Chung (2001). Os valores de d foram confrontados com os limites críticos da tabela de Durbin-Watson (1977).

2.2 Teste de DW *bootstrap* (DWB)

Para cada amostra simulada foi aplicado o teste de DW na sua versão de *bootstrap*. Utilizou-se para isso o procedimento de *bootstrap* recursivo apresentado por Jeong e Chung (2001) e descrito por:

- obter uma estimativa de $\underline{\beta}$ por quadrados mínimos ordinários utilizando a equação (2.3). Obter a estimativa $\hat{\underline{u}}$ pela equação (2.4);
- estimar ρ a partir de $\hat{\underline{u}}$ e calcular $\hat{\underline{\rho}}$. O estimador de quadrados mínimos de ρ (Drapper e Smith, 1998) é dado por:

$$\hat{\rho} = \frac{\sum_{i=1}^n (\hat{u}_i \hat{u}_{i-1})}{\sum_{i=2}^n \hat{u}_i^2}; \quad (2.6)$$

- reamostrar $\hat{\underline{\rho}}$ para construir um vetor de *bootstrap* residual $\hat{\underline{\rho}}^*$;
- recursivamente construir o vetor de resíduos $\hat{\underline{u}}^*$ utilizando a equação $u_i = \rho u_{i-1} + e_i$. Neste estágio, $H_0 : \rho = 0$ deve ser imposta artificialmente para a construção da distribuição nula; portanto, $\hat{\underline{u}}^*$ é igual a $\hat{\underline{\rho}}^*$;
- usando \underline{x} e $\hat{\underline{u}}^*$, criar os pseudovalores \underline{y}^* pela equação (2.1);
- usando \underline{x} e \underline{y}^* , calcular a estatística de DWB (d^*) por:

$$d^* = \frac{\hat{\underline{u}}^{*'} A \hat{\underline{u}}^*}{\hat{\underline{u}}^{*'} \hat{\underline{u}}^*}; \quad (2.7)$$

- repetir 2.000 vezes os passos de (c) a (f) para construir a distribuição de *bootstrap* de d^* (F_{d^*});

- h) testar H_0 comparando os valores críticos de F_d com o valor de d estimado a partir dos dados originais.

A proporção de simulações Monte Carlo em que $H_0: \rho = 0$ foi rejeitada é armazenada. Um total de 4×10^6 análises foi realizado para cada configuração. É conveniente salientar que esse teste está sendo aplicado sob a hipótese nula de $\rho = 0$, pois não existe um valor nulo da distribuição da estatística de Durbin-Watson.

2.3 Teste de *bootstrap* ρ ($B\rho$)

O teste de *bootstrap* é aplicado diretamente no estimador de quadrados mínimos da autocorrelação de primeira ordem. Para isso os passos de (a) a (c) da seção 2.2 são repetidos. Os passos necessários são dados por:

- a) a c) mesmo do DWB;
 d) a partir do resíduo e_1^* , fez-se $\hat{u}_1^* = e_1^*$; os demais valores do vetor \underline{u}^* são recriados recursivamente por:

$$\hat{u}_i^* = \rho \hat{u}_{i-1}^* + e_i^*, \quad i = 2, 3, \dots, n \quad (2.8)$$

- e) uma estimativa de ρ^* é obtida por quadrados mínimos ordinários:

$$\hat{\rho}^* = \frac{\sum_{i=1}^n (\hat{u}_i^* \hat{u}_{i-1}^*)}{\sum_{i=2}^n \hat{u}_i^{*2}}; \quad (2.9)$$

- f) repetir os passos de (c) a (e) para construir a distribuição de *bootstrap* de $\hat{\rho}^* (F_{\hat{\rho}^*})$;
 g) testar $H_0: \rho = 0$ comparando os valores críticos de $F_{\hat{\rho}^*}$ com o valor nulo de ρ (zero).

Novamente foi computada a proporção de rejeições de H_0 nas 2.000 amostras de Monte Carlo geradas por configuração.

2.4 Teste *bootstrap* de ρ com correção de viés acelerado ($B\rho Ca$)

No item 2.3 foi apresentado o estimador de ρ ($\hat{\rho}$) na amostra original e os estimadores de *bootstrap* ($\hat{\rho}_b$). Em cada reamostragem verificou-se $\hat{\rho}_b > \hat{\rho}$. A proporção de vezes em que essa desigualdade foi verificada nas 2.000 reamostragens de *bootstrap* foi computada. Assim, o estimador dessa proporção é definido por:

$$\hat{P}(\hat{\rho}_b > \hat{\rho}) = P \quad (2.10)$$

O próximo passo é estimar Z_0 por :

$$Z_0 = \Phi^{-1}(1-P), \quad (2.11)$$

em que $\Phi^{-1}(t)$ é a inversa da função de distribuição da normal padrão para o argumento t . Da mesma forma foi obtido Z_α (quantil α superior da normal padrão), uma vez que o teste é unilateral, por:

$$Z_\alpha = \Phi^{-1}(1-\alpha). \quad (2.12)$$

A constante a , necessária para a obtenção da estatística do teste, não pode ser derivada de uma forma simples. Uma aproximação é apresentada por Efron e Tibshirani (1993), baseada em Jackknife é dada por:

$$a \approx \frac{\sum_{i=1}^n (\hat{\rho}_i - \hat{\rho}_{-i})^3}{6 \left[\sum_{i=1}^n (\hat{\rho}_i - \hat{\rho}_{-i})^2 \right]^{3/2}}, \quad (2.13)$$

em que $\hat{\rho}_{-i}$ é estimador de ρ da amostra original simulada, eliminando-se a i -ésima observação; e $\hat{\rho}$ é dado por:

$$\hat{\rho} = \frac{\sum_{i=1}^n \hat{\rho}_{-i}}{n}. \quad (2.14)$$

Com esses valores determinam-se, em seguida, as seguintes quantidades:

$$Z_1 = \frac{Z_0 - Z_\alpha}{[1 - a(Z_0 - Z_\alpha)] + Z_0} \quad (2.15)$$

e

$$\rho_1 = \Phi(Z_1). \quad (2.16)$$

Com os valores de ρ_1 e a amostra ordenada obteve-se o quantil $100\rho_1\%$ da distribuição de *bootstrap* de ρ . Esse limite foi denominado de $\hat{\rho}_1$. Assim, a significância empírica foi estimada computando-se a proporção de vezes em que o valor hipotético de ρ ($H_0 : \rho = 0$) foi inferior a $\hat{\rho}_1$ nas 2.000 simulações de Monte Carlo para cada configuração.

2.5 Teste t ($t\rho$)

Para cada simulação realizada aplicou-se o teste t (Mood, Graybill e Boes, 1974) para $\rho = 0$ de uma distribuição normal bivariada, o qual, adaptado para autocorrelação, é dado por:

$$t_c = \frac{\hat{\rho}\sqrt{n-3}}{\sqrt{1-\hat{\rho}^2}}. \quad (2.17)$$

Sob $H_0 : \rho = 0$ a distribuição de t_c é aproximadamente uma t de Student com $\nu = n - 3$ graus de liberdade. Computou-se a proporção de vezes em que o teste apresentou-se significativo nos níveis nominais adotados da mesma forma como foi realizado anteriormente.

2.6 Teste t de *bootstrap* ($t\rho B$)

Em cada simulação realizada, o procedimento de *bootstrap* da seção 2.2 foi implementado e as estimativas de $\hat{\rho}_b$ foram obtidas. A essas 2.000 estimativas o teste de equação (2.17) foi aplicado, substituindo $\hat{\rho}$ por $\hat{\rho}_b$, obtendo t_b . Computou-se a proporção de vezes em que o t_b superou t_c da amostra original para obter a significância de *bootstrap*. Nas 2.000 amostras de Monte Carlo computou-se a proporção de vezes em que essa significância foi inferior aos níveis nominais adotados (1% e 5%). Essas foram as estimativas do erro tipo I ou do poder para esse teste.

2.7 Teste normal para ρ ($N\rho$)

Inicialmente, o estimador do erro padrão de $\hat{\rho}$ apresentado por De Carlo e Tryon (1993), o qual foi proposto por Moran (1948), foi estimado por:

$$S_{(\hat{\rho})} = \frac{n-2}{n\sqrt{n-1}}; \quad (2.18)$$

Posteriormente, aplicou-se um teste normal padrão sob $H_0: \rho = 0$ da seguinte forma:

$$Z_c = \frac{\hat{\rho}\sqrt{n-1}}{n-2}. \quad (2.19)$$

Rejeitou-se $H_0 : \rho = 0$ se $P(Z > Z_c) \leq \alpha$, em que α é o valor nominal de significância. Computou-se a proporção de vezes em que a hipótese foi rejeitada nas 2.000 simulações.

2.8 Teste normal para ρ com correção de viés ($N\rho c$)

O estimador $\hat{\rho}$ é viesado (De Carlo e Tryon, 1993), então se aplica a correção de viés dada por:

$$r_{1+} = \hat{\rho} + \frac{1}{n} \quad (2.20)$$

Da mesma forma como apresentado na seção (2.7), aplicou-se o teste normal-padrão para $H_0 : \rho = 0$. Para isso, substituiu-se $\hat{\rho}$ por r_{1+} na equação (2.19). Os demais passos foram exatamente os mesmos descritos em 2.7.

2.9 Teste normal para C de Young (1941) – (NC)

Os estimadores $\hat{\rho}$ e r_{1+} são baseados em somas de produtos. A estatística C proposta por Young (1941) é um estimador de ρ baseado em somas de diferenças. Esse estimador foi utilizado e é dado por:

$$C = 1 - \frac{\sum_{i=1}^{n-1} (\hat{u}_i - \hat{u}_{i+1})^2}{2 \sum_{i=1}^n (\hat{u}_i - \bar{u})^2}, \quad (2.21)$$

em que $\bar{u} = \sum_{i=1}^n \hat{u}_i / n = 0$ (por definição).

Esse estimador pode ser obtido utilizando a sua relação com $\hat{\rho}$ da equação (2.6) por:

$$C = \hat{r} + \frac{(\hat{u}_n - \bar{u}_g)^2 + (\hat{u}_1 - \bar{u}_g)^2}{2 \sum_{i=1}^n (\hat{u}_i - \bar{u}_g)^2} = 1 - \frac{d}{2}. \quad (2.22)$$

O estimador de erro-padrão de C é dado por:

$$S_{(C)} = \sqrt{\frac{n-2}{n^2-1}}. \quad (2.23)$$

Para o teste de hipótese $H_0 : r = 0$, aplicou-se o teste:

$$Z_c = \frac{C \sqrt{n^2 - 1}}{\sqrt{n - 2}}, \quad (2.24)$$

o qual segue a normal-padrão assintoticamente. Se $Z_c > Z_\alpha$ ou $P(Z_c > Z_\alpha) \leq \alpha$, rejeita-se a hipótese nula. Novamente, computou-se a proporção de vezes em que a hipótese nula foi rejeitada nas 2.000 simulações.

2.10 Teste *bootstrap* para C (BC)

Em cada simulação aplicou-se o procedimento de *bootstrap* da seção 2.2 e obteve-se a estimativa C da equação (2.22), aqui denominada C_b . O seu valor padronizado foi obtido utilizando a equação (2.24), substituindo C por C_b e denominando de Z_{C_b} .

Computou-se a proporção de vezes em que Z_{Cb} superou Z_c da amostra original de Monte Carlo. Nas 2.000 amostras de Monte Carlo, calculou-se a proporção de vezes em que essa significância de *bootstrap* foi inferior aos níveis nominais (1% e 5%).

2.11 Viés e Erro Quadrático Médio (EQM) dos três estimadores de ρ

O viés e o erro quadrático médio dos três estimadores de ρ (\hat{f} , r_{1+} e C) foram computados nas 2.000 simulações, conforme descrição apresentada a seguir. Seja \hat{Q} um estimador de ρ , dado por \hat{f} , r_{1+} ou C ; então, o viés é dado por:

$$\text{Viés}(\hat{Q}) = \frac{\sum_{m=1}^{2000} \hat{Q}_m}{2000} - r. \quad (2.25)$$

E o erro quadrático médio (EQM) de \hat{Q} foi obtido por:

$$\text{EQM}(\hat{Q}) = \frac{\sum_{m=1}^{2000} (\hat{Q}_m - r)^2}{2000}. \quad (2.26)$$

3 Resultados e discussão

3.1 Propriedades dos estimadores de ρ

Na Tabela 3.1.1 estão apresentados o viés dos estimadores $\hat{\rho}$, r_{1+} e C do parâmetro de autocorrelação ρ em função do tamanho da amostra. Para a situação de $\rho = 0$, o viés de $\hat{\rho}$ é negativo e em módulo superior aos vieses de r_{1+} e C . Assim, o $\hat{\rho}$ subestima ρ e, em geral, os estimadores r_{1+} e C superestimam o valor paramétrico. Com o aumento do valor de n , como é apontado por De Carlo e Tryon (1993), os estimadores r_{1+} e C apresentaram estimativas que se aproximaram daquelas obtidas quando se utilizou o estimador $\hat{\rho}$. Assim, os vieses tendem a ser semelhantes. Para pequenas amostras com $\rho = 0$, os vieses de r_{1+} e C são menores do que o viés de $\hat{\rho}$. Sendo nas situações de pequenas amostras, mais críticas, os estimadores r_{1+} e C , por apresentarem menores vieses, são recomendados.

Com o aumento do valor paramétrico de ρ observou-se, para os três estimadores, um aumento do viés em módulo. Os vieses para essas situações dos três estimadores são negativos. Os vieses de C e r_{1+} são bastante similares para todos os valores de ρ e n para $k = 3$. O aumento do valor de n diminui consideravelmente os vieses observados.

Tabela 3.3.1 - Viés dos estimadores $\hat{\rho}$, r_{I+} e C do parâmetro do autocorrelação ρ , $k = 3$ em função do tamanho da amostra

ρ	n					
	10			20		
	$\hat{\rho}$	r_{I+}	C	$\hat{\rho}$	r_{I+}	C
0,0	-0,0954	0,0046	0,00365	-0,0478	0,0022	0,0021
0,1	-0,1526	-0,0526	-0,0497	-0,0729	-0,0230	-0,0230
0,5	-0,3791	-0,2791	-0,2750	-0,1935	-0,1435	-0,1428
0,9	-0,6504	-0,5504	-0,5222	-0,3582	-0,3082	-0,2942
ρ	50					
	10			100		
	$\hat{\rho}$	r_{I+}	C	$\hat{\rho}$	r_{I+}	C
0,0	-0,0132	0,0068	0,0070	-0,0121	-0,0021	-0,0022
0,1	-0,0306	-0,0105	-0,0104	-0,0193	-0,0093	-0,0088
0,5	-0,0833	-0,0633	-0,0644	-0,0431	-0,0331	-0,0336
0,9	-0,1454	-0,1254	-0,1250	-0,0690	-0,0590	-0,0603

Na Tabela 3.1.2 estão apresentados os vieses de $\hat{\rho}$, r_{I+} e C para $k = 10$. Para os estimadores C e r_{I+} , com $\rho = 0$, nota-se que deve haver um ponto de equilíbrio considerando n e k em que o viés será nulo. À medida que ρ afasta de zero, o viés para esses estimadores aumenta, considerando n fixado, e também com o aumento de k .

Tabela 3.1.2 - Viés dos estimadores $\hat{\rho}$, r_{I+} e C do parâmetro do autocorrelação ρ , para $k = 10$ em função de n e ρ

ρ	n					
	10			20		
	$\hat{\rho}$	r_{I+}	C	$\hat{\rho}$	r_{I+}	C
0,0	-	-	-	-0,0518	-0,0018	-0,0011
0,1	-	-	-	-0,1051	-0,0550	-0,0540
0,5	-	-	-	-0,3526	-0,3026	-0,3025
0,9	-	-	-	-0,6177	-0,5676	-0,5604
ρ	50					
	10			100		
	$\hat{\rho}$	r_{I+}	C	$\hat{\rho}$	r_{I+}	C
0,0	-0,0230	-0,0030	-0,0040	-0,0117	-0,0017	-0,0016
0,1	-0,0412	-0,0212	-0,0210	-0,0225	-0,0125	-0,0124
0,5	-0,1488	-0,1288	-0,1305	-0,0714	-0,0614	-0,0622
0,9	-0,2655	-0,2455	-0,2448	-0,1304	-0,1204	-0,1217

Na Tabela 3.1.3 está apresentado EQM dos estimadores $\hat{\rho}$, r_{I+} e C de ρ para $k = 3$ em função de n e de ρ . Pode-se observar, considerando $\rho = 0$, que os EQM de $\hat{\rho}$, r_{I+} e C são bastante similares, embora o r_{I+} apresente valores menores de EQM que os seus competidores. Devido ao fato de o EQM ser função do viés ao quadrado e da variância do estimador, é natural que seu valor diminua com o aumento de n . Como os EQM's são parecidos para $\rho = 0$, infere-se que as variâncias dos três estimadores sejam similares. Da

mesma forma, espera-se que os EQM's de r_{I+} e C sejam ligeiramente menores do que o de $\hat{\rho}$. Esses resultados podem ser comprovados com a consulta da Tabela 3.1.3.

Tabela 3.1.3 - EQM dos estimadores $\hat{\rho}$, r_{I+} e C de ρ para $k = 3$ em função de n e de ρ

ρ	n					
	10			20		
	$\hat{\rho}$	r_{I+}	C	$\hat{\rho}$	r_{I+}	C
0,0	0,0806	0,0715	0,0804	0,0432242	0,0409	0,0429
0,1	0,0976	0,0779	0,0851	0,0482994	0,0435	0,0458
0,5	0,2188	0,1529	0,1546	0,0776732	0,0608	0,0601
0,9	0,4947	0,3746	0,3507	0,1606625	0,1273	0,1171
ρ	50			100		
	$\hat{\rho}$	r_{I+}	C	$\hat{\rho}$	r_{I+}	C
	0,0	0,0192	0,0191	0,0195	0,0099	0,0097
0,1	0,0202	0,0194	0,0199	0,0106	0,0103	0,0137
0,5	0,0241	0,0212	0,0210	0,0098	0,0090	0,0090
0,9	0,0309	0,0255	0,0248	0,0083	0,0070	0,0069

Na Tabela 3.1.4 estão apresentados os EQM's para $k = 10$ dos estimadores $\hat{\rho}$, r_{I+} e C . Para $\rho = 0$, poucas mudanças no EQM foram observadas, pois o viés nessa situação é bem pequeno. Assim, embora os comentários anteriormente realizados para $k = 3$ continuem valendo, em geral, para $k = 10$, a alteração ocorreu somente nos níveis de EQM que foram maiores na mesma situação de $k = 10$ (tabelas 3.1.3 e 3.1.4). As exceções ocorrem geralmente para grandes valores de n . Esses resultados diferem daqueles apresentados por De Carlo e Tryon (1993) devido a esses autores não considerarem o estudo desse problema em situações de regressão, pois utilizaram apenas uma série temporal simples.

Tabela 3.1.4 - EQM dos estimadores de $\hat{\rho}$, r_{I+} e C de ρ para $k = 10$ em função de n e ρ

ρ	n					
	10			20		
	$\hat{\rho}$	r_{I+}	C	$\hat{\rho}$	R_{I+}	C
0,0	-	-	-	0,0464	0,0437	0,0459
0,1	-	-	-	0,0541	0,0461	0,0477
0,5	-	-	-	0,1674	0,1346	0,1355
0,9	-	-	-	0,4256	0,3663	0,3588
ρ	50			100		
	$\hat{\rho}$	r_{I+}	C	$\hat{\rho}$	r_{I+}	C
	0,0	0,0193	0,0188	0,0191	0,0097	0,0096
0,1	0,0205	0,0192	0,0195	0,0101	0,0098	0,0099
0,5	0,0386	0,0330	0,0334	0,0134	0,0121	0,0121
0,9	0,0831	0,0729	0,0721	0,0214	0,0189	0,0191

3.2 Erro tipo I

Nas Tabelas 3.2.1, 3.2.2 e 3.2.3 estão apresentadas as taxas de erro tipo I para os dez testes estudados neste trabalho, considerando valores nominais de 5% de significância e distribuição normal para os resíduos. O teste de DW apresentou taxas de erro tipo I significativamente inferiores ao valor nominal de $\alpha = 5\%$ para todos os valores de n e k simulados. O intervalo de 99% de confiança para testar a hipótese de igualdade da taxa de erro do teste ao valor nominal de 5% foi de 3,83% e 6,39%. Essa hipótese foi testada para eliminar os erros de Monte Carlo presentes na amostra de tamanho 2.000 para cada configuração.

Dessa forma, o teste DW foi considerado rigoroso, ou seja, controlou as taxas de erro tipo I em um nível real de significância abaixo do valor nominal adotado. Com o aumento do valor de k , para um mesmo valor de n , verifica-se que a taxa de erro tipo I do teste DW diminuiu consideravelmente (Tabelas 3.2.1, 3.2.2 e 3.2.3). Da mesma forma, fixando o valor k e aumentando o valor de n , verifica-se que a taxa de erro tipo I empírica deste teste (DW) aumenta, embora com $n \leq 200$ ainda continue sendo significativamente ($P < 0,01$) inferior ao valor nominal.

Tabela 3.2.1 - Taxas de erro tipo I para os dez testes de autocorrelação em função do número de variáveis $k = 3$ e do tamanho da amostra para distribuição normal considerando $\alpha = 5\%$

Testes	n				
	10	20	50	100	200
DW	0,0015*	0,0065*	0,0225*	0,0275*	0,0265*
DWB	0,0515	0,0440	0,0555	0,0510	-
B ρ	0,0015*	0,0135*	0,03050*	0,0420	0,0325*
B ρ Ca	0,0310*	0,0415	0,0450	0,0450	0,0355*
t ρ	0,0025*	0,0140*	0,0360	0,0430	-
t ρ B	0,0520	0,0450	0,0585	0,0525	-
N ρ	0,0240*	0,0270*	0,0455	0,0425	-
N ρ c	0,0590	0,0470	0,0590	0,0500	-
NC	0,0510	0,0480	0,0575	0,0500	-
BC	0,0515	0,0440	0,0555	0,0510	-

Tabela 3.2.2 - Taxas de erro tipo I para os dez testes de autocorrelação em função do número de variáveis $k = 10$ e do tamanho da amostra para distribuição normal considerando $\alpha = 5\%$

Testes	n				
	10	20	50	100	200
DW	-	0,0000*	0,0005*	0,0010*	0,0065*
DWB	-	0,0490	0,0515	0,0485	-
B ρ	-	0,0125*	0,0300*	0,0420	0,0365*
B ρ Ca	-	0,0395	0,0405	0,0465	0,0380*
t ρ	-	0,0165*	0,0265*	0,0385	-
t ρ B	-	0,0490	0,0560	0,0495	-
N ρ	-	0,0290*	0,0365*	0,0410	-
N ρ c	-	0,0500	0,0535	0,0500	-
NC	-	0,0490	0,0500	0,0500	-
BC	-	0,0490	0,0515	0,0485	-

Tabela 3.2.3 - Taxas de erro tipo I para os dez testes de autocorrelação em função do número de variáveis $k = 20$ e do tamanho da amostra para distribuição normal considerando $\alpha = 5\%$

Testes	n				
	10	20	50	100	200
DW	-	-	0,0000*	0,0000*	0,0005*
DWB	-	-	0,0420	0,0490	-
B ρ	-	-	0,0260*	0,0370*	0,0340*
B ρ Ca	-	-	0,0375*	0,0395	0,0330*
$\tau\rho$	-	-	0,0250*	0,0400	-
$\tau\rho$ B	-	-	0,0430	0,0480	-
N ρ	-	-	0,0305*	0,0420	-
N ρ c	-	-	0,0445	0,0490	-
NC	-	-	0,0435	0,0515	-
BC	-	-	0,0420	0,0490	-

O teste DW possui essa deficiência, possivelmente, devido ao fato de não apresentar resultados conclusivos quanto à rejeição ou não da hipótese de nulidade para a correlação serial em algumas regiões dos valores da estatística d observada. Nessa simulação foi mensurada a taxa de resultados não conclusivos do teste que podem ser observados na Figura 3.2.1.

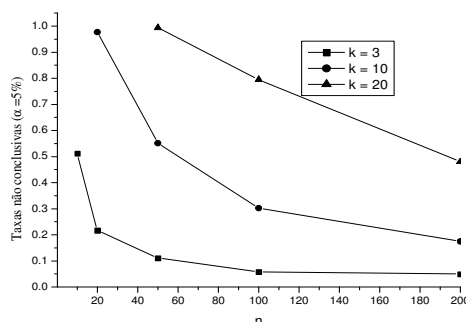


Figura 3.2.1 - Taxas não conclusivas do teste DW para o valor nominal de significância de 5%.

O teste DW na versão *bootstrap* foi nomeado DWB e os resultados encontrados para as taxas de erro tipo I permitem inferir que uma sensível melhora foi obtida. Isso ocorreu devido ao fato de que todas as taxas de erro tipo I observadas para n variando de 10 a 200 e k variando de 3 a 20 foram iguais ao valor nominal de 5%, exceto pelos erros de Monte Carlo presentes nas estimativas dessas taxas. Assim, conclui-se que a versão *bootstrap* desse teste foi eficiente para o controle da taxa de erro tipo I e que o teste DWB apresentou tamanho do teste igual ao valor nominal $\alpha = 5\%$.

Os testes DW e DWB são testes indiretos para a autocorrelação serial. Os testes baseados diretamente na autocorrelação serial estudados nesse trabalho são $\tau\rho$, N ρ , N ρ c e NC. As taxas de erro tipo I desses testes e de suas respectivas versões de *bootstrap* se encontram nas Tabelas 3.2.1, 3.2.2 e 3.2.3. Os testes baseados nos estimadores viesados

de ρ são $t\rho$ e $N\rho$. Para esses dois testes, as taxas de erro tipo I tenderam a ser significativamente diferentes e menores do que o valor nominal $\alpha = 5\%$ para amostras pequenas. O teste normal para ρ apresentou taxas que convergiam mais rapidamente para o valor nominal com o aumento de n do que o teste $t\rho$.

O teste $N\rho_c$, o qual é baseado em um estimador não viesado (r_{1+}), apresentou taxas de erro tipo I iguais ao valor nominal em todas as situações estudadas referentes aos valores de k e n . A simples correção de viés melhorou consideravelmente o resultado do teste $N\rho$. O mesmo resultado foi obtido para o teste NC, em que C é também um estimador de ρ com menor viés do que $\hat{\rho}$ (De Carlo e Tryon, 1993). Assim, os testes $N\rho_c$ e NC possuem tamanhos iguais ao valor nominal de 5% para todas as situações estudadas, sob distribuição normal.

Dentre os testes de *bootstrap* para ρ , o teste Bp foi o que apresentou pior resultado. Para $k = 3$, por exemplo, somente com $n = 100$ a significância encontrada foi semelhante ao valor nominal $\alpha = 5\%$ ($P > 0,01$). Esse teste tendeu a ser rigoroso e deve apresentar menor poder (Mood, Graybill e Boes, 1974). O teste de *bootstrap* para ρ com correção de viés (BpCa) se apresentou melhor no controle da taxa de erro tipo I, em relação Bp. Esse teste tendeu a ser mais rigoroso com pequenas amostras.

Nas Tabelas 3.2.4, 3.2.5 e 3.2.6 estão apresentadas as taxas de erro tipo I para os testes considerando a distribuição uniforme e valor nominal de significância de 5% . Os resultados, em geral, concordam com aqueles obtidos para a distribuição normal. As maiores alterações foram observadas para o teste BpCa. Em algumas situações de k e n houve uma piora (teste rigoroso) em relação à situação de normalidade, em outras, os valores das taxas de erro tipo I se igualaram ($P > 0,01$) ao valor nominal de $\alpha = 5\%$. Isso pode ser visualizado confrontando as Tabelas 3.2.1, 3.2.2 e 3.2.3 com as Tabelas 3.2.4, 3.2.5 e 3.2.6. Os resultados não conclusivos para o teste de DW foram similares aos da distribuição normal e isso não foi apresentado. Assim, por não alterarem a sua *performance* com a mudança de distribuição, os testes foram considerados robustos.

Tabela 3.2.4 - Taxas de erro tipo I para os dez testes de autocorrelação em função do número de variáveis $k = 3$ e do tamanho da amostra para distribuição uniforme considerando $\alpha = 5\%$

Testes	n				
	10	20	50	100	200
DW	0,0000*	0,0075*	0,0215*	0,0040*	-
DWB	0,0470	0,0585	0,0550	0,0565	-
Bp	0,0040*	0,0130*	0,0325*	0,0355*	0,0420
BpCa	0,0400	0,0345*	0,0400	0,0365*	0,0420
$t\rho$	0,0035*	0,0225*	0,0360*	0,0430	-
$t\rho_B$	0,0485	0,0590	0,0555	0,0580	-
$N\rho$	0,0180*	0,0405	0,0415	0,0475	-
$N\rho_c$	0,0505	0,0660	0,0580	0,0595	-
NC	0,0465	0,0605	0,0545	0,0575	-
BC	0,0470	0,0585	0,0550	0,0565	-

Tabela 3.2.5 - Taxas de erro tipo I para os dez testes de autocorrelação em função do número de variáveis $k = 10$ e do tamanho da amostra para distribuição uniforme considerando $\alpha = 5\%$

Testes	n				
	10	20	50	100	200
DW	-	0,0000*	0,0010*	0,0030*	-
DWB	-	0,0435	0,0575	0,0455	-
Bp	-	0,0130*	0,0310*	0,0355*	0,0420
BpCa	-	0,0405	0,0380	0,0365*	0,0420
tp	-	0,0155*	0,0385	0,0375*	-
tpB	-	0,0455	0,0560	0,0470	-
Np	-	0,0265*	0,0455	0,0380*	-
Npc	-	0,0475	0,0575	0,0475	-
NC	-	0,0415	0,0555	0,0445	-
BC	-	0,0435	0,0575	0,0455	-

Tabela 3.2.6 - Taxas de erro tipo I para os dez testes de autocorrelação em função do número de variáveis $k = 20$ e do tamanho da amostra para distribuição uniforme considerando $\alpha = 5\%$

Testes	n				
	10	20	50	100	200
DW	-	-	0,0000*	0,0000*	-
DWB	-	-	0,0485	0,052	-
Bp	-	-	0,0330*	0,0315*	0,0370*
pCa	-	-	0,0440	0,0350*	0,0370*
tp	-	-	0,0325*	0,0365*	-
tpB	-	-	0,0525	0,0490	-
Np	-	-	0,0360*	0,0430	-
Npc	-	-	0,0520	0,0510	-
NC	-	-	0,0510	0,0535	-
BC	-	-	0,0485	0,0520	-

Na Tabela 3.2.7 estão apresentados os resultados das taxas de erro tipo I para os testes de autocorrelação considerando $k = 3$, $\alpha = 5\%$ e distribuição exponencial para gerar uma condição mais inapropriada para distribuição dos resíduos, a qual é uma distribuição assimétrica à direita. Embora essa condição atípica tenha sido simulada, os testes apresentaram praticamente os mesmos resultados obtidos para as distribuições normal e uniforme. Assim, os testes DWB, tpB, NpC, NC e Bc apresentaram tamanho igual ao nominal de 5%. Os demais testes, em geral, para a maioria das situações, foram considerados rigorosos, ou seja, apresentaram taxas de erro tipo I inferiores ($P < 0,001$) ao valor nominal de 5%.

Tabela 3.2.7 - Taxas de erro tipo I para os dez testes de autocorrelação em função do número de variáveis $k = 3$ e do tamanho da amostra para distribuição exponencial considerando $\alpha = 5\%$

Testes	n				
	10	20	50	100	200
DW	0,0025*	0,0080*	0,0170*	0,0270*	0,0285*
DWB	0,0505	0,0525	0,0500	0,0485	-
Bp	0,0025*	0,0140*	0,0275*	0,0385	0,0365*
BpCa	0,0335*	0,0365*	0,0350*	0,0400	0,0380*
tp	0,0040*	0,0195*	0,0340*	0,0360*	-
tpB	0,0495	0,0545	0,0535	0,0480	-
Np	0,0230*	0,0340*	0,0415	0,0405	-
Npc	0,0540	0,0570	0,0545	0,0485	-
NC	0,0535	0,0510	0,0530	0,0495	-
BC	0,0505	0,0525	0,0500	0,0485	-

Na Tabela 3.2.8 estão apresentadas as taxas de erro tipo I para os testes em função de n , para $k = 3$, distribuição normal e valor nominal de significância de 1%. Para esse caso o intervalo de 99% de confiança para a taxa de erro tipo I é de 0,52 e 1,73.

Os testes considerados rigorosos para $\alpha = 5\%$ continuaram rigorosos para $\alpha = 1\%$. O teste DW, da mesma forma como ocorreu para 5%, apresentou, para todos os valores de n estudados, taxas empíricas inferiores ($P < 0,01$) ao valor nominal de 1%. Os melhores testes, ou seja, aqueles que para todos os n apresentaram tamanhos idênticos ao nominal foram também os mesmos. A exceção ocorreu para o teste de NC, o qual, com $n \leq 20$, mostrou-se rigoroso. Assim, o único teste não baseado em *bootstrap* que se apresentou com controle do erro tipo I foi o Npc.

Tabela 3.2.8 - Taxas de erro tipo I para os dez testes de autocorrelação em função do número de variáveis $k = 3$ e do tamanho da amostra para distribuição normal considerando $\alpha = 1\%$

Testes	n				
	10	20	50	100	200
DW	0,0050*	0,0005*	0,0005*	0,0045*	0,0050*
DWB	0,0085	0,0070	0,0075	0,0085	0,0090
Bp	0,0000*	0,0050*	0,0020*	0,0060	0,0060
BpCa	0,0095	0,0060	0,0050*	0,0070	0,0060
tp	0,0000*	0,0020*	0,0025*	0,0040*	0,0070
tpB	0,0085	0,0085	0,0100	0,0090	0,0095
Np	0,0020*	0,0040*	0,0045*	0,0040*	0,0070
Npc	0,0065	0,0095	0,0070	0,0085	0,0085
NC	0,0045*	0,0050*	0,0085	0,0070	0,0090
BC	0,0085	0,0070	0,0075	0,0085	0,0090

É conveniente salientar que o comportamento dos 5 testes que se destacaram nesse trabalho em relação ao controle do erro tipo I para as distribuições uniforme e exponencial foi similar àquele da distribuição normal. Esse resultado evidencia uma *performance* robusta desses testes.

3.3 Poder

Na Tabela 3.3.1 está apresentado o poder dos 10 testes para correlação serial obtidos mediante simulação Monte Carlo sob H_1 , considerando ρ igual 0,1, 0,5 e 0,9 em função de n . De maneira geral, pode-se verificar que o aumento de n provoca aumento do poder, o que é esperado pela teoria dos testes de hipóteses (Mood, Graybill e Boes, 1974). Também se verifica que o aumento de ρ de 0,1 para 0,9 provoca grande aumento no poder dos testes, como é esperado pela teoria. O poder dos testes para $\rho = 0,1$, mesmo quando $n = 20$, pode ser considerado baixo (inferior a 4%). Poder da ordem de 80% ou mais para $\rho = 0,5$ é encontrado com n entre 20 e 50 e para $\rho = 0,9$ com n entre 10 e 20. Para grandes valores de n e ρ , os poderes dos testes tendem a se igualar e se aproximar de 100%.

O teste de DW apresentou o menor poder dentre os 10 testes estudados em todas as configurações apresentadas nas Tabelas 3.3.1, 3.3.2 e 3.3.3. A diferença em relação aos três testes mais poderosos se destaca principalmente para pequenos valores de ρ e n . Em algumas situações, o poder do teste é significativamente ($P < 0,01$) inferior ao valor nominal de significância de 5%. Os resultados para este teste não são similares aos encontrados por Jeong e Chung (2001). O teste Bp foi inferior a todos os outros, superando apenas em poder o teste DW original. Da mesma forma que o teste DW, Bp apresentou poder inferior ao valor nominal de 5% para pequenos valores de n e ρ . O mesmo teste com correção de viés, BpCa, por outro lado, apresentou poder bem maior que o teste sem correção, Bp. Grandes diferenças foram observadas, independentemente do valor de ρ , para pequenas amostras, o que se torna uma grande vantagem desse teste sobre o teste sem correção de viés.

Tabela 3.3.1 - Poder dos dez testes de autocorrelação para $k = 3$, $\rho = 0,1$ em função do tamanho da amostra para distribuição normal considerando $\alpha = 5\%$

Testes	n				
	10	20	50	100	200
DW	0,0040	0,0170	0,0745	0,1600	0,3200
DWB	0,0670	0,0955	0,1645	0,2425	-
Bp	0,0015	0,0030	0,0955	0,1975	0,3640
BpCa	0,0485	0,0425	0,1275	0,2115	0,3695
tp	0,0055	0,0365	0,1210	0,2020	-
tpB	0,0725	0,0995	0,1715	0,2450	-
Np	0,0355	0,0700	0,1415	0,2130	-
Npc	0,0820	0,1040	0,1720	0,2420	-
NC	0,0740	0,1025	0,1660	0,2410	-
BC	0,0670	0,0955	0,1645	0,2425	-

Tabela 3.3.2 - Poder dos dez testes de autocorrelação em função do número de variáveis $k = 3$, $\rho = 0,5$ e do tamanho da amostra para distribuição normal considerando $\alpha = 5\%$

Testes	n				
	10	20	50	100	200
DW	0,0215	0,2525	0,8680	0,9995	1,0000
DWB	0,2110	0,5455	0,9365	0,9975	-
Bp	0,0210	0,3335	0,9230	0,9985	1,0000
BpCa	0,1685	0,4910	0,9375	0,9985	1,0000
tp	0,0425	0,3715	0,9115	0,9970	-
tpB	0,2185	0,5530	0,9365	0,9975	-
Np	0,1285	0,4645	0,9205	0,9970	-
Npc	0,2265	0,5595	0,9365	0,9970	-
NC	0,2090	0,5490	0,9355	0,9975	-
BC	0,2110	0,5455	0,9365	0,9975	-

Tabela 3.3.3 - Poder dos dez testes de autocorrelação em função do número de variáveis $k = 3$, $\rho = 0,9$ e do tamanho da amostra para distribuição normal considerando $\alpha = 5\%$

Testes	n				
	10	20	50	100	200
DW	0,0755	0,7585	0,9990	1,0000	1,0000
DWB	0,4380	0,9145	0,9995	1,0000	-
Bp	0,0635	0,7710	1,0000	1,0000	1,0000
BpCa	0,3530	0,8615	1,0000	1,0000	1,0000
tp	0,0945	0,8085	0,9995	1,0000	-
tpB	0,4040	0,9070	0,9995	1,0000	-
Np	0,2635	0,8710	0,9995	1,0000	-
Npc	0,4110	0,9075	0,9995	1,0000	-
NC	0,4295	0,9135	1,0000	1,0000	-
BC	0,4380	0,9145	0,9995	1,0000	-

O teste tp foi, em ordem crescente de poder, o terceiro colocado, superando os testes DW e Bp. Teve pequeno poder como esses outros dois, para pequenos valores de ρ e n . Já o teste Np apresentou *performance* imediatamente superior ao tp inclusive para valor de n pequeno e grandes valores de ρ tendo poder muito superior ao tp. É conveniente salientar que se pode observar que os poderes dos testes DWB e BC foram exatamente os mesmos. As taxas de erro tipo I também são iguais para esses dois testes. Depreende-se, assim, que esses dois testes de *bootstrap* na verdade são equivalentes. Isso se deve ao fato de que existe uma relação linear perfeita monótona decrescente entre C e d , ou seja, $C = 1 - 0,5d$. Assim, efetivamente têm-se 9 testes diferentes nesse trabalho.

Na Figura 3.3.1 estão apresentados os poderes dos testes em função de n para $k = 5$ e $\rho = 0,5$, para distribuição normal. O mesmo comportamento ocorre em relação à *performance* dos testes quanto ao poder, sendo o pior teste o DW e os melhores os testes NC, Npc DWB, tpB e BC, sendo este último equivalente a DWB e não está apresentado na Figura 3.3.1.

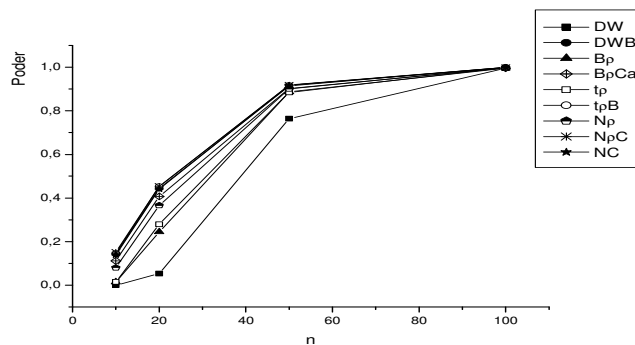


Figura 3.3.1 - Poder dos testes para distribuição normal dos resíduos em função de n com $\rho = 0,5$, $k = 5$ para $\alpha = 0,05$.

Na Figura 3.3.2 estão apresentados os poderes dos 9 testes para $k = 5$ em função de ρ para diferentes valores de n (Figura 3.3.2 a, b, c, e d). Deve ser observado que existem diferenças nas escalas das figuras. Os testes se diferenciam em maior magnitude quando n é pequeno ($n = 10$) e ρ é grande ($\rho = 0,9$). Se n for grande, os testes terão maior diferenciação se ρ for pequeno. Em todos os casos, o teste DW original foi o que apresentou menor poder.

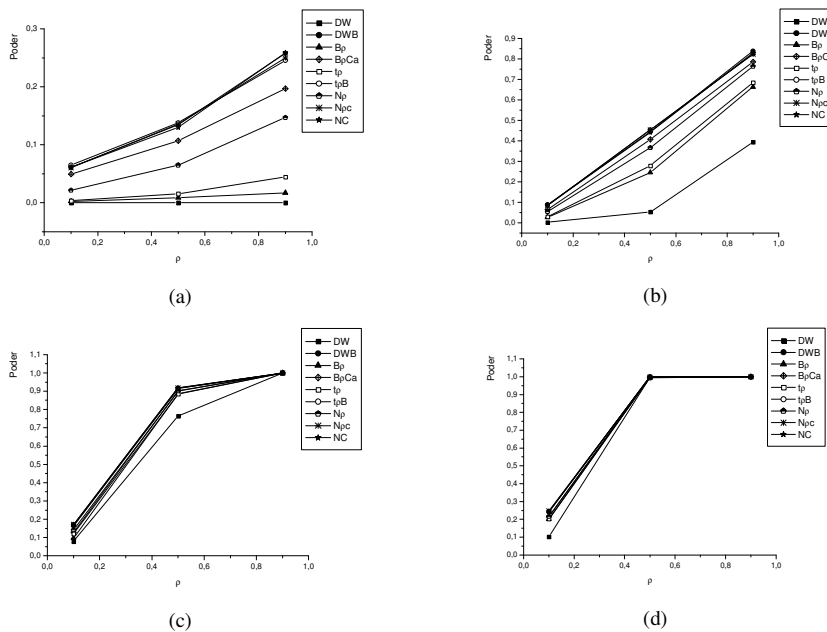


Figura 3.3.2 - Poder dos testes para distribuição normal dos resíduos em função de ρ com $k = 5$, para $\alpha = 0,05$ (a) $n = 10$, (b) $n = 20$, (c) $n = 50$, (d) $n = 100$.

Na Figura 3.3.3 o poder dos testes é apresentado em função de k , para algumas configurações de ρ e de n para distribuição normal. Outro fato que chama a atenção e que ainda não foi destacado refere-se à grande diferença entre os testes B_p e B_pCa . A correção de viés provocou uma grande melhoria no teste.

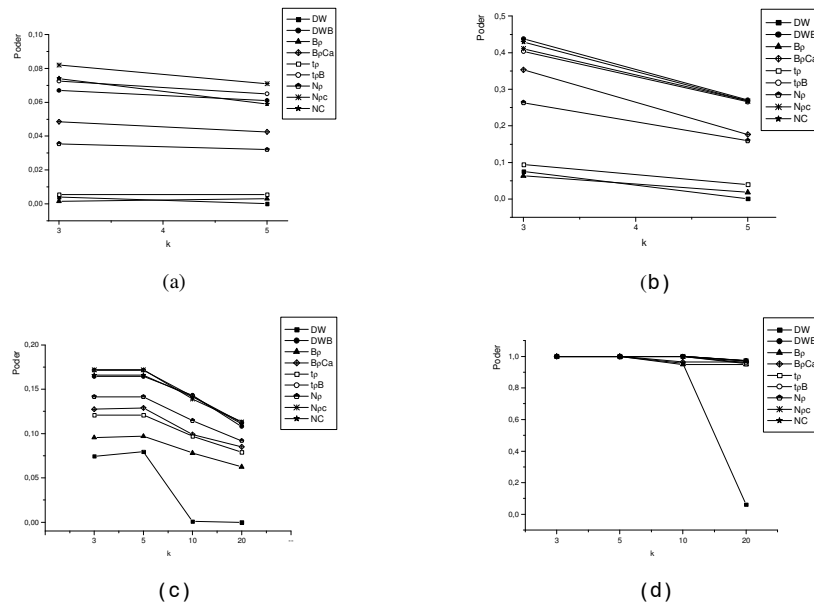


Figura 3.3.3 - Poder dos testes para distribuição normal dos resíduos em função de k para $\alpha=0,05$ (a) $n=10, \rho=0,1$ (b) $n=10, \rho=0,9$ (c) $n=50, \rho=0,1$ (d) $n=50, \rho=0,9$.

Para a distribuição uniforme, as mesmas situações apresentadas para a distribuição normal (Figuras 3.3.1, 3.3.2 e 3.3.3) foram obtidas e estão contempladas nas Figuras 3.3.4, 3.3.5 e 3.3.6. Devido ao fato de os resultados serem praticamente os mesmos, nenhum comentário adicional é feito. Apenas deve ser destacado que os testes, por esse motivo, são considerados robustos, uma vez que é violada a pressuposição de normalidade e os resultados são mantidos.

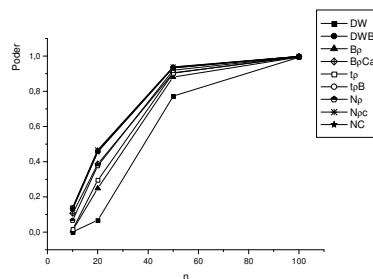


Figura 3.3.4 - Poder dos testes para distribuição uniforme dos resíduos em função de n , com $k=5$ e $\rho=0$, para $\alpha=0,05$.

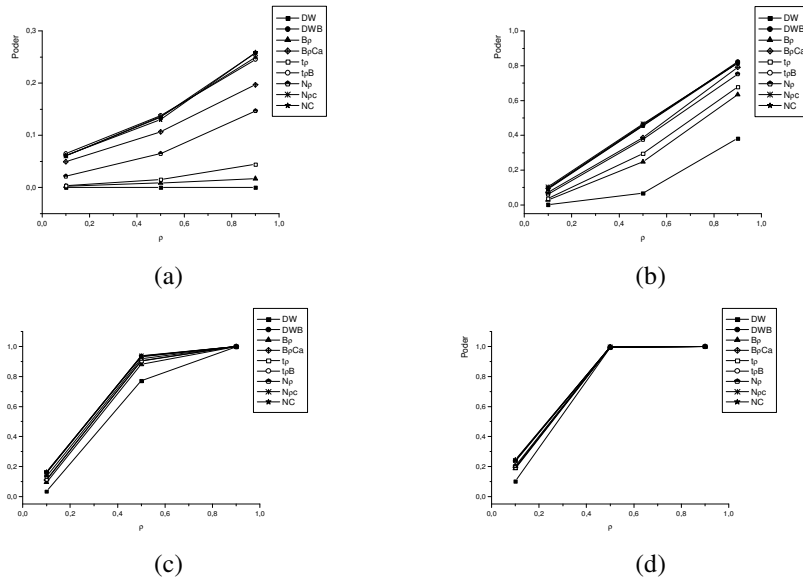


Figura 3.3.5 - Poder dos testes para distribuição uniforme dos resíduos em função de ρ , com $k = 5$ para $\alpha = 0,05$ (a) $n = 10$ (b) $n = 20$ (c) $n = 50$ (d) $n = 100$.

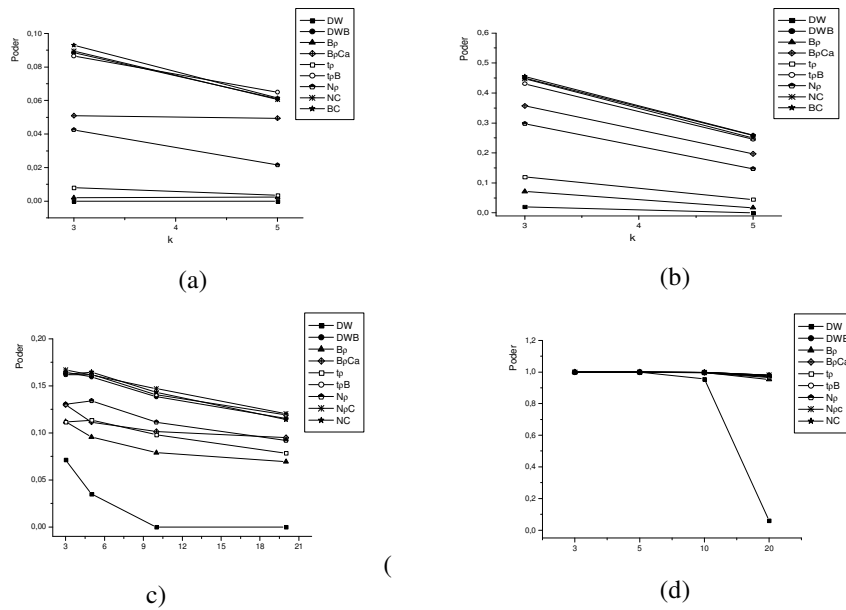


Figura 3.3.6 - Poder dos testes para distribuição uniforme dos resíduos em função de k para $\alpha = 0,05$ (a) $n = 10, \rho = 0,1$ (b) $n = 10, \rho = 0,9$ (c) $n = 50, \rho = 0,1$ (d) $n = 50, \rho = 0,9$.

Nas Figuras 3.3.7 (a), (b), (c) e (d) estão apresentados os poderes dos testes para $\alpha = 1\%$ e distribuição normal. Verificou-se o mesmo padrão de respostas encontrado para o valor nominal de 5%. A exceção novamente ocorre para $n = 10$ com o teste NC, que apresentou menor poder do que os concorrentes tpB, DWB, BC e Npc. Esse resultado está de acordo com os resultados observados para o erro tipo I desse teste, o qual se apresentou rigoroso para pequenas amostras. Para valores de n maiores do que 10 essa diferença praticamente desapareceu.

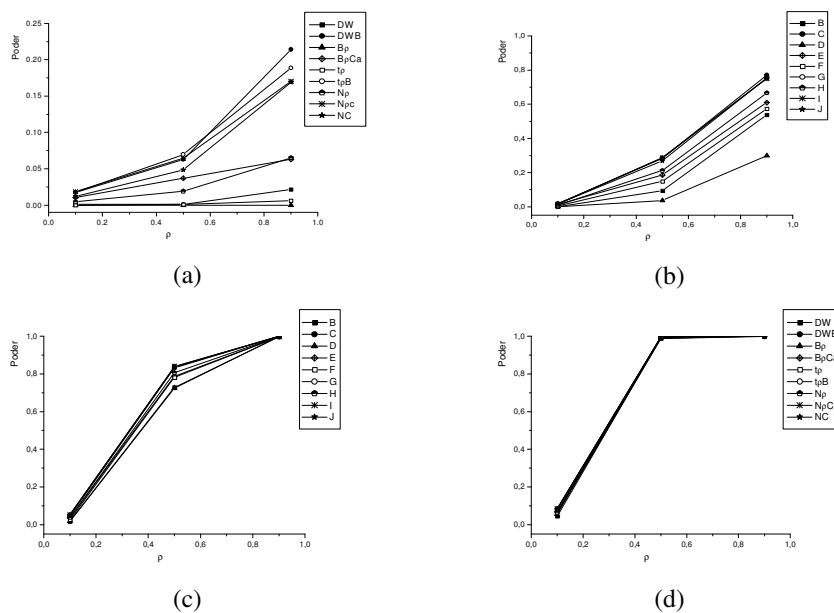


Figura 3.3.7 - Poder dos testes para distribuição normal dos resíduos em função de ρ com $k = 3$ para $\alpha = 0,01$ (a) $n = 10$ (b) $n = 20$ (c) $n = 50$ (d) $n = 100$.

3.4 Considerações finais

Os três estimadores ($\hat{\rho}$, r_{1+} e C) de ρ foram comparados e observou-se que os três apresentaram vies. Os estimadores r_{1+} e C apresentaram menores vieses, principalmente com pequenas amostras, situação em que o viés é teoricamente maior. Esse fato é importante, pois pode influenciar os testes de hipóteses que se baseiam nesses estimadores. Observou-se que as principais mudanças no EQM desses estimadores eram causadas pelo viés, permanecendo constante a variabilidade dos mesmos quando o valor paramétrico de ρ e o valor de k mudavam para um mesmo n . Esse resultado para k é importante, pois não foi comentado por De Carlo e Tryon (1993) e vem mostrar que o número de variáveis regressoras pode afetar o poder e as taxas de erro tipo I.

A comparação das taxas de erro tipo I dos testes mostrou que 5 deles apresentaram tamanhos de testes iguais aos valores nominais de 5% e 1%. Esses testes foram o DWB, BC, tpB, Npc e NC. Assim, os melhores testes envolveram de alguma forma estimadores corrigidos para vies de ρ . A exceção foi o teste tpB, que envolveu uma estatística pivô t ,

mas baseada em um estimador que não possui correção para viés. Manly (1998) aponta que esse tipo de teste de *bootstrap* tem apresentado melhores resultados do que seus competidores.

A *performance* desses testes também foi avaliada sob ambientes anormais para o teste (distribuição uniforme e exponencial para o resíduo). Os resultados obtidos foram os mesmos no ambiente favorável que o da distribuição normal. Isso mostra que os teste são robustos à violação da normalidade dos resíduos. Outras distribuições devem ser contempladas em futuros trabalhos para que esses resultados possam ser generalizados de forma mais ampla.

O número de variáveis (k) influenciou o poder e as taxas de erro tipo I. Em geral, o aumento de k provocou reduções no poder e redução nas taxas de erro tipo I de alguns dos testes, exceto nos cinco testes considerados superiores.

Conclusões

O viés dos estimadores $\hat{\rho}$, r_{1+} e C aumenta com o aumento de k e de ρ . A variância dos estimadores $\hat{\rho}$, r_{1+} e C não é afetada pelo aumento de k e de ρ . Os testes DW, $N\rho$, $t\rho$, $B\rho$ e $B\rho Ca$ são rigorosos e menos poderosos do que seus competidores. O teste DW é o mais rigoroso de todos e apresentou menor poder. Os testes DWB e BC são equivalentes. Os testes DWB, BC, $t\rho B$, $N\rho c$ e NC são considerados os melhores por apresentarem tamanho de teste idênticos ao valor nominal, maior poder e por serem robustos. O teste NC foi considerado rigoroso para pequenos valores de α e de n . Além das propriedades destacadas na conclusão (f), o teste $N\rho c$ é recomendado neste trabalho devido à simplicidade e facilidade para ser aplicado.

MIRANDA, V. F. L.; FERREIRA, D. F. Monte Carlo evaluation of asymptotic and *bootstrap* tests for residual autocorrelation. *Rev. Mat. Estat.*, São Paulo, v.24, n.1, p.29-52, 2006.

- **ABSTRACT:** *Autocorrelation diagnosis can be formally made by using Durbin-Watson's test (DW). This study aimed to evaluate through Monte Carlo simulation, the Durbin-Watson's bootstrap approach (DWB), direct bootstrap test for autocorrelation parameter (ρ) with ($B\rho$) and without accelerated bias ($B\rho Ca$) correction, t test ($t\rho$), bootstrap t test ($t\rho B$), normal ρ test ($N\rho$), normal ρ with bias correction ($N\rho c$), normal test for Young's C statistics (NC) and bootstrap C test approach (BC). Additionally, the quality of the three estimators ($\hat{\rho}$, r_{1+} and C) of ρ was studied by evaluating bias and the mean square error. An autocorrelation structure of first order was simulated. As an evaluation criterion type-I error rate and the power of these tests were compared with the test of DW. The main conclusions are: the bias of the $\hat{\rho}$, r_{1+} and C estimators increase with the increase of the number of k covariates and ρ autocorrelation parameter; the variance of the three estimators are not affected by increasing k and ρ ; DW, $N\rho$, $t\rho$, $B\rho$ and $B\rho Ca$ tests are rigorous and less powerful than their competitors; the DW test is the most rigorous of all and presented the smallest power; DWB and BC tests are equivalent; DWB, BC, $t\rho B$, $N\rho c$ and NC tests are considered the best because they have sizes that are not significantly different from the nominal level (α), larger power and because they are robust; NC test was considered rigorous for small values of α and of sample size n ; $N\rho c$ test is recommended here due to its simplicity and ease of application.*
- **KEYWORDS:** *Autocorrelation; Durbin-Watson's test; bootstrap; Monte Carlo.*

Referências

- BUSSAB, W. O.; MORETTIN, P. A. *Estatística básica* 5. ed. São Paulo: Atual, 2003. 526 p.
- DE CARLO, L. T.; TRYON, W. W. Estimating and testing autocorrelation with small samples: A comparison of the C-statistic to a modified estimator. *Behav. Ther.*, Oxford, v.31, n.8, p.781-788, 1993.
- DRAPER, R. N.; SMITH, H. *Applied regression analysis*. 3. ed. New York: John Wiley, 1998. 706 p.
- DURBIN, J.; WATSON G. S. Testing for serial correlation in least squares regression I. *Biometrika*, London, v.37, n.3/4, p.409-428, 1950.
- DURBIN, J.; WATSON G. S. Testing for serial correlation in least squares regression II. *Biometrika*, London, v.38, n.1/2, p.159-178, 1951.
- EFRON, B.; TIBSHIRANI, R. *An introduction to the bootstrap*. New York: Chapman & Hall, 1993. 436 p.
- HOFFMANN, R.; VIERA, S. *Uma introdução à econometria*. 2. ed. São Paulo: Hucitec, 1987. 379 p.
- HUITEMA, B. E.; MCKEAN, J. W. Autocorrelation estimation and inference with small samples. *Psychol. Bull.*, Washington, v.110, n.2, p.291-304, 1991.
- JEONG, J.; CHUNG, S. Bootstrap tests for autocorrelation. *Comput. Stat. Anal.*, Amsterdam, v.38, p.49-69, 2001.
- MANLY, B. F. J. *Randomization, bootstrap and Monte Carlo methods in biology*. 2. ed. London: Chapman-Hall, 1998. 399 p.
- MOOD, A. M.; GRAYBILL, F. A.; BOES D. C. *Introduction to the theory of statistics*. 3. ed. New York: McGraw Hill, 1974, 842p.
- MORAN, P. A. P. Some theorems on time series II. The significance of the series correlation coefficient. *Biometrika*, London, v.35, n.3/4, p.255-260, 1948.
- NETER, J.; WASSERMAN, W.; KUTNER, H. M. *Applied linear statistical models: regression, analysis of variance, and experimental designs*, 2. ed. Illinois: Richard D. Irwin, 1974. 1184p.
- YOUNG, L. C. On randomness in ordered sequences. *Ann. Math. Stat.*, Baltimore, v.12, p.296-300, 1941.

Recebido em 15.06.2004.

Aprovado após revisão em 08.02.2006.