

5. ANÁLISE DE RESÍDUOS

No Capítulo 4 foram propostas metodologias para estimar o volume de óleo recuperável. Porém, após investigar o modelo que melhor se ajusta aos dados históricos, deve-se analisar ainda o resíduo, que é a diferença entre os dados observados e os valores ajustados pelo modelo (calculado).

O objetivo principal deste capítulo é fornecer uma metodologia para verificar se os modelos propostos no Capítulo 4 capturam toda a estrutura de dependência na variável a ser modelada, ou seja, se os modelos capturam toda a informação contida nos dados.

Se o modelo for adequado, espera-se que os resíduos não apresentem nenhum padrão bem definido, i.e., possam ser classificados como ruído branco. Caso contrário, se existir estrutura de dependência no resíduo e este puder ser modelado, não podemos classificá-lo como ruído branco e, conseqüentemente, o modelo deverá ser modificado para incluir esta explicação adicional.

Para classificar o resíduo como sendo ruído branco, deve-se investigar se os resíduos são independentes e identicamente distribuídos (i.i.d.) de média zero. Para tanto, utiliza-se o Teste BDS (Brock *et al.*, 1987) que tem como hipótese nula (H_0) a classificação do resíduo como sendo i.i.d. e hipótese alternativa (H_1) a presença de dependência linear ou não linear nos resíduos.

Em Estatística, define-se hipótese nula como sendo uma hipótese que é presumida verdadeira até que provas estatísticas sob a forma de testes de hipóteses indiquem o contrário. Por outro lado, a hipótese alternativa representa a negação da hipótese nula.

A necessidade de caracterizar dependência não linear em séries temporais estimulou o desenvolvimento do Teste BDS que levou o nome dos pesquisadores que o criaram: William Brock, Davis Dechert e José Alexandre Sheinkman.

Segundo Gazola & Caratori (2003), o BDS passou a ser amplamente utilizado em diversas áreas e tornou-se o teste mais conhecido para detectar estruturas não lineares presentes em uma série temporal.

5.1. A INTEGRAL DE CORRELAÇÃO

O Teste BDS utiliza o conceito da correlação espacial dos termos da série dentro de um espaço de dimensão “m”. Baseia-se numa integral de correlação definida pela expressão:

$$C_{m,T}(\varepsilon) = \sum_{t < s} I_{\varepsilon}(x_t^m, x_s^m) \cdot \left[\frac{2}{T_m(T_m - 1)} \right] \quad (1)$$

Onde:

T é o tamanho da amostra;

$T_m = T - m + 1$ representa o número de vetores x_t^m ;

$$x_t^m = (x_t, x_{t+1}, \dots, x_{t+m-1}) ;$$

$$I_{\varepsilon}(x_t^m, x_s^m) = 1, \text{ se } \|x_t^m - x_s^m\| < \varepsilon ;$$

$$= 0, \text{ caso contrário.}$$

ε = distância arbitrária;

t e s são instantes de tempo com $s = t + 1$.

Para uma melhor compreensão, considere o exemplo proposto por Gazola & Caratori (2003), cuja série histórica está apresentada na Tabela 2.

Tabela 1 – Série histórica proposta composta por 10 observações.

t=1	-0,617527
t=2	1,492318
t=3	-1,931585
t=4	-0,782284
t=5	0,461741
t=6	-0,244256
t=7	0,454955
t=8	-1,970666
t=9	-2,166198
t=10	0,385075

Fonte: Gazola & Caratori (2003).

Da tabela, obtém-se o tamanho da amostra $T = 10$. Considerando um espaço de dimensão $m=3$, calcula-se o valor $T_m = (T - m + 1) \Rightarrow T_m = (10 - 3 + 1) \Rightarrow T_m = 8$ vetores $x_t^m = (x_t, x_{t+1}, \dots, x_{t+m-1})$. Por exemplo, para $t=1$ e $s=2$, temos:

$$x_t^m = x_1^3 = (x_1, x_2, x_{1+3-1}) \Rightarrow x_1^3 = (x_1, x_2, x_3)$$

$$x_s^m = x_2^3 = (x_2, x_3, x_{2+3-1}) \Rightarrow x_2^3 = (x_2, x_3, x_4)$$

Analogamente,

$$x_t^m = x_1^3 = (x_1, x_2, x_3) \quad \Rightarrow \quad x_s^m = x_2^3 = (x_2, x_3, x_4)$$

$$x_t^m = x_2^3 = (x_2, x_3, x_4) \quad \Rightarrow \quad x_s^m = x_3^3 = (x_3, x_4, x_5)$$

$$x_t^m = x_3^3 = (x_3, x_4, x_5) \quad \Rightarrow \quad x_s^m = x_4^3 = (x_4, x_5, x_6)$$

⋮

$$x_t^m = x_7^3 = (x_7, x_8, x_9) \quad \Rightarrow \quad x_s^m = x_8^3 = (x_8, x_9, x_{10})$$

$$x_{T_m}^m = x_8^3 = (x_8, x_9, x_{10}) \quad \Rightarrow \quad \text{---}$$

Calcula-se $I_{ts} = \|x_t^m - x_s^m\| = \|x_t - x_s, x_{t+1} - x_{s+1}, \dots, x_{t+m-1} - x_{s+m-1}\|$. Para $t=1$ e $s=2$, temos:

$$I_{12} = \|x_1^3 - x_2^3\| = \|x_1 - x_2, x_2 - x_3, x_3 - x_4\| = \|-2,1098; 3,4239; -1,1493\|$$

Repete-se este procedimento para todos os pares I_{ts} (de I_{12} a I_{78}). A Tabela 3 mostra o cálculo realizado para alguns destes pares.

Tabela 2 – Cálculo dos pares I_{ts} .

$I_{t,s}$	$I_{1,3}$	$I_{1,4}$...	$I_{1,8}$	$I_{2,3}$...	$I_{2,8}$	$I_{3,4}$...	$I_{7,8}$
$x_t - x_s$	1,3140	0,1647	...	1,3531	3,4239	...	3,4629	-1,1493	...	2,4256
$x_{t+1} - x_{s+1}$	2,2746	1,0305	...	3,6585	-1,1493	...	0,2346	-1,2440	...	0,1955
$x_{t+m-1} - x_{s+m-1}$	-2,3933	-1,6873	-2,3166	-1,2440	-1,1673	0,7059	-2,5512

Fonte: o autor.

Em seguida, calcula-se a norma de cada um dos vetores I_{ts} . Por exemplo, para a norma I_{12} , temos:

$$I_{12} = \sqrt[3]{(-2,10985)^2 + (3,42390)^2 + (-1,14930)^2} = 4,182756$$

Analogamente, a Tabela 4 mostra o cálculo realizado da norma de alguns pares I_{ts} .

Tabela 3 – Cálculo das normas dos pares I_{ts} .

$I_{1,3}$	$I_{1,4}$...	$I_{1,8}$	$I_{2,3}$...	$I_{2,8}$	$I_{3,4}$...	$I_{7,8}$
3,5536	1,9840	...	4,5368	3,8198	...	3,6619	1,8349	...	3,5257

Fonte: o autor.

O próximo passo refere-se à escolha da distância arbitrária de tamanho “ ε ”. Esta distância deve atender à seguinte condição: $0 < \varepsilon < \max x_t - \min x_t$. Caso a norma seja menor do que “ ε ”, então $I_\varepsilon(x_t^m, x_s^m) = 1$. Caso contrário, $I_\varepsilon(x_t^m, x_s^m) = 0$.

Finalmente, calcula-se a integral de correlação definida anteriormente (Equação 72). Para a série dada, o valor da integral de correlação é 0,29. Este valor representa a razão entre o número de pares de pontos onde a distância entre eles é menor do que “ ε ” e o número de pares total.

Quanto maior o valor da integral de correlação $C_{m,T}(\varepsilon)$, maior a estrutura de dependência na série. Processos com estrutura de dependência geram séries cujos termos encontram-se mais espacialmente correlacionados do que em séries geradas por processos aleatórios, como o ruído branco.

Se o valor de “ ε ” é escolhido tal que todos os pares atendam à condição acima, então $C_{m,T}(\varepsilon) = 1$. Da mesma forma, se “ ε ” for escolhido tal que nenhum par atenda a condição, então $C_{m,T}(\varepsilon) = 0$.

Além disso, se x_t^m e x_s^m estiverem muito “próximos”, então a integral de correlação assumirá um valor próximo a 1. Caso contrário, se x_t^m e x_s^m estiverem “longe”, então a integral de correlação assumirá um valor próximo de zero.

5.2. A ESTATÍSTICA BDS

Considere a probabilidade de qualquer par de observações (x_i, x_j) diste ao máximo de “ ε ” um do outro. Então,

$$P_1 \equiv P\left(\|x_i - x_j\| \leq \varepsilon\right) \quad \text{para todo inteiro } i \neq j \quad (2)$$

Definindo a probabilidade “ P_2 ” como sendo a probabilidade de duas observações estarem próximas uma da outra, assim como pela probabilidade de seus predecessores estarem próximos um do outro, temos:

$$P_2 \equiv P\left(\|x_i - x_j\| \leq \varepsilon, \|x_{i-1} - x_{j-1}\| \leq \varepsilon\right) \quad \text{para todo inteiro } i \neq j \quad (3)$$

Quando a série x_t for i.i.d, $P_1 = P\left(\|x_i - x_j\| \leq \varepsilon\right) = P\left(\|x_{i-1} - x_{j-1}\| \leq \varepsilon\right)$, então:

$$P_2 = P_1^2 \quad (\text{se } x_t = \text{i.i.d}) \quad (4)$$

A partir deste princípio, o teste BDS sobre uma dimensão “ m ” tem como hipótese nula que as probabilidades acima são iguais. Logo, temos:

$$H_0 : P_m = P_1^m \quad H_1 : P_m \neq P_1^m \quad (5)$$

que equivale a testar:

$$H_0 : x_t = \text{i.i.d.} \quad (6)$$

A probabilidade P_m é estimada pela integral de correlação $C_{m,T}(\varepsilon)$. Conforme visto anteriormente, este valor representa a fração dos pares de pontos (x_t^m, x_s^m) que distam no máximo “ ε ” um do outro, isto é, $\|x_t^m, x_s^m\| < \varepsilon$.

Brock *et al.* (1987) demonstraram, que sob a hipótese nula das observações i.i.d., $C_m(\varepsilon) = [C_1(\varepsilon)]^m$ e que $C_{m,T}(\varepsilon) - [C_{1,T}(\varepsilon)]^m$ possui uma distribuição assintótica normal com média 0 e variância:

$$\sigma_{m,T}^2(\varepsilon) = 4 \left[K^m + 2 \sum_{j=1}^{m-1} K^{m-j} C^{2j} + (m-1)^2 C^{2m} - m^2 K C^{2m-2} \right] \quad (7)$$

Onde:

$C = C_{1,T}(\varepsilon) \Rightarrow C$ é estimado pela integral de correlação de tamanho 1.

$$K_T(\varepsilon) = \frac{2}{T_m(T_m - 1)(T_m - 2)} \sum_{t < s < r} \{I_\varepsilon(X_t, X_s)I_\varepsilon(X_s, X_r) + I_\varepsilon(X_t, X_r)I_\varepsilon(X_r, X_s) + I_\varepsilon(X_s, X_t)I_\varepsilon(X_t, X_r)\} \quad (8)$$

Portanto, sob a hipótese nula (i.i.d.), a estatística do teste BDS tem uma distribuição assintótica normal definida como sendo:

$$BDS_{m,T}(\varepsilon) = \frac{\sqrt{T} [C_{m,T}(\varepsilon) - (C_{1,T}(\varepsilon))^m]}{\sigma_{m,T}(\varepsilon)} \quad (9)$$

Caso $C_{m,T}(\varepsilon) - [C_{1,T}(\varepsilon)]^m$ seja significativo, implica num valor “grande” da estatística $BDS_{m,T}(\varepsilon)$, caindo na região crítica (RC) da distribuição normal ilustrada na Figura 28, na qual rejeita-se H_0 . Por outro lado, para $C_{m,T}(\varepsilon) - [C_{1,T}(\varepsilon)]^m$ pouco significativo, a estatística $BDS_{m,T}(\varepsilon)$ é “pequena”, caindo na região de aceitação (RA), logo não há evidência para rejeitar H_0 .

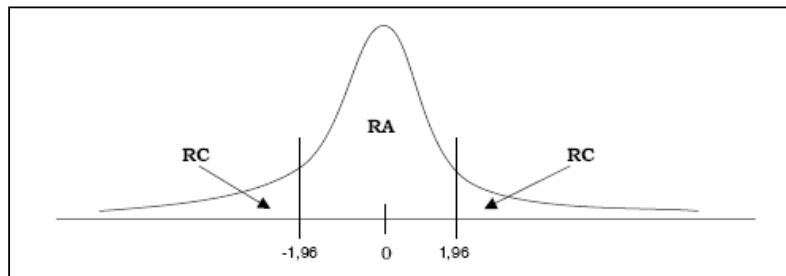


Figura 1 – Região de aceitação e crítica do Teste BDS para o intervalo de confiança 95%.

Fonte: Gazola & Caratori (2003).

Para mostrar os limites da região de aceitação (RA), utilizou-se na Figura 28 um intervalo de confiança de 95,0% ($Z = \pm 1,96$). Assim sendo, se os valores da estatística $BDS_{m,T}(\varepsilon)$ encontram-se no intervalo $[-1,96; 1,96]$, para diferentes valores de “m” e “ε”, então o teste aceita a hipótese nula (H_0) com um nível de significância $\alpha = 5\%$.

Em Estatística, um resultado é significativo se for improvável que tenha ocorrido por acaso. A significância de um teste, também chamada de *p-value*, é a probabilidade máxima de rejeitar acidentalmente uma hipótese verdadeira (uma decisão conhecida como erro de tipo I).

É relevante observar que um resultado “significante ao nível de 1%” é mais significativo do que um resultado ao nível de 5%. No entanto, um teste ao nível de 1% é mais susceptível de padecer do erro de tipo II (falsamente aceitar uma hipótese inválida) do que um teste de 5%. O melhor resultado que se pode obter é um equilíbrio entre significância e poder, em outras palavras, entre os erros de tipo I e tipo II.

5.3. ESCOLHA DE PARÂMETROS

Nota-se pela Equação 80 que a estatística BDS é uma função de dois parâmetros a serem arbitrados: “m” e “ε”. Segundo Gazola & Caratori (2003), não existe qualquer teoria totalmente satisfatória para a escolha destes parâmetros.

Brock *et. al.* (1991) realizaram estudos empíricos através da simulação de Monte Carlo, buscando estabelecer um conjunto de valores viáveis para as escolhas dos parâmetros “m” e “ε”. Escolhe-se o parâmetro “ε” em unidades do desvio padrão da série, usualmente entre $0,5\sigma$ e $2,0\sigma$. Em relação à dimensão “m”, Brock *et. al.* (1991) recomendam usar “m” entre 2 e 10. Para séries com 200 a 500 observações, “m” deve ser escolhido entre 2 e 5.

Entretanto, segundo Brock *et. al.* (1991), a Equação 80 é somente adequada para valores $T/m > 200$. Em amostras pequenas, a estatística BDS pode apresentar uma distribuição diferente da distribuição normal. Como o presente estudo trata de séries anuais, dificilmente encontraremos um valor $T/m > 200$.

Neste caso, Brock *et. al.* (1991) propõem a utilização da técnica *bootstrap*¹ para o cálculo da estatística BDS. O *software* utilizado no Capítulo 6 para a execução do Teste BDS oferece uma opção de cálculo dos *p-values* através da técnica *bootstrap*.

Quando a técnica *bootstrap* é solicitada, realiza-se uma série de repetições onde, para cada repetição, um conjunto de observações é extraído aleatoriamente sem substituição dos dados originais.

¹Utilizou-se Bootstrap com 10000 repetições para os conjuntos de dados testados.