

ALGUNS TESTES PARA DETECTAR
MUDANÇAS EM MODELOS DE REGRESSÃO

ROSA GOEHLER PAIT

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ORIENTADOR:

PROF. DR. WILTON DE OLIVEIRA BUSSAB

- SÃO PAULO, JULHO DE 1979 -

Aos meus filhos

Felipe e

Heloisa

AGRADECIMENTOS

Queremos agradecer a todos os que direta ou indiretamente colaboraram para a realização deste trabalho, em particular

ao Professor Doutor Wilton de Oliveira Bussab, que sugeriu o tema desta dissertação e nos orientou de forma efetiva durante todo o trabalho.

ao Professor Doutor Adolpho Walter Pimazoni Canton pelas sugestões apresentadas.

ao Professor Victor Mirshawka, professor titular da Universidade Mackenzie, pela amizade e incentivo constante.

aos colegas da Universidade Mackenzie.

aos amigos do Instituto de Matemática e Estatística da Universidade de São Paulo.

Rosa Goehler Pait

ÍNDICE

	Pg.
1. CAPÍTULO 1	1
1.1 - Introdução.....	1
1.1.1 - Ponto de intersecção conhecido...	3
1.1.2 - Ponto de intersecção desconhecido	5
1.2 - A Proposta do Trabalho.....	7
2. CAPÍTULO 2	11
2.1 - Introdução	11
2.2 - Formulação do Problema	11
2.3 - O Teste Cusum de Page.....	13
2.4 - Exemplos	16
3. CAPÍTULO 3	23
3.1 - Introdução.....	23
3.2 - Descrição do Modelo	28
3.3 - Resíduos Recursivos e suas Propriedades..	29
3.4 - O Teste Cusum.....	30
3.5 - Exemplo.....	32
4. CAPÍTULO 4.....	35
4.1 - Introdução.....	35
4.2 - Modelo de Análise e Desenvolvimento do Mé- todo.....	35
4.3 - Exemplo.....	41
5. CAPÍTULO 5.....	42
5.1 - Introdução	42
5.2 - Determinação do Modelo e Desenvolvimento do Método.....	42
5.3 - Exemplo	45
5.4 - Conclusões	48

	Pg.
6. CAPÍTULO 6	50
6.1 - Introdução	50
6.2 - Determinação do Modelo.....	51
6.3 - Quatro tipos de União na Solução Global.	52
6.4 - Resolução do Problema.....	55
6.5 - Exemplos	61
7. CAPÍTULO 7	71
7.1 - Conclusões	71
A. APÊNDICE A	77
A.1 - Propriedades do W_r	77
A.2 - A Determinação do Par de Retas.....	80
B. APÊNDICE B	83
B.1 - Algumas Demonstrações.....	83
C. APÊNDICE C	85
C.1 - 1º Teorema	85
C.2 - Ajuste de duas Curvas Condicionadas a se Encontrar no Ponto $X = \zeta$	87
C.3 - 2º Teorema	90
C.4 - 3º Teorema	92
REFERÊNCIAS BIBLIOGRÁFICAS	93

CAPÍTULO 1

1.1 - INTRODUÇÃO

É muito comum no estudo do ajuste de um modelo aos dados, nos depararmos com uma situação em que nosso modelo é composto de 2 submodêlos, por exemplo, duas retas, que se cruzam num determinado ponto, havendo, uma mudança na estrutura do modelo.

Este problema de mudança de estrutura em regressão aparece em vários campos. É comum em problemas econômicos: um modelo de regressão para comportamento de mercado pode mostrar um salto quando a demanda excede a oferta. O ponto a partir do qual isto ocorre terá com certeza um significado econômico, e sua localização é muito importante.

Estudando-se as variações na demanda para um certo produto, poderemos detectar durante qual período a regressão que descreve o mercado foi estável, obtendo-se assim uma

base mais firme para previsões futuras.

Também em biologia o fenômeno é usual. No estudo do crescimento, assume-se comumente que existe uma relação log-linear entre o tamanho de duas partes do corpo e esta relação permanece, durante períodos estáveis de crescimento. Uma mudança estrutural nesta relação pode indicar que nova fase de crescimento se iniciou. Saber se durante o período estudado houve a passagem para uma nova fase do crescimento e quando isto ocorreu, é fundamental.

Algo análogo ocorre no estudo do metabolismo. Verificou-se que o metabolismo, para vários indivíduos e espécies, cresce log-linearmente com o tamanho do corpo, durante regimes estáveis e detectar uma mudança pode ser de interesse.

Comumente nos problemas práticos não sabemos se o nosso modelo sofre mudança de estrutura ou não, se houve ou não mudança de regime. E no caso de haver mudança, em geral não se sabe onde essa mudança ocorreu e precisamos estimar este ponto. Porém a título de introdução vamos estudar como é tratado um caso muito simples: o caso em que o fenômeno pode ser explicado por 2 retas e se sabe quando ocorreu a mudança, quer a abcissa exata deste ponto, quer a abcissa do 1º ponto que passa a ser explicado pela 2ª reta. O tratamento é o tradicional e é devido a NH Draper e H Smith (1966).

1.1.1 - Ponto de Intersecção Conhecido

Através de um exemplo irreal gerado sem erro, vamos verificar como ajustar 2 retas, com declives desconhecidos, a um fenômeno, quando o ponto de intersecção é conhecido. Suporemos que foram medidas as quantidades de produtos consumidos em uma determinada cidade, nos anos de 1968, 1969 ... 1976, obtendo-se as seguintes quantidades, y , a menos de uma constante.

$$y = 4, 8, 12, 16, 20, 22, 24, 26, 28$$

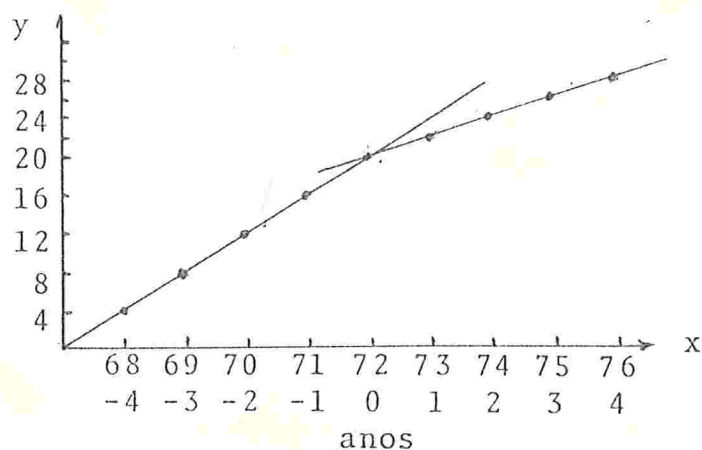


Figura 1.1

Pelo fig.1.1 vê-se claramente que 2 retas se ajustam ao fenômeno, sendo 1972 o ano da mudança.

Consideremos o modelo $y = \beta_0 y_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$

Podemos expressar este ajuste usando os dados como na matriz abaixo, onde x_1 e x_2 são 2 variáveis "fantasma".

Nº obs.	x_0	x_1	x_2	y		
1	1	-4	0	4		
2	1	-3	0	8		
3	1	-2	0	12		
4	1	-1	0	16		
5	1	0	0	20	=> ponto de intersecção	$x_1=x_2=0$
6	1	0	1	22		$y = 20$
7	1	0	2	24		
8	1	0	3	26		
9	1	0	4	28		

Sendo $\hat{\beta}_i$ os estimadores de mínimos quadrados de β_i , $i=0,1,2$ as equações normais para o modelo são:

$$\left. \begin{array}{l} 9\hat{\beta}_0 + (-10)\hat{\beta}_1 + 10\hat{\beta}_2 = 160 \\ -10\hat{\beta}_0 + 30\hat{\beta}_1 = -80 \\ 10\hat{\beta}_0 + 30\hat{\beta}_2 = 260 \end{array} \right\} \Rightarrow \begin{array}{l} \hat{\beta}_0 = 20 \\ \hat{\beta}_1 = 4 \\ \hat{\beta}_2 = 2 \end{array}$$

O modelo é exato, visto que foi criado sem erro. O estimador $\hat{\beta}_0$ é o valor de \hat{y} em $x_1 = x_2 = 0$, isto é, \hat{y} é o ponto da intersecção das retas. $\hat{\beta}_1$ é o declive da 1.^a reta e $\hat{\beta}_2$ é o declive da 2.^a reta.

1.1.2 - Ponto de Intersecção Desconhecido

Vejamos agora o caso em que o ponto de intersecção e os declives são desconhecidos. Somente se sabe até que ponto ajustamos o 1º modelo, isto é, sabemos a partir de que ponto o modelo muda mas desconhece-se a sua verdadeira localização. Novamente construiremos um exemplo irreal, exato. Para os anos de 1968, 1969...1976 o consumo, y , de um dado produto, foi, respectivamente

4, 8, 12, 16, 19, 21, 23, 25, 27

Consideremos o modelo

$$y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

A nossa matriz deverá ter uma 3ª variável "fantasma", β_3 , a fim de considerar o salto que aparece na 5ª observação. β_3 representa a distância vertical entre as 2 retas no ponto $x = 72$, distância em que a 2ª excede, está acima, da 1ª reta. Caso β_3 seja negativo, significará que a 2ª reta de regressão localiza-se abaixo da 1ª. A figura 2 mostra bem a situação descrita,

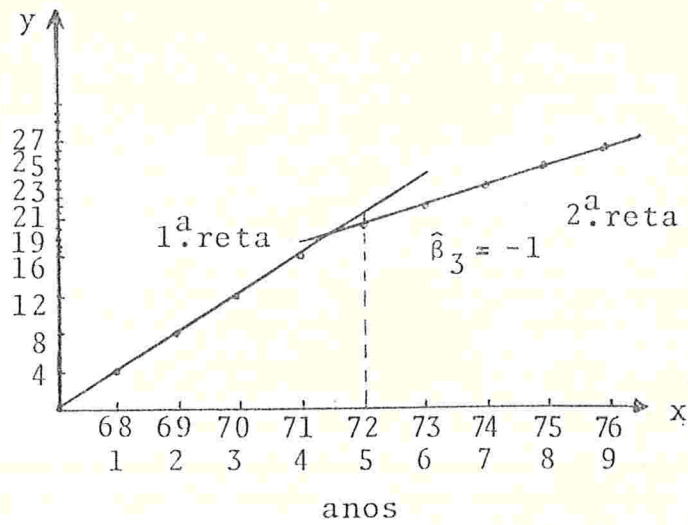


Figura 1.2

A matriz será então

Nº obs	x_0	x_1	x_2	x_3	y
1	1	1	0	0	4
2	1	2	0	0	8
3	1	3	0	0	12
4	1	4	0	0	16
5	1	5	0	1	19
6	1	5	1	1	21
7	1	5	2	1	23
8	1	5	3	1	25
9	1	5	4	1	27

Note que há uma pequena diferença nas colunas x_1 e x_2 em relação ao exemplo do ítem 1.1.1, porém o exemplo do

Ítem 1.1.1 poderia ser escrito de maneira análoga a este .

Aqui β_0 é o intercepto da 1.^a reta.

As equações normais serão

$$\left. \begin{array}{l} 9\hat{\beta}_0 + 35\hat{\beta}_1 + 10\hat{\beta}_2 + 5\hat{\beta}_3 = 155 \\ 35\hat{\beta}_0 + 155\hat{\beta}_1 + 50\hat{\beta}_2 + 25\hat{\beta}_3 = 695 \\ 10\hat{\beta}_0 + 50\hat{\beta}_1 + 30\hat{\beta}_2 + 10\hat{\beta}_3 = 250 \\ 5\hat{\beta}_0 + 25\hat{\beta}_1 + 10\hat{\beta}_2 + 5\hat{\beta}_3 = 115 \end{array} \right\} \begin{array}{l} \hat{\beta}_0=0 \\ \hat{\beta}_1=4 \rightarrow \text{declive de 1.ª reta} \\ \hat{\beta}_2=2 \rightarrow \text{declive de 2.ª reta} \\ \hat{\beta}_3=-1 \end{array}$$

O sinal negativo de $\hat{\beta}_3$ e o fato de que $\hat{\beta}_1 > \hat{\beta}_2$, indicam que o ponto de intersecção das 2 linhas está à esquerda do 5º ponto.

Isto ocorre realmente quando $x_1 = 4,5$.

1.2 - A PROPOSTA DO TRABALHO

O caso descrito nestas linhas preliminares tem como objetivo situar melhor o problema. Porém, como já dissemos, geralmente, na prática, não só não sabemos onde houve mudança de estrutura, no modelo, como nem sequer sabemos se houve mudança ou não.

Nossa proposta é estudar o ajuste de um modelo aos dados quando temos razão para supor que em determinado momento o modelo muda de estrutura. Faz-se necessário então, antes de mais nada, testar se o modelo permanece estável ou

se houve mudança de estrutura, e, caso seja aceita a hipótese de que houve mudança, estimar a localização do ponto de mudança.

Não trataremos da importante questão de escolher entre vários modelos de regressão. Assumiremos que nosso modelo, com suas possíveis mudanças estruturais, faz sentido, mesmo que possa existir outro modelo, isto é, uma regressão de outro tipo que também se ajustaria aos dados.

Para maior facilidade de leitura, colocamos no apêndice muitas demonstrações, deixando nosso apêndice talvez muito volumoso, porém despojando a 1.^a leitura de quebras e interrupções, tornando-a assim no nosso entender, mais clara e didática.

No intuito de alcançar nosso objetivo, nos propusemos mostrar estudos de vários autores sobre o problema da mudança de estrutura.

Primeiramente veremos o enfoque de 2 autores E.S. Page(1954) e Brown et al. (1974), para o problema de testar a estabilidade do modelo, cada enfoque tendo sua indicação específica. A seguir desenvolveremos um teste localmente mais poderoso para testar a mudança.

Caso uma mudança seja detectada, cumpre-nos estimar o ponto onde ocorrerá mudança. Para tal estudaremos a proposta de 2 outros autores, R.E.Quant(1958) D.J.Hudson(1960).

Ainda teceremos comentários sobre alguns recentes artigos sobre o assunto, alguns enfocando a distribuição assintótica do estimador do ponto de mudança.

Detalhando mais, vamos expor a proposta de cada capítulo:

Capítulo 2: Considerando uma amostra aleatória de observações independentes, na ordem em que são obtidas, forneceremos um teste, devido a E.S. Page (1954), que nos permite verificar se as observações provêm da mesma função distribuição, $F(x/\theta)$, contra a alternativa de que x_1, \dots, x_{t_0} vêm de $F(x/\theta)$ e x_{t_0+1}, \dots, x_T vêm de $F(x/\theta')$, $\theta \neq \theta'$, com t_0 desconhecido. O teste é desenvolvido para o caso em que θ é conhecido e θ' não, fato comum na prática. Introduziremos aqui o conceito de SOMAS CUMULATIVAS (CUSUMS), técnica retomada por vários autores.

Capítulo 3: Estudaremos a estabilidade no tempo em relações de regressão, de acordo com a proposta de Brown, Durbin e Evans (1975) que, para realizarem o teste usaram também uma técnica baseada em SOMAS CUMULATIVAS (CUSUMS) e Resíduos Recursivos.

Capítulo 4: Será aqui desenvolvido um teste localmente mais poderoso para testar a hipótese de que um coeficiente de declive, em um modelo linear, numa série temporal, é estável, contra a alternativa de que o declive muda exata

mente uma vez em algum lugar na série. A proposta se deve a John Farley e Melvin J. Hinich (1970).

Capítulo 5: Uma maneira de estimar os parâmetros de um sistema de regressões lineares obedecendo a 2 regimes separados é aqui apresentada. Como apesar de se supor a existência de 2 regimes, isto é, uma mudança de estrutura, não se sabe a abcissa do ponto de mudança de um submodelo para outro, será indicado um método para estimá-la: determinar o estimador de máxima verossimilhança, baseado no exame direto da função de verossimilhança. O assunto foi tratado por Richard E. Quant (1958).

Capítulo 6: Estudaremos a solução proposta por Derek J. Hudson (1966) para a situação em que a curva completa consiste de 2 ou mais submodêlos unido-se em pontos cujas abcissas devem ser estimadas, apresentando-se a solução de mínimos quadrados global.

Capítulo 7: Neste capítulo, alguns trabalhos recentes que desenvolveram mais aprofundadamente as técnicas aqui expostas são apresentados.

CAPÍTULO 2

2.1 - INTRODUÇÃO

Desenvolvemos neste capítulo um teste para detectar se houve ou não mudança num parâmetro de uma distribuição . Será o 1º passo para atingir o objetivo fundamental deste trabalho que é estudar a mudança em funções de regressão. Entretanto o capítulo aqui desenvolvido é de fundamental importância como introdutor de técnicas denominadas SOMAS CUMULATIVAS (CUSUMS = CUMULATIVES SUMS), técnicas estas posteriormente retomadas por vários autores no estudo do nosso problema e amplamente recomendados. O teste deve-se a Page (1954).

2.2 - FORMULAÇÃO DO PROBLEMA

Consideremos uma amostra de observações independentes, na ordem em que são obtidas.

Queremos testar:

$$\left\{ \begin{array}{l} H_0: \text{Todas as observações são retiradas de uma mesma população com função de distribuição } F(x/\theta), \theta \text{ conhecido.} \\ H_a: x_1, \dots, x_{t_0} \text{ vêm de } F(x/\theta); \\ \quad x_{t_0+1}, \dots, x_T \text{ vêm de } F(x/\theta') \quad \theta \neq \theta' \end{array} \right.$$

θ , o parâmetro em questão, aqui será a média.

Se t_0 fosse conhecido seria em problema direto de teste de duas médias. Estudaremos o caso de t_0 desconhecido. O teste é proposto para o caso do parâmetro θ ser conhecido. Como já foi dito suporemos que o parâmetro sob consideração é a média.

O esquema proposto por Page consiste em tomar somas cumulativas (CUSUMS), S_r , com

$$S_r = \sum_{j=1}^r (x_j - \theta) \text{ e com } S_0 = 0 \quad 0 \leq r \leq T$$

Diremos que houve uma mudança no parâmetro quando

$$S_r - \min_{0 \leq i \leq r} S_i \geq h$$

isto é, quando o caminho amostral alcança uma altura h acima de seu valor mínimo anterior, o que é intuitivo.

O teste de significância é baseado na estatística

$$m = \max_{0 \leq r \leq T} \{S_r - \min_{0 \leq i \leq r} S_i\}$$

Rejeita-se H_0 se $m \geq h$, sendo h um valor crítico.

O presente teste é uma forma truncado do esquema de inspeção, sendo então suas propriedades difíceis de calcular.

Um caso particular, que não oferece tanta dificuldade, é aquêles em que as observações são variáveis binomiais 0 ou 1, e que veremos a seguir.

2.3 - O TESTE CUSUM DE PAGE

Consideremos um teste para o caso geral usando variáveis binomiais. Para isto faz-se $y_i = a$, se $x_i - \theta \geq 0$ e $y_i = -b$, se $x_i - \theta < 0$ e escolhe-se a e b (> 0) tais que $E(y_i / \theta) = 0$ ($i=1 \dots T$)

Com a estatística

$$m = \max_{0 \leq r \leq T} \left\{ S_r - \min_{0 \leq i \leq r} S_i \right\} \text{ onde agora } S_r = \sum_{j=1}^r y_j; S_0 = 0$$

rejeita-se H_0 se $m \geq h$. Para simplificar ainda mais consideremos somente o caso onde a distribuição de x_i é simétrica e desta forma poderemos tomar $a=b=1$. Então y_i reduz-se a $y_i = \text{sinal}(x_i - \theta)$.

Vamos avaliar as propriedades do teste. Chamemos

$$m_r = S_r - \min_{0 \leq i \leq r} S_i \quad \text{e}$$

$P_{r,i}$ = prob de $m_r = i$ com todos os m anteriores inferiores a h , isto é,

$$P_{r,i} = P(m_r = i) \quad (i = 0, \dots, h-1) \quad \text{e} \quad m_s < h \\ \text{para} \quad 1 \leq s \leq r$$

A probabilidade de que H_0 seja rejeitada é

$$1 - \sum_{i=1}^{h-1} P_{T,i} \quad (2.1)$$

Façamos $P(y_i = 1) = p = 1-q$. Considerando o resultado da próxima observação, temos as relações

$$P_{r+1,0} = q(P_{r,0} + P_{r,1})$$

$$P_{r+1,i} = p \cdot P_{r,i-1} + qP_{r,i+1} \quad 1 \leq i < h-1$$

$$P_{r+1,h-1} = p \cdot P_{r,h-2}$$

Em notação matricial isto nos dá

$$P_{r+1} = \mathbb{P} \cdot p_r \quad \text{onde} \quad \mathbb{P} \text{ é a matriz}$$

$$\mathbb{P} = \begin{pmatrix} q & q & 0 & 0 & \dots & 0 & 0 \\ p & 0 & q & 0 & \dots & 0 & 0 \\ 0 & p & 0 & q & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & p & 0 \end{pmatrix}$$

Inicialmente

$$\begin{aligned} p_{0,0} &= 1 \\ p_{0,i} &= 0, \quad i \neq 0 \end{aligned} \quad p_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

Usamos implicitamente o fato de que m_r são variáveis numa cadeia de Markov, com h estados. Claro esta que

$$\boxed{p_r = P^r \cdot p_0} \quad (2.2)$$

Sob H_0 , $p = \frac{1}{2} = \text{cte.}$

O poder do teste depende tanto do valor de p após a mudança quanto da posição da mudança. Se a Prob. $(y_i = +1)$ é constantemente igual a p , de tal sorte que a mudança pode ser considerada como tendo ocorrido imediatamente, a probabilidade de que H_0 seja rejeitada em uma amostra de T observações é dada por (2.1) e (2.2) com $r = T$. Se a mudança ocorre após a k -ésima observação, o valor de P_T a ser usado em (2.1) é dado por $P_T = P_2^{T-k} P_1^k p_0$, onde P_1 e P_2 são obtidos da matriz P com $p = 1/2$ e $p = p$, respectivamente.

Estes conhecimentos nos permitem criar tabelas que relacionem valores de h e T (tamanho de amostra) para os quais as probabilidades de erro tipo I seja α , e nos permitem comparar o poder deste teste para um determinado p , com o poder de outros testes, mostrando suas vantagens.

A tabela 2.1, retirada do trabalho de PAGE(1954)

nos fornece valores de T e h para $\alpha = 0,05$ e $\alpha = 0,01$.

TABELA 2.1

Tabela de valores de T e h

$\alpha = 0,05$				$\alpha = 0,01$	
T	h	T	h	T	h
21	10	75	19	20	12
26	11	83	20	27	14
31	12	91	21	35	16
36	13	100	22	43	18
41	14	119	24	53	20
47	15	139	26	64	22
54	16	161	28	76	24
60	17	185	30	89	26
67	18			103	28
				118	30

2.4 - EXEMPLOS

2.4.1 - Exemplo 1

Mostraremos o procedimento que deve ser seguido a través de um exemplo empírico obtido pela geração de uma amostra aleatória normal. Foi construído uma amostra de 40 observações, as primeiras 20 tendo média 5 e variância 1, e as últimas 20, com média 6 e variância 1. Testaremos

$$\begin{cases} H_0 : \mu = 5 \text{ constantemente} \\ H_1 : \mu > 5 \text{ para } x_{t_0+1} \dots x_T \end{cases}$$

Na tabela 2.2 encontram-se os dados e as estatísticas necessárias para o cálculo do teste.

TABELA 2.2

Número da Observação	Valor de x_i	$y_i = \text{sign} \text{ de } (x_i - 5)$	S_r	$S_r - \min S_i$
1	3,95	-1	-1	0
2	5,96	1	0	1
3	6,22	1	1	2
4	5,58	1	2	3
5	4,02	-1	1	2
6	4,97	-1	0	1
7	3,46	-1	-1	0
8	4,29	-1	-2	0
9	4,65	-1	-3	0
10	5,66	1	-2	1
11	5,44	1	-1	2
12	5,91	1	0	3
13	4,98	-1	-1	2
14	3,58	-1	-2	1
15	5,26	1	-1	2
16	3,98	-1	-2	1
17	4,19	-1	-3	0
18	6,66	1	-2	1
19	6,05	1	-1	2
20	5,97	1	0	3
21	7,14	1	1	4
22	6,22	1	2	5
23	4,76	-1	1	4
24	6,60	1	2	5
25	5,72	1	3	6
26	4,88	-1	2	5
27	5,44	1	3	6
28	5,03	1	4	7
29	5,66	1	5	8
30	5,56	1	6	9
31	6,37	1	7	10
32	6,66	1	8	11
33	5,10	1	9	12
34	5,80	1	10	13
35	6,29	1	11	14
36	5,49	1	12	15
37	4,93	-1	11	14
38	6,18	1	12	15
39	8,29	1	13	16
40	6,84	1	14	17

O maior valor de $S_r - \min_{0 < i \leq r} S_i$, no exemplo, é 17. Verifica-se através da tabela (2.1), que para $T = 40$, 17 é o valor de h para aproximadamente 1% como valor de α . Podemos ainda tirar uma outra indicação deste teste. A posição do último zero de $S_r - \min_{0 < i \leq r} S_i$ é um estimador, viesado, da posição da mudança. Neste exemplo suspeitamos que a mudança ocorreu perto da 17ª observação. (Na realidade ocorreu após a 20ª).

2.4.2 - Exemplo 2

Vamos estudar agora o caso em que se quer testar se todas as observações podem ser expressas pelo modelo $y_i = \alpha + \beta x_i$.

Os dados agora foram gerados da seguinte maneira: Produziu-se uma amostra de 9 elementos dos quais os quatro primeiros foram gerados por $y = x$ e as 5 últimos por $y = 2,7 + 0,1x$.

Os erros acrescentados são v.a.i. normalmente distribuídas com média 0 e variância $(0,03)^2$. Estes valores encontram-se na tabela (2.3).

Queremos testar:

$$\begin{cases} H_0: \text{O modelo } y=x \text{ se ajusta aos dados todos.} \\ H_a: \text{Há uma mudança, após um ponto desconhecido, para um} \\ \text{ novo modelo que também se desconhece.} \end{cases}$$

TABELA 2.3

x	y	\hat{y} sob H_0	signal de $\hat{y}-y_i$
0	0,01	0	-
1	1,01	1	-
2	1,99	2	+
3	2,92	3	+
4	3,10	4	+
5	3,21	5	+
6	3,29	6	+
7	3,42	7	+
8	3,51	8	+

TABELA 2.4

Estatísticas necessárias para se efetuar o teste

x_i	signal de $\hat{y}-y_i$	S_r	$S_r - \min S_r$
0	-	-1	0
1	-	-2	0
2	+	-1	1
3	+	0	2
4	+	1	3
5	+	2	4
6	+	3	5
7	+	4	6
8	+	5	7

Mesmo nossa tabela (2.1) não nos fornecendo valores de T menores que 20, extrapolação grosseira nos daria para T=9 e h=7 e temos sinais evidentes de que se rejeita H_0 . Como o último zero em $S_r - \min S_i$ se encontra no local

correspondente à 2.^a observação, temos uma estimativa "viesada" do ponto de mudança, após a 2.^a observação.

2.5 - OBSERVAÇÕES

2.5.1 - Devemos tomar alguns cuidados ao testar ajuste de modelos por esta técnica. O teste proposto por Page é unilateral, isto é, testamos no ex. se a média é constantemente $= \theta$ ou se a partir de t_0+1 , passa a ser $\theta' > \theta$. Se quisermos testar

$$\begin{cases} H_0: \mu = \theta \text{ constantemente} \\ H_1: \mu < \theta \text{ para } x_{t_0+1} \dots x_T \end{cases}$$

isto é, se se suspeita desta mudança, o que se pode verificar através de um estudo gráfico prévio dos pontos (x, y) , deveríamos fazer nosso teste tomando o valor de $y_i = \text{sinal}(\theta - x_i)$.

Neste exemplo, através de um estudo gráfico, vê-se que o declive da 2.^a reta é menor, ficando ela abaixo da 1.^a reta. Daí termos tomado $y_i = \text{sinal}(\hat{y} - y_i)$. Se não tivéssemos feito nenhum estudo prévio, e tomássemos $y_i = \text{sinal}(y_i - \hat{y})$ teríamos o exemplo assim desenvolvido

TABELA 2.5

x_i	0	1	2	3	4	5	6	7	8
y_i	+	+	-	-	-	-	-	-	-
S_r	1	2	1	0	-1	-2	-3	-4	-5
$S_r - \min S_i$	1	2	1	0	0	0	0	0	0

com zeros de um ponto em diante, visto que os S_r finais são todos negativos. Isto ocorrendo deve imediatamente nos sugerir a retomada do problema com $y_i = \text{sinal}(\hat{y} - y_i)$ como feito anteriormente.

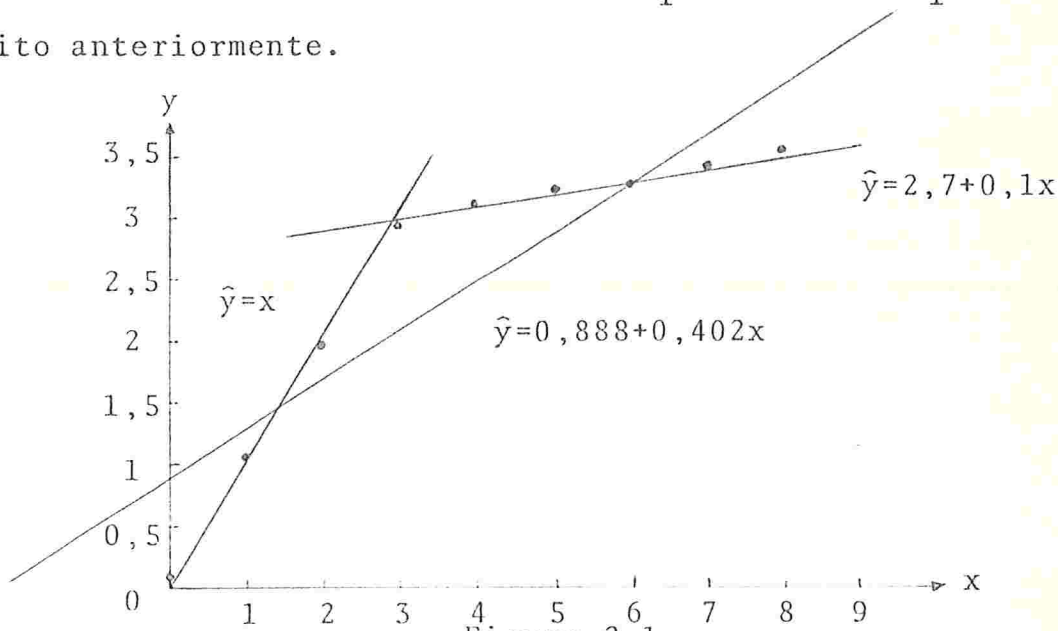


Figura 2.1

2.5.2 - A técnica de Page é usada quando o parâmetro θ , inicial é conhecido, no caso de teste para mudança de distribuição, ou, para quando se conhece o submodelo inicial para o caso de estudo da estabilidade de um modelo de regressão. Se não for este o caso, isto é, se para o exemplo que acabamos de estudar desconhecêssemos o submodelo $y=x$, inicial, talvez nos inclinássemos a proceder da seguinte forma.

Testaríamos a

$$\begin{cases} H_0 : E(y/x) = \alpha + \beta x & \text{constantemente} \\ H_1 : \text{Há uma mudança em algum ponto desconhecido,} \end{cases}$$

através da análise do resíduo fornecido pelo ajuste da reta de mínimos quadrados usando o conjunto de todos os dados. A reta assim ajustada, é $\hat{y} = 0,888 + 0,402x$, produzindo os números da tabela 2.6.

TABELA 2.6

x	\hat{y}	sinal $y_i - \hat{y}_i$	S_r	$S_r - \min S_i$
0	0,88756	-	-1	0
1	1,22956	-	-2	0
2	1,69160	+	-1	1
3	2,09550	+	0	2
4	2,49550	+	1	3
5	2,89755	+	2	4
6	3,29955	-	1	3
7	3,70155	-	0	2
8	4,10355	-	-1	1

Entretanto veremos que esta não é uma boa maneira de proceder visto que os efeitos residuais da mudança, que se espalham sobre todo o conjunto dos resíduos, vão influir de tal forma que seremos levados à decisão errada. Pela tabela (2.6) vemos que não se pode dizer que haja evidências de que se possa rejeitar H_0 , o que nos leva realmente à conclusão errada.

No próximo capítulo nos deteremos na explicação de uma maneira de resolver melhor o problema que surge quando desconhecemos o submodelo inicial.

CAPÍTULO 3

3.1 - INTRODUÇÃO

Neste capítulo usaremos a técnica apresentada no capítulo anterior (CUSUMS) para detetar mudanças em funções de regressão, para dados em série de tempo. Como vimos no capítulo 2, para se usar as técnicas de Page é necessário que se conheça o parâmetro θ , inicial, ou, em caso de ajuste de modelo, é necessário conhecer o 1º submodelo, que se ajusta à parte inicial dos dados. Com as técnicas desenvolvidas aqui, isto não é necessário.

O teste, que é devido a Brown et al (1974/5), baseia-se na soma cumulativa dos resíduos recursivos.

Iniciaremos com uma definição: Resíduos recursivos são resíduos padronizados da regressão para cada observação y_t , sendo os coeficientes da regressão calculados das $t-1$ observações anteriores, y_1, \dots, y_{t-1} para $t=k+1, \dots, T$, onde

k é o número de regressores e T o número de observações.

Sob a hipótese nula H_0 de que não houve mudança, veremos que os resíduos recursivos são não correlacionados, com média zero e variância constante, e se ainda mais as sumimos normalidade, serão independentes, o que levará a uma simplificação no nosso trabalho. Daí a definição de resíduos recursivos. Outros métodos de transformar resíduos de mínimos quadrados em variáveis $N(0, \sigma^2)$, independentes foram dados por vários autores, como Theil (1965, 1968) e Durbin (1970). Entretanto, os resíduos recursivos parecem melhores para detetar uma mudança no modelo durante o tempo pois, quando uma mudança ocorre, devido à sua própria definição, espera-se que sinais desta mudança sejam logo visíveis.

Já com outros métodos, como um tratamento com resíduos ordinários de mínimos quadrados, ou mesmo com CUSUMS destes resíduos, acontece comumente que os efeitos residuais da mudança se espalhem sobre todo o conjunto dos resíduos, impossibilitando mesmo detectar a mudança. Foi o que ocorreu no exemplo do capítulo 1, quando supusemos que o modelo inicial era desconhecido, e ajustando um modelo único a todos os dados, entramos com a técnica de Page.

O teste aqui apresentado deve ser visto mais como uma ferramenta, aliás poderosíssima, para detectar mu

danças, do que como um teste formal. Baseia-se fundamentalmente no gráfico dos CUSUMS dos resíduos recursivos como função do tempo.

3.2 - DESCRIÇÃO DO MODELO

O nosso modelo de regressão é

$$y_t = \underline{x}'_t \underline{\beta}_t + \epsilon_t \quad t=1, \dots, T$$

onde, no tempo t , y_t é a observação da variável dependente e \underline{x}_t é o vetor coluna das observações em k regressores. \underline{x}_t é suposto sem erro. O vetor dos parâmetros, $\underline{\beta}_t$, é escrito com o subscrito t para indicar que ele pode variar com o tempo.

$$\underline{x}_t = \begin{pmatrix} x_{at} \\ x_{bt} \\ \vdots \\ x_{kt} \end{pmatrix} \quad \underline{\beta}_t = \begin{pmatrix} \beta_{at} \\ \beta_{bt} \\ \vdots \\ \beta_{kt} \end{pmatrix} \quad \underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}$$

Vamos supor que os erros ϵ_t são v.a.i. $\sim N(0, \sigma_t^2)$, $t=1, \dots, T$. A hipótese nula será

$$H_0 = \begin{cases} \underline{\beta}_1 = \underline{\beta}_2 = \dots = \underline{\beta}_T = \underline{\beta} \\ \sigma_1^2 = \sigma_2^2 = \dots = \sigma_T^2 = \sigma^2 \end{cases}$$

No momento estaremos apenas preocupados com a primeira

parte da hipótese, isto é, estamos mais preocupados em detectar diferenças entre os $\underline{\beta}$'s do que entre os σ 's.

3.3 - RESÍDUOS RECURSIVOS E SUAS PROPRIEDADES

Sob H_0 , denotando por \underline{b}_r o estimador de mínimos quadrados de $\underline{\beta}$, baseado nas primeiras r observações, temos para \underline{b}_r a expressão

$$\underline{b}_r = (X_r' X_r)^{-1} X_r' \underline{y}_r \quad \text{onde} \quad \begin{cases} \underline{y}_r' = (y_1, \dots, y_r) \\ X_r' = (\underline{x}_1, \dots, \underline{x}_r) \end{cases}$$

isto é, X_r' é a matriz retangular composta dos vetores das observações nos sucessivos tempos 1, 2, ... r.

A matriz $X_r' X_r$ é suposta não singular. Assim, para $r = k+1, \dots, T$ temos os resíduos recursivos w_r dados por

$$w_r = \frac{y_r - \underline{x}_r' \underline{b}_{r-1}}{\sqrt{1 + \underline{x}_r' (X_{r-1}' X_{r-1})^{-1} \underline{x}_r}}$$

onde

$$X_{r-1}' = \left[\underline{x}_1, \dots, \underline{x}_{r-1} \right]$$

Vê-se que o resíduo recursivo w_r , é a previsão de erro padronizado de y_r , quando previsto através de $y_1 \dots y_{r-1}$.

Propriedades de w_r .

1) Sob H_0 , w_{k+1}, \dots, w_T são v.a.i. - $N(0, \sigma^2)$

$$2) (X'_r X_r)^{-1} = (X'_{r-1} X_{r-1})^{-1} - \frac{(X'_{r-1} X_{r-1})^{-1} \underline{x}_r \underline{x}'_r (X'_{r-1} X_{r-1})^{-1}}{1 + \underline{x}'_r (X'_{r-1} X_{r-1})^{-1} \underline{x}_r}$$

$$3) \underline{b}_r = \underline{b}_{r-1} + (X'_r X_r)^{-1} \underline{x}_r (y_r - \underline{x}'_r \underline{b}_{r-1})$$

$$4) \bar{S}_r = S_{r-1} + w_r^2, \quad r = k+1, \dots, T$$

onde S_r é a SQR depois de ajustar o modelo às primeiras r observações, assumindo H_0 verdadeiro, isto é,

$$S_r = (\underline{y}_r - X_r \underline{b}_r)' (\underline{y}_r - X_r \underline{b}_r)$$

As propriedades 2, 3 e 4 são relações recursivas para facilitar os cálculos. As provas se encontram no apêndice A-A.1.

3.4 - O TESTE "CUSUM"

Se $\underline{\beta}_t$ é constante até $t=t_0$ e difere deste valor a partir de $t > t_0$, os w_r terão média zero para r até $t=t_0$, mas, terão média não zero para $t > t_0$. Isto sugere o exame do gráfico dos CUSUMS

$$W_r = \frac{1}{\hat{\sigma}} \sum_{j=k+1}^r w_j \quad \text{para } r = k+1, \dots, T$$

com

$$\hat{\sigma}^2 = \frac{S_T}{T-k}$$

Queremos um teste para determinar quando o caminho amostral de W_r desvia-se de sua média, $E(W_r)=0$. Uma maneira seria encontrar um par de linhas simétricas acima e abaixo de $W_r=0$, de tal sorte que a probabilidade de cruzar uma linha (ou ambas as linhas) seja α , um nível de significância pré fixado.

Das propriedades dos w_r sob H_0 , a sequência W_{k+1}, \dots, W_r é uma sequência de variáveis aproximadamente normais tais que

$$E(W_r)=0, \text{ Var}(W_r)=r-k \quad \text{e} \quad C_{OV}(W_r, W_s)=\min(r,s)-k$$

como boa aproximação.

Para derivar o teste aproxima-se W_r pelo processo Gaussiano (z_t) com os parâmetros acima. Este é um processo Browniano partindo do zero no tempo $t=k$.

O par de retas simétricas será escrito na forma $y = \pm a \sqrt{T-k} \pm \frac{2a}{\sqrt{T-k}} (t-k)$ onde a é uma constante a ser determinada para que o nível de significância seja α .

Daremos abaixo uma tabela devida a Brown, Durbin

e Evans (1974) que relaciona a e α , cuja determinação baseou-se em resultados conhecidos sobre movimentos Brownianos. Vide apêndice A-A.2 onde a determinação das duas retas está explicada.

Assumiu-se, o que é justificável, que a probabilidade que W_T cruze ambas as linhas é negligenciável para valores de α normalmente usados

TABELA 3.1

$\alpha = 0,01$	$a = 1,143$
$\alpha = 0,05$	$a = 0,948$
$\alpha = 0,10$	$a = 0,850$

Procede-se da seguinte maneira: Fazemos o gráfico dos CUSUMS como função do tempo e verificamos se o caminho cruza uma das retas $y = \pm a\sqrt{T-k} \pm \frac{2a}{\sqrt{T-k}}(t-k)$. Se isto acontece rejeita-se H_0 .

3.5 - EXEMPLO

O próprio Durbin forneceu-nos um exemplo levado a efeito no Statistics and Business Research Department of the Post Office onde um modelo de regressão foi desenvolvido para explicar o crescimento no número de telefonemas locais (isto é, diferenças entre o número de chamadas em anos consecutivos) em termos de um modelo linear

envolvendo uma constante e 4 variáveis independentes, a saber:

- 1) uma medida da atividade econômica;
- 2) número de telefones residenciais;
- 3) preço de chamadas locais;
- 4) preço de telefones residenciais.

Dados de 1950/51 até 1971/72 foram medidos, suspeitando-se de um aumento de chamadas locais em 1964/65, sendo de importância fundamental para futuras previsões investigar a estabilidade no tempo.

Daremos os gráficos dos resíduos ordinários de mínimos quadrados Fig. 3.1, dos CUSUMS dos resíduos de mínimos quadrados Fig. 3.2 e dos CUSUMS dos resíduos recursivos Fig. 3.3. Estes gráficos vão mostrando evidências crescentes de instabilidade

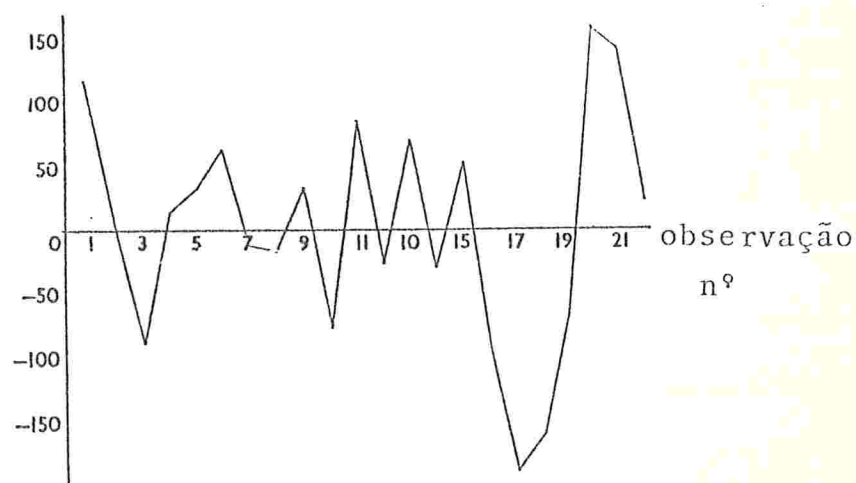


Figura 3.1 - Resíduos ordinários de mínimos quadrados

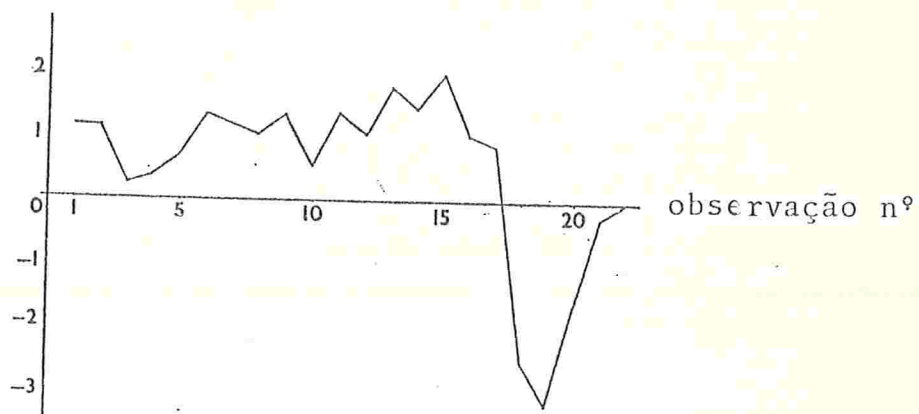


Figura 3.2 - CUSUMS dos resíduos ordinários de mínimos quadrados.

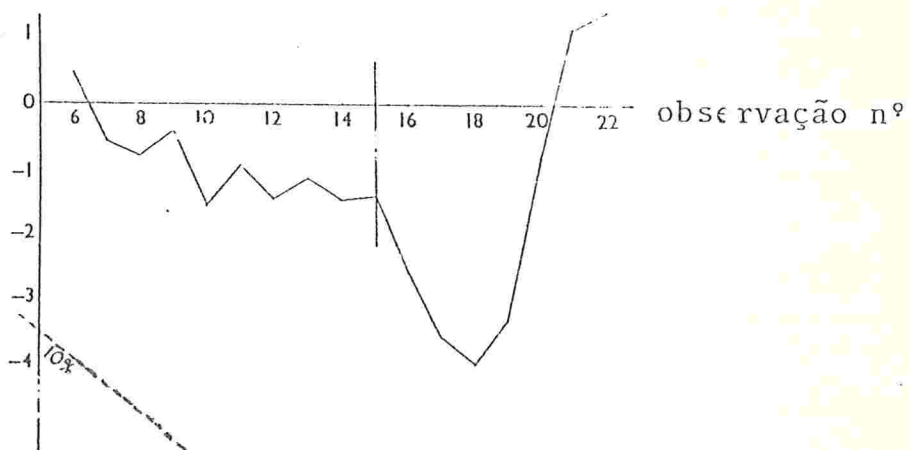


Figura 3.3 - CUSUMS dos resíduos recursivos

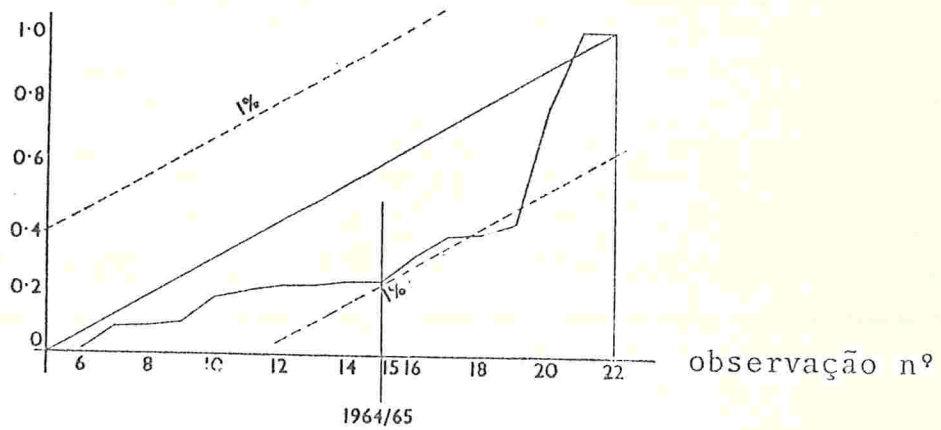


Figura 3.4 - CUSUMS dos quadrados dos resíduos recursivos.

Não se vê sinal de instabilidade até o ano de 1964/65, havendo sinais de instabilidade após este ano.

Entretanto não se alcança o nível de significância no gráfico dos CUSUMS dos resíduos recursivos.

A figura 3.4 mostra o gráfico dos CUSUMS dos quadrados dos resíduos recursivos, S_r onde

$$S_r = \frac{\sum_{j=k+1}^r w_j^2}{\sum_{j=k+1}^T w_j^2} \quad r = k+1, \dots, T$$

Este teste gráfico é um complemento ao nosso teste ante

rior e segue no seu desenvolvimento as mesmas linhas do anterior.

O gráfico 3.4, nos dá mais mostras da evidência de instabilidade depois de 1964/65, atingindo o nível de significância de 1%. O fato da significância ser atingida pelo CUSUM dos quadrados dos resíduos recursivos, mas não pelo CUSUM dos resíduos recursivos, sugere que a instabilidade pode ser devida a uma mudança na variância residual, mais do que nos coeficientes de regressão. Entretanto, posterior análise através de regressões que se movem, que definiremos a seguir, demonstram o contrário.

Como verificou-se posteriormente que a previsão para 1971/72, baseada dos dados até 1964/65 era muito próxima do realmente ocorrido, a decisão tomada foi ignorar os estimativas suspeitas entre 1964/65 e 1971/72 e ajustar o modelo aos outros dados para futuras previsões.

É ainda devido a Durbin um outro teste que aqui citaremos brevemente: Regressões que se movem.

Para fazer este teste ajusta-se a regressão a um curto segmento de n observações sucessivas e move-se este segmento através da série. O gráfico dos coeficientes resultantes, contra o tempo, nos dão mais uma maneira de detetar mudanças.

Além disto, com este método, a variância resi

dual estimada pode ser computada e colocada num gráfico, para investigar a constância de σ^2 . O teste encontra-se desenvolvido no artigo de Brown, Durbin e Evans (1974/75).

Juntamente com os outros dois testes vistos neste capítulo temos em mãos uma arma eficaz para detectar mudanças.

CAPÍTULO 4

4.1 - INTRODUÇÃO

Trata-se aqui, como nos dois capítulos anteriores, de detectar se houve ou não mudança em modelos de regressão. Aqui procuraremos um teste localmente mais poderoso para isto. O teste é desenvolvido para o caso específico de testar a hipótese de que o declive de um modelo linear, numa série de tempo, não muda, contra a alternativa de que o parâmetro muda exatamente uma vez. É um teste sensível, indicado ainda quando a mudança potencial é pequena relativa à variância do erro. É devido a Farley e Hinich (1970) e será desenvolvido a seguir, sendo baseado no princípio da máxima verossimilhança.

4.2 - MODELO DE ANÁLISE E DESENVOLVIMENTO DO MÉTODO.

O modelo básico $y_t = \underline{x}'_t \underline{\beta}_t + \epsilon_t$, aqui será escrito

$y_t = \alpha_t + \beta_t x_t + \epsilon_t$, $t=1, \dots, T$, uma simples reta, sendo x uma variável suposta medida sem erro, e ϵ_t , v.a. $i \sim N(0, \sigma^2)$. Temos ainda

$$\beta_t = \begin{cases} \beta & \text{se } 1 \leq t \leq T_0 \\ \beta + \delta & \text{se } T_0 \leq t \leq T \end{cases} \quad \alpha_t = \begin{cases} \alpha & \text{se } 1 \leq t \leq T_0 \\ \alpha - \delta x_{T_0} & \text{se } T_0 \leq t \leq T \end{cases}$$

O ponto de mudança T_0 é uma v.a. que vamos supor ter distribuição uniforme.

Sem dúvida poderíamos pensar em outra distribuição para T_0 , se tivéssemos alguma informação prévia sobre a distribuição do ponto de mudança, o que, infelizmente, em geral não ocorre.

Assim, a probabilidade é suposta igualmente distribuída para qualquer um dos pontos, isto é,

$$P(T_0 = t_0) = \frac{1}{T} \quad t_0 = 1, \dots, T$$

Testaremos

H_0 : Não há mudança de declive,

H_1 : Ocorre uma mudança de tamanho δ em $T_0 = t_0$.

Sob H_0

$E_0(\underline{y}) = \alpha \underline{1} + \beta \underline{x}$ onde \underline{x} e \underline{y} são os vetores colunas das ob

servações e $\underline{1}$ é um vetor coluna formado de uns.

Se um salto ocorre, de tamanho δ , quando $T_0=t_0$, então, sendo $E_1(\)$ a esperança calculada sob H_1

$$E_1(\underline{y} | T_0=t_0) = \alpha \underline{1} + \beta \underline{x} + \delta \underline{z}(t_0) \text{ onde}$$

$$\underline{z}(t_0)' = (0, \dots, 0, x_{t_0+1}-x_{t_0}, \dots, x_T-x_{t_0})$$

e a esperança incondicional será

$$E_1(\underline{y}) = \frac{1}{T} \sum_{s=1}^T E_1(\underline{y} | T_0=s) = \alpha \underline{1} + \beta \underline{x} + \frac{\delta}{T} \sum_{s=1}^T \underline{z}(s)$$

Vamos definir

$$\underline{\theta}' = (0, \frac{1}{T} \sum_{j=1}^2 (x_2-x_j), \dots, \frac{1}{T} \sum_{j=1}^T (x_T-x_j)).$$

Assim

$$E_1(\underline{y}) = \alpha \underline{1} + \beta \underline{x} + \delta \underline{\theta}$$

Ora, sob H_0 , a função de verossimilhança será

$$f_0(\underline{y}) = (2\pi\sigma^2)^{-\frac{T}{2}} \exp \left[-\frac{1}{2\sigma^2} (\underline{y}-\underline{\mu})' (\underline{y}-\underline{\mu}) \right]$$

Sob H_1

$$f_1(\underline{y}) = (2\pi\sigma^2)^{-\frac{T}{2}} \sum_{s=1}^T T^{-1} \exp \left[-\frac{1}{2\sigma^2} (\underline{y}-\underline{\mu}-\delta \underline{z}(s))' (\underline{y}-\underline{\mu}-\delta \underline{z}(s)) \right]$$

A razão de verossimilhança $f_1(\underline{y})/f_0(\underline{y})$ é então

$$\lambda(\underline{y}/\alpha, \beta, \delta) = \sum_{s=1}^T T^{-1} \exp \left[\frac{\delta}{\sigma^2} (\underline{y}-\underline{\mu})' \underline{z}(s) - \frac{\delta^2}{2\sigma^2} \underline{z}(s)' \underline{z}(s) \right]$$

Expandindo λ em torno de $\delta=0$

$$\lambda = 1 + \sum_{s=1}^T T^{-1} \frac{\delta}{\sigma^2} (\underline{y}-\underline{\mu})' \underline{z}(s) + o(\delta^2) = 1 + \frac{\delta}{\sigma^2} (\underline{y}-\underline{\mu})' \underline{\theta} + o(\delta^2)$$

Trocando a ordem de $\underline{\theta}$ e $\underline{y}-\underline{\mu}$, a aproximação de 1.^a ordem para o teste da razão de verossimilhança é um teste que rejeita H_0 se

$S = \underline{\theta}'(\underline{y}-\underline{\alpha}\underline{1}-\beta\underline{x})$ for significativamente diferente de zero.

Para α, β, δ e σ^2 fixos o teste é localmente mais poderoso à medida que $\delta \rightarrow 0$. (Capon (1961) mostrou que um teste localmente mais poderoso, sob certas condições, de regularidade para a função distribuição da população, que no nosso caso estão preenchidas, é assintoticamente eficiente comparado com o teste da razão de verossimilhança).

Os parâmetros α e β são em geral desconhecidos. Vamos estimá-los usando estimadores de mínimos quadrados sob H_0 : $\hat{\alpha}$ e $\hat{\beta}$. Em notação matricial temos, (segundo Searle)

$$\tilde{S} = \underline{\theta}'(\underline{y} - \hat{\alpha}\underline{1} - \hat{\beta}\underline{x}) = \underline{\theta}'(I-A)\underline{y}, \text{ onde}$$

A é a matriz:

$$A = \frac{1}{T} (\underline{1}\underline{1}') + \frac{1}{\underline{x}'\underline{x} - T\bar{x}} (\underline{x} - \bar{x})(\underline{x} - \bar{x})'$$

\bar{x} tendo como cada elemento a média aritmética da variável explanatória, isto é, tendo $\bar{x} = \frac{1}{T} (\underline{x}'\underline{1})$ como cada elemento.

$E_0(\tilde{S}) = 0$ visto que $\hat{\alpha}$ e $\hat{\beta}$ são não viesados

$$E_1(S) = \delta \underline{\theta}'(I-A)\underline{\theta} \quad ; \quad \sigma^{-2}V_0(S) = \underline{\theta}'(I-A)\underline{\theta}$$

$$\sigma^{-2}V_1(\tilde{S}) = \underline{\theta}'(I-A)\underline{\theta} + 0\left(\frac{\delta^2}{\sigma^2}\right) = \sigma^{-2}V_0(\tilde{S}) + 0\left(\frac{\delta^2}{\sigma^2}\right)$$

Vide provas no apêndice B.

Sob H_0 , \tilde{S} é normalmente distribuída com média 0 e variância $V_0(\tilde{S}) = \underline{\theta}'(I-A)\underline{\theta}$.

Dado um salto em $T_0 = t_0$, \tilde{S} é normalmente distribuído com média $\delta \underline{\theta}'(I-A)\underline{z}(t_0)$ e variância $V_0(\tilde{S})$. Logo sob H_1 , a distribuição de \tilde{S} é uma média sobre t_0 destas distribuições normais, que em geral é multimodal.

Ao nível de significância α , H_0 é rejeitada se $|\tilde{S}| > z_{\alpha/2} \sigma \sqrt{\underline{\theta}'(I-A)\underline{\theta}}$.

Acontece porém que na maioria dos casos σ^2 é desconhecida. Então usamos seu estimador usual

$$s^2 = \frac{1}{T-2} \sum_{t=1}^T (y_t - \hat{\alpha} - \hat{\beta} x_t)^2 = \frac{1}{T-2} \underline{y}'(I-A)\underline{y}$$

$$\text{Sob } H_0 \quad \frac{(T-2)s^2}{\sigma^2} \sim \chi^2_{v = T-2}$$

Dada uma mudança em $T_0 = t_0$

$$\frac{(T-2)s^2}{\sigma^2} \sim \chi^2(T-2, \lambda)$$

$$\lambda = \frac{1}{2} \delta^2 \underline{z}(t_0)' (I-A) \underline{z}(t_0)$$

onde λ é parâmetro de não centralidade.

Entretanto \bar{S} e s são correlacionados mesmo sob H_0 , e então a distribuição de $\frac{\bar{S}}{s}$ não é uma distribuição de Student. Entretanto sob H_0 $s^2 \rightarrow \sigma^2$ (s^2 tende em probabilidade a σ^2) à medida que $t \rightarrow \infty$.

Então

$$\text{Prob} \left[\frac{|\bar{S}|}{s \sqrt{\underline{\theta}' (I-A) \underline{\theta}}} > Z_{\alpha/2} \right] \rightarrow \alpha$$

4.3 - EXEMPLO

Através de um exemplo veremos o uso do teste. Para maior compreensão dos métodos, usamos o mesmo exemplo do capítulo II, parágrafo 2.4.2 - Exemplo 2.

Como sabemos, pela maneira como os dados foram gerados, temos duas retas, isto é, há uma mudança no declive, que sabemos se situa após o 3º ponto. Vamos verificar que o nosso teste detecta uma mudança (os erros são supostos $v a i \sim N(0; 0,03)$). Temos

$$\underline{y}' = (0,01; 1,01; 1,99; 2,92; 3,10; 3,21; 3,29; 3,42; 3,51)$$

$$\underline{x}' = (0; 1; 2; 3; 4; 5; 6; 7; 8).$$

Fazendo os cálculos teremos:

$$\underline{\varrho}' = (0; \frac{1}{9}(1); \frac{1}{9}(1+2); \frac{1}{9}(1+2+3); \dots \frac{1}{9}(1+2+3+4+5+6+7+8))$$

$$\underline{\varrho}' = \frac{1}{9} (0; 1; 3; 6; 10; 15; 21; 28; 36)$$

$$A = \frac{1}{9} \begin{bmatrix} 1 & 1 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} + \frac{1}{[0, \dots, 8] \begin{bmatrix} 0 \\ \vdots \\ 8 \end{bmatrix} - 9(4)^2} \begin{pmatrix} -4 \\ -3 \\ \vdots \\ 4 \end{pmatrix} [-4 \dots 4]$$

$$\underline{\varrho}'(I - A)\underline{\varrho} = 143,02$$

$$\sigma = 0,03$$

$$S = \underline{\varrho}'(I - A)\underline{y} = 245,13$$

$$\text{Prob} \left[\frac{|S|}{\sigma \sqrt{\underline{\varrho}'(I - A)\underline{\varrho}}} > Z_{\alpha/2} \right] = 0,10$$

$$Z_{\alpha/2} \cdot \sigma \sqrt{\underline{\varrho}'(I - A)\underline{\varrho}} = 1,645 \cdot 0,03 \cdot \sqrt{143,2} = 0,59 \quad . \quad \text{Como}$$

$|S| > 0,59$. Rejeita-se H_0 , isto é há uma mudança no declive.

CAPÍTULO 5

5.1 - INTRODUÇÃO

É obvio que uma vez detectada uma mudança no mo
dêlo, cumpre-nos estimar em que ponto ocorreu esta mudan
ça. Neste capítulo estudaremos a proposta de Quant(1958)pa
ra determinar o estimador de máxima verossimilhança do
ponto de mudança, em relações lineares.

5.2 - DETERMINAÇÃO DO MODELO E DESENVOLVIMENTO DO MÉTODO.

Temos um total de T observações, x_1, \dots, x_T . O
nosso modelo toma a forma

$$\begin{aligned}y_t &= \alpha_1 + \beta_1 x_t + \varepsilon_1 && \text{para } t \leq t_0 \\y_t &= \alpha_2 + \beta_2 x_t + \varepsilon_2 && \text{para } t_0 + 1 \leq t \leq T \\ \varepsilon_1, \varepsilon_2 & \text{ v.a. } i \sim N(0, \sigma_i) && i = 1, 2\end{aligned}$$

e t_0 é desconhecido.

O nosso problema é estimar t_0 , o ponto a partir do qual o sistema muda de um regime para outro. Queremos ainda estimar os parâmetros $\alpha_{1,2}$ e $\beta_{1,2}$

A função de verossimilhança para a amostra toda, condicionada a um certo t_0 , é:

$$\left(\frac{1}{\sqrt{2\pi\sigma_1^2}} \right)^{t_0} \left(\frac{1}{\sqrt{2\pi\sigma_2^2}} \right)^{T-t_0} \exp \left[- \frac{1}{2\sigma_1^2} \sum_{t=1}^{t_0} (y_t - \alpha_1 - \beta_1 x_t)^2 - \frac{1}{2\sigma_2^2} \sum_{t=t_0+1}^T (y_t - \alpha_2 - \beta_2 x_t)^2 \right]$$

e a função log de verossimilhança é

$$L = -T \log \sqrt{2\pi} - t_0 \log \sigma_1 - (T-t_0) \log \sigma_2 - \frac{1}{2\sigma_1^2} \sum_{t=1}^{t_0} (y_t - \alpha_1 - \beta_1 x_t)^2 - \frac{1}{2\sigma_2^2} \sum_{t=t_0+1}^T (y_t - \alpha_2 - \beta_2 x_t)^2$$

Os estimadores de máxima verossimilhança para α_1 , β_1 , α_2 e β_2 são obtidos pelos métodos usuais e são dados por:

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{t_0} x_t y_t - \frac{\sum_{t=1}^{t_0} x_t \sum_{t=1}^{t_0} y_t}{t_0}}{\sum_{t=1}^{t_0} x_t^2 - \frac{(\sum_{t=1}^{t_0} x_t)^2}{t_0}} ; \hat{\alpha}_1 = \frac{\sum_{t=1}^{t_0} y_t}{t_0} - \hat{\beta}_1 \frac{\sum_{t=1}^{t_0} x_t}{t_0}$$

$$\hat{\beta}_2 = \frac{(T-t_0) \sum_{t=t_0+1}^T x_t y_t - \sum_{t=t_0+1}^T x_t \sum_{t=t_0+1}^T y_t}{(T-t_0) \sum_{t=t_0+1}^T x_t^2 - \left(\sum_{t=t_0+1}^T x_t \right)^2} ;$$

$$\hat{\alpha}_2 = \frac{\sum_{t=t_0+1}^T y_t}{T-t_0} - \hat{\beta}_2 \frac{\sum_{t=t_0+1}^T x_t}{T-t_0}$$

E ainda os estimadores de σ_1^2 e σ_2^2 são fornecidos

por

$$\hat{\sigma}_1^2 = \frac{\sum_{t=1}^{t_0} (y_t - \hat{\alpha}_1 - \hat{\beta}_1 x_t)^2}{t_0} ;$$

$$\hat{\sigma}_2^2 = \frac{\sum_{t=t_0+1}^T (y_t - \hat{\alpha}_2 - \hat{\beta}_2 x_t)^2}{T-t_0}$$

O logaritmo da função de máxima verossimilhança será:

$$L(t_0) = -T \log \sqrt{2\pi} - t_0 \log \hat{\sigma}_1 - (T-t_0) \log \hat{\sigma}_2 - \frac{T}{2} \quad (5.1)$$

e é função de t_0 somente (para T dado).

Devemos agora determinar o valor t_0 que maximize $L(t_0)$. Derivar em relação a t_0 e igualar a zero não é apropriado pois t_0 não é contínua.

Procurar um determinado valor t_0 tal que

$$L(t_0-1) < L(t_0) \quad \text{e} \quad L(t_0+1) < L(t_0)$$

também não parece apropriado pois poderemos ter vários máximos. Então recomenda-se que se calcule o valor do log da função de máxima verossimilhança para todos os valores possíveis de t_0 e procure aquele t_0 , que denotaremos \hat{t}_0 que nos dê o máximo dos máximos.

Para isto devemos usar o seguinte procedimento:

i) Com os dados ordenados de acordo com o período de tempo, $(x_1, y_1) \dots (x_T, y_T)$, dividir o conjunto de pontos em dois subconjuntos

$$\{(x_1, y_1) \dots (x_{t_0}, y_{t_0})\} \text{ e } \{(x_{t_0+1}, y_{t_0+1}) \dots (x_T, y_T)\}$$

ii) Estimar as duas retas de regressão, uma para cada subconjunto. Calcule para esta divisão o valor da função 5.1.

iii) Então movimente o ponto de divisão entre os 2 subconjuntos e proceda como no ítem ii)

O estimador \hat{t}_0 de máxima verossimilhança para t_0 é aquele valor t_0 para o qual temos o maior valor de 5.1. E os estimadores $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\alpha}_2$ e $\hat{\beta}_2$ serão os correspondentes estimadores da regressão com ponto de mudança \hat{t}_0 .

5.3 - EXEMPLO

Ilustraremos o procedimento com os dados da tabela 5.1. devido a Quant (1958).

Estes dados foram gerados da seguinte maneira: 20 pontos (x), com valores de um a 20 foram sorteados e colocados na tabela na ordem em que foram obtidos, ou, o que seria equivalente, foi sorteada uma permutação dos 20 pontos de 1 a 20. Os primeiros 12 pontos forneceram valores de y obtidos através da expressão $y_t = 2,5 + 0,7x_t + \epsilon_1$. Os últimos 8 valores das observações x forneceram valores de y gerados por $y_t = 5 + 0,5x_t + \epsilon_2$. ϵ_1 e ϵ_2 são erros normalmente distribuídos, independentes.

- Os pares x e y estão na tabela abaixo

t	x	y
1	4	3,473
2	13	11,555
3	5	5,714
4	2	5,710
5	6	6,046
6	8	7,650
7	1	3,140
8	12	10,312
9	17	13,353
10	20	17,197
11	15	13,036
12	11	8,264
13	3	7,612
14	14	11,802
15	16	12,551
16	10	10,296
17	7	10,014
18	19	15,472
19	18	15,650
20	9	9,871

Tabela 5.1

Vemos que os valores de x não estão ordenados, por exemplo em ordem crescente, porém, na ordem em que foram obtidos. O problema é determinar a posição do verdadeiro ponto de mudança (que, no nosso problema criado, é entre a 12.^a e a 13.^a observação, onde mudamos de modelo).

Regressões separadas foram computadas para todas as possíveis maneiras nas quais as observações podem ser repartidas em dois grupos, sem modificar sua ordem. Logicamente nenhuma regressão foi computada para grupos com menos do que três pontos.

E para todas estas situações, calculou-se o valor da função de verossimilhança 5.1, e os valores foram colocados num gráfico (a menos de uma constante aditiva).

Como vemos pela fig. 5.1, claramente o máximo dos máximos é atingido no ponto $t=12$, que corresponde ao verdadeiro ponto de mudança. Entretanto a função de verossimilhança tem dois outros máximos, em 10 e em 15, e três mínimos, em 4, 11 e 13. O comportamento irregular da função de verossimilhança em direção aos dois extremos não é inesperada pois nos extremos as estimativas da variância do erro são baseadas em muito poucas observações.

Em muitos casos a análise qualitativa do problema pode até sugerir a adoção de um máximo relativo para o ponto de mudança, em vez do máximo absoluto.

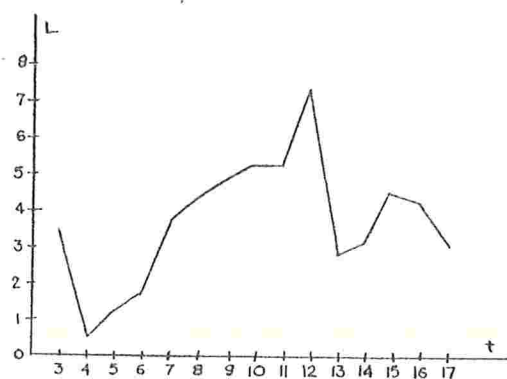


Fig. 5.1 Gráfico da função de verossimilhança para os vários valores de t.

O gráfico mostra-se de enorme utilidade para se determinar o verdadeiro ponto de mudança.

5.4 - CONCLUSÕES

Poderíamos usar um teste, baseado neste processo, para resolver o problema de detectar mudanças, estudado nos capítulos anteriores.

Escolhendo t_0 como o ponto que maximize a função de verossimilhança conjunta das duas regressões, 5.1, um teste χ^2 poderia ser usado numa razão de verossimilhança, como um teste aproximado para testar se os parâmetros de uma única regressão grupada para todos os dados dife

rem dos das duas subregressões, porém esta distribuição difere marcadamente de uma χ^2 , sob a hipótese nula de que não há mudança. Daí Farley e Hinich, terem desenvolvido seu teste mais poderoso, que vimos no capítulo IV.

CAPÍTULO VI

6.1 - INTRODUÇÃO

No capítulo anterior vimos a maneira dada por Quant para estimar o ponto de mudança quando o modelo consiste em duas retas que se encontram num ponto desconhecido. Vimos que o processo exigia que se tentassem todos os pontos t_0 ($2 \leq t_0 \leq T-2$) como possíveis candidatos ao ponto de mudança e que se escolhesse aquele que nos desse o máximo dos máximos na função de verossimilhança.

Neste capítulo, além de não nos restringirmos obrigatoriamente a duas retas, veremos uma maneira de reduzir os trabalhosos cálculos anteriores. A proposta se deve a Derek J. Hudson (1966). Será dada a solução de mínimos quadrados para dois modelos quaisquer, mas pode-se facilmente generalizar para mais do que dois. Os modelos aqui precisam se interceptar, isto é, o modelo é contínuo no ponto de mu

dança, diferindo do problema de Quant que não fazia esta restrição.

6.2 - DETERMINAÇÃO DO MODELO

O modelo será definido por

$$\begin{cases} y_1 = f_1(x, \beta_1) & \text{para } x_1 \leq x \leq x_{t_0} \\ y_2 = f_2(x, \beta_2) & \text{para } x_{t_0+1} \leq x \leq x_T \end{cases}$$

os dois modelos se unem em $x = \zeta$

Queremos encontrar os vetores $\hat{\beta}_1$ e $\hat{\beta}_2$ e os valores $\hat{\zeta}$ e \hat{t}_0 sujeitos às condições

$$\left. \begin{array}{l} 1) f_1(\hat{\zeta}, \hat{\beta}_1) = f_2(\hat{\zeta}, \hat{\beta}_2) \\ 2) x_{\hat{t}_0} \leq \hat{\zeta} \leq x_{\hat{t}_0+1} \end{array} \right\} \begin{array}{l} \text{condições de continuidade no} \\ \text{ponto de mudança} \end{array}$$

e que minimizem os resíduos de mínimos quadrados.

$$R(\beta_1, \beta_2, \zeta, t_0) = \sum_{t=1}^{t_0} \left[y_t - f_1(x_t, \beta_1) \right]^2 + \sum_{t=t_0+1}^T \left[y_t - f_2(x_t, \beta_2) \right]^2$$

Necessitamos de algumas definições e notações, a saber: a solução de mínimos quadrados para os parâmetros e para a curva ajustada será denotada por $\{\hat{\beta}_1, \hat{\beta}_2, \hat{\zeta}\}$ e $\hat{f}(x)$ (ou \hat{y}), respectivamente.

Se $\underline{\beta}_1(\underline{\beta}_2)$ é estimado por mínimos quadrados usando somente os dados num determinado intervalo, então o estimador de $\underline{\beta}_1(\underline{\beta}_2)$, que é único, e denotaremos por $\underline{\beta}_1^*(\underline{\beta}_2^*)$ é chamado estimador local de mínimos quadrados. A função ajustada correspondente será denotada por

$$f_1^* = f_1^*(x, \underline{\beta}_1^*) \quad (f_2^* = f_2^*(x, \underline{\beta}_2^*))$$

Como se sabe o estimador de mínimos quadrados será também o de máxima verossimilhança, se os erros forem v.a.i. $\sim N(\mu, \sigma^2)$.

6.3 - QUATRO TIPOS DE UNIÃO NA SOLUÇÃO GLOBAL.

Os dois submodêlos da solução global se unem em $x = \hat{\zeta}$. Como é evidente, eliminaremos de discussão futura os casos em que $\hat{\zeta} < x_1$ ou $\hat{\zeta} > x_T$. Com relação à posição de $\hat{\zeta}$ e à maneira como esta intersecção ocorre, podemos dividir a união (intersecção dos dois modêlos) em três tipos conforme quadro abaixo.

	$\hat{\zeta} \neq x_{t_0}$	$\hat{\zeta} = x_{t_0}$
$g'(\hat{\zeta}) \neq 0$	I	II
$g'(\hat{\zeta}) = 0$	III	II

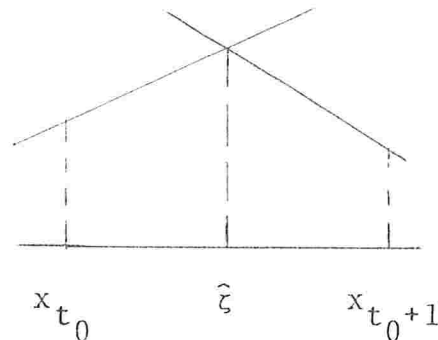
onde $g(\hat{\zeta}) = f_1(\hat{\zeta}, \underline{\beta}_1) - f_2(\hat{\zeta}, \underline{\beta}_2)$

Derivando com relação a $\hat{\zeta}$ temos

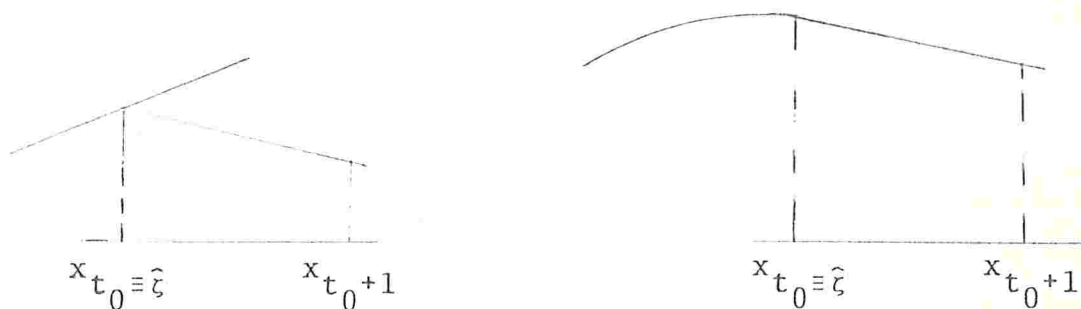
$$g'(\hat{\zeta}) = f'_1(\hat{\zeta}, \hat{\beta}_1) - f'_2(\hat{\zeta}, \hat{\beta}_2)$$

A intersecção do tipo I ocorre quando o ponto de intersecção ($\hat{\zeta}$) não coincide com nenhum t_0 e os declives dos dois submodêlos são diferentes.

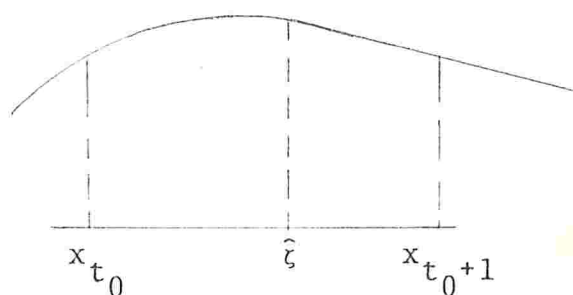
No caso em que o modelo consiste de duas retas, a figura abaixo ilustra a situação.



A intersecção é do tipo II quando o ponto intersecção $\hat{\zeta}$ coincide com algum x_{t_0} , sejam os declives iguais ou diferentes, não se fazendo distinção para estes dois subcasos como ilustram as figuras



A intersecção do tipo III quando $\hat{\zeta}$ não coincide com x_{t_0} mas os declives são iguais.



Em linhas gerais o raciocínio que vamos usar para resolver o nosso problema de determinar os estimadores $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\zeta}$ e \hat{t}_0 é o seguinte: Como primeiro Passo, suporemos que a intersecção é do tipo I e que t_0 é conhecido. Obteremos a solução de mínimos quadrados para este caso e determinaremos a Soma dos Quadrados dos Resíduos. A seguir abandona-se a suposição de que t_0 é conhecido e passa-se à investigação para todos os t_0 , ainda dentro da suposição de união tipo I. Determina-se a solução de mínimos quadrados pa

ra todos êstes possíveis t_0 , calculando as respectivas Somas dos Quadrados dos Resíduos (SQR). Passa-se sô então ã suposição de que a união é do tipo II (2º Passo) inicialmente ainda com t_0 conhecido. Determina-se a solução de mínimos quadrados para este tipo de união, que será uma solução de mínimos quadrados condicionada ao fato da intersecção ser em t_0 . Determina-se a SQR e repete-se estes cálculos para todos os possíveis valores de t_0 , em união tipo II. Então vamos ao 3º Passo, repetindo o raciocínio, agora usando um tratamento individual, para cada situação a fim de estimar $\underline{\beta}$, e encontrar SQR como função de ζ partindo ainda primeiramente da suposição de que t_0 é conhecido.

O valor \hat{t}_0 e o seu correspondente $\hat{\zeta}$ que der menor SQR, computando para todos os tipos de intersecção é a solução procurada. Veremos no estudo do método que não será necessário pesquisar todas as possíveis possibilidades para $\hat{\zeta}$, pois alguns pontos podem ser previamente abandonados graças a conclusões obtidas em passos anteriores.

6.4 - RESOLUÇÃO DO PROBLEMA

1º Passo

A união é suposta do tipo I, isto é, $\hat{\zeta} \neq x_{t_0}$ e $g'(\zeta) \neq 0$, e t_0 é suposto conhecido.

Vamos provar através de um Teorema(1) que a solu-

ção global consiste simplesmente dos estimadores locais de mínimos quadrados, isto é,

Para minimizar

$$R = \sum_{t=1}^{t_0} \left[y_t - f_1(x_t, \underline{\beta}_1) \right]^2 + \sum_{t=t_0+1}^T \left[y_t - f_2(x_t, \underline{\beta}_2) \right]^2$$

com respeito a β_1 e β_2 , sujeito às restrições

$$\begin{cases} 1) & f_1(\tilde{\zeta}, \hat{\beta}_1) = f_2(\tilde{\zeta}, \hat{\beta}_2) \\ 2) & x_{t_0} < \tilde{\zeta} < x_{t_0+1} \end{cases}$$

com $\frac{\partial f_1(\tilde{\zeta}, \hat{\beta}_1)}{\partial x} \neq \frac{\partial f_2(\tilde{\zeta}, \hat{\beta}_2)}{\partial x}$

é suficiente minimizar as duas somas separadamente, o que é provado através do seguinte teorema:

1º Teorema

Se i) $x_{t_0} < \tilde{\zeta} < x_{t_0+1}$

ii) $g'(\tilde{\zeta}) \neq 0$

Então $\begin{cases} \hat{\beta}_1 = \beta_1^*(t_0) \\ \hat{\beta}_2 = \beta_2^*(t_0) \end{cases}$ e $\begin{cases} \hat{f}_1 = f_1^* \\ \hat{f}_2 = f_2^* \end{cases}$

iv) $\tilde{\zeta} = \zeta^*(t_0)$

onde $\underline{\beta}_1^*(t_0)$ e $\underline{\beta}_2^*(t_0)$ são os estimadores locais de mínimos quadrados (não condicionados a que as duas funções se interceptem num determinado ponto), denotados anteriormente por $\underline{\beta}_1^*$ e $\underline{\beta}_2^*$, onde se colocou entre parêntesis t_0 para indicar que foram calculados com o ponto de mudança situado em t_0 . $\tilde{\zeta}^*(t_0)$ é o ponto real da intersecção de f_1^* com f_2^* (já definidas anteriormente) situado entre x_{t_0} e x_{t_0+1} .

Para a prova deste teorema vide Apêndice , C.Cl.

O que se conclui é que se a intersecção é do tipo I, com t_0 conhecido, para achar os estimadores globais de mínimos quadrados que procuramos, $\hat{\underline{\beta}}_1$, $\hat{\underline{\beta}}_2$ e $\hat{\zeta}$, basta determinarmos os estimadores de mínimos quadrados locais, $\underline{\beta}_1 = \underline{\beta}_1^*(t_0)$ e $\underline{\beta}_2 = \underline{\beta}_2^*(t_0)$ usando para o 1º estimador os dados (x_1, \dots, x_{t_0}) e para o 2º os dados (x_{t_0+1}, \dots, x_T)

Para determinar $\hat{\zeta}$ substituímos estes valores na restrição 1, resolvendo

$$f_1[\zeta; \underline{\beta}_1^*(t_0)] = f_2[\zeta, \underline{\beta}_2^*(t_0)].$$

Encontraremos $\zeta = \zeta^*(t_0)$. Por hipótese existe ao menos uma solução real ζ . Se $x_{t_0} < \zeta^*(t_0) < x_{t_0+1}$ diremos que as curvas tem uma intersecção $\zeta^*(t_0)$ no lugar certo.

Como na realidade não sabemos onde está t_0 , faremos os cálculos acima para todos os possíveis valores de

t_0 e para cada um dêles calculamos a $SQR^*(t_0)$ (Soma dos Quadrados dos Resíduos para um determinado t_0 , que denotaremos por $SQR^*(t_0)$), que nada mais é do que a soma abaixo

$$SQR^*(t_0) = SQR_1^*(t_0) + SQR_2^*(t_0)$$

onde $SQR_1^*(t_0)$ e $SQR_2^*(t_0)$ são as somas dos quadrados dos resíduos locais.

Se as curvas se encontram no lugar certo, chamaremos a $SQR^*(t_0)$ de $R(t_0) < \infty$. Caso contrário daremos $R(t_0)$ o valor ∞ , sendo neste caso $R(t_0) \neq SQR^*(t_0)$. Isto feito para todos os possíveis t_0 , passemos ao 2º Passo.

2º Passo

A união é suposta do tipo II, isto é, $\tilde{\zeta} \equiv x_{t_0}$

Inicialmente supõe-se t_0 conhecido. Determina-se a solução de mínimos quadrados, sujeita à restrição linear.

$$f_1(x_{t_0}, \beta_1) = f_2(x_{t_0}, \beta_2)$$

Esta será a solução de mínimos quadrados condicionada que se encontra explicada em detalhes no apêndice C.C2. Por esta explicação poderemos facilmente determinar a SQR global, que para este caso de intersecção tipo II, vamos denotar por $S(t_0)$. Como t_0 é desconhecido, faremos os cálculos e determinaremos $S(t_0)$ para todos os possíveis valo-

res de t_0 . Isto feito passaremos ao 3º Passo.

Suposta a união tipo III, e suposto t_0 conhecido, e tendo

$$\left\{ \begin{array}{l} x_{t_0} < \hat{\zeta} < x_{t_0+1} \\ f_1(\hat{\zeta}, \hat{\beta}_1) = f_2(\hat{\zeta}, \hat{\beta}_2) \\ f'_1(\hat{\zeta}, \hat{\beta}_1) = f'_2(\hat{\zeta}, \hat{\beta}_2) \end{array} \right\} \begin{array}{l} \text{definição de união tipo} \\ \text{III e restrições do pro-} \\ \text{blema genérico} \end{array}$$

determinaremos o estimador dos parâmetros. Estes casos entre tanto requerem um tratamento especial para cada caso.

A recomendação é resolvê-los por aproximações sucessivas, com métodos iterativos; se por exemplo $f_1(x)$ é uma parábola e $f_2(x)$ uma reta, escolhemos valores experimentais para ζ , fazemos $z = x - \zeta$ e reescrevemos o modelo como

$$\begin{array}{ll} f_1(z, \beta) = \beta_0 + \beta_1 z + \beta_2 z^2 & \text{para } x_1 - \zeta \leq z \leq 0 \\ f_2(z, \beta) = \beta_0 + \beta_1 z & \text{para } 0 \leq z \leq T - \zeta \end{array}$$

com $g(0) = 0 = g'(0)$. Estima-se agora β e encontra-se SQR como função de ζ . A SQR mínima pode ser encontrada por sucessivas aproximações. Parece entretanto, e felizmente, que união tipo III é um caso raro de acontecer.

Uma vez descritos os tres passos, e como não se sabe que tipo de união realmente temos, devemos iniciar procurando soluções tipo I (1º Passo) e marcar todas as SQR* (t_0) e

caso sejam casos com união no lugar certo marquem $SQR^*(t_0) = R(t_0) < \infty$. Caso contrário faz-se $R(t_0) = \infty \neq SQR^*(t_0)$. Em seguida partimos para a procura de soluções tipo II (2º passo).

Vamos ver que alguns valores de $\hat{\zeta} = x_{t_0}$ podem ser descartados sem necessidade de experimentá-los e esta é uma simplificação valiosa. Se no 1º passo encontrarmos

$$x_{t_0} < \zeta^*(t_0) < x_{t_0+1},$$

isto é, se $\zeta^*(t_0)$ estiver no lugar certo, não precisaremos tentar $\hat{\zeta} = x_{t_0}$ nem $\hat{\zeta} = x_{t_0+1}$ no 2º passo. E mesmo se $\zeta^*(t_0)$ não estiver no lugar certo, sendo então $R(t_0) = \infty$, podemos abandonar x_{t_0} e x_{t_0+1} como possíveis soluções para $\hat{\zeta}$ no 2º passo se o ajuste das curvas de mínimos quadrados locais, no 1º passo, for tão má que $SQR^*(t_0)$ for maior que algum $R(t_0')$ (SQR no 1º passo com ζ no lugar certo) ou que $S(t_0'')$ (SQR no 2º passo com mínimos quadrados condicionados).

Conclui-se então que no 2º passo, necessitamos somente tentar $\hat{\zeta} = x_{t_0}$ se para ambos $m = t_0$ e $m = t_0 - 1$ tivermos $R(m) = \infty$, isto é, intersecção no lugar errado e $SQR^*(m) < \frac{\text{Min}\{R(j)\}}{p/\text{todoj}}$ e ainda $SQR^*(m) < \text{Min}\{S(k)\}$.
p/os k previamente determinados

Estas conclusões podem ser tiradas do 2º Teorema (partes a, b e c) enunciados e demonstrados no apêndice C.C3.

Uma vez cumprido 2º passo, quando então iniciamos a procura de solução tipo III, podemos novamente abandonar algu

mas possibilidades. Na verdade no 3º passo somente necessitamos tentar encontrar uma união tipo III, com $x_{t_0} < \zeta < x_{t_0+1}$, para algum t_0 , se $R(t_0) = \infty$ e $SQR^*(t_0)$ for menor que o mínimo entre $\{R(j), S(j)\}$, Em outras palavras, sã necessitamos tentar uma solução tipo III com $x_{t_0} < \zeta < x_{t_0+1}$, para algum t_0 , se no 1º passo, para t_0 , a intersecção for no lugar "errado", porém a $SQR^*(t_0)$, correspondente, for menor que todos os outros valores de $R(j)$ ($SQR^*(j)$ com j no lugar certo) e também menor que todas as SQR para intersecção tipo II. ($S(j)$). Esta conclusão baseia-se no 3º teorema enunciado e demonstrado no apêndice C.C4.

6.5 - EXEMPLOS

Somente poderemos compreender bem o nosso método se o aplicarmos a alguns exemplos, e é o que faremos a seguir.

6.5.1 - Exemplo 1:

Caso de duas retas unidas em $x = \zeta$, com união tipo I

x	1	2	3	4	5	6
y	1	2	4	4	3	1

1º Passo

Examinemos as 3 possibilidades $\zeta \in (2,3)$ ou $\zeta \in (3,4)$ ou $\zeta \in (4,5)$, isto é, encontremos $R(2)$, $R(3)$ e $R(4)$. As curvas 2 na tabela 6.1 não se encontram no lugar certo, isto é, no intervalo $(2,3)$; então para a SQR daremos o valor $R(2) = \infty$. Na

figura 6.1 estão em linhas interrompidas, "tentando" se encontrar em (2.3), mas não se encontrando. $R(3)$ e $R(4)$ são finitas, e então sabemos do 2º teorema (a) que $\hat{\zeta} \neq 3, 4$, ou 5. No 2º passo tentamos $\hat{\zeta} = 2$ se bem que o 2º teorema (b) nos faça prever que a SQR para $\zeta = 2$ ($S(2)$) será maior que o mínimo $R(k)$ que é $R(3) = \frac{1}{3}$ pois

$$SQR_1^*(2) + SQR_2^*(2) = 1 > \frac{1}{3}. \text{ De fato } S(2) = 5, 9.$$

Os valores para os estimadores se encontram na tabela 6.1; as figuras 6.1 e 6.2 elucidam o que foi feito.

Curva n°	posição de ζ	$f_1(x)$	$f_2(x)$	Ponto união	SQR_1^*	SQR_2^*	R ou S
2	$2 < \zeta < 3$	x	$7,5 - x$	$3\frac{3}{4}$	0	1	∞
3	$3 < \zeta < 4$	$-\frac{2}{3} + 1,5x$	$10\frac{1}{6} - 1,5x$	$3\frac{11}{18}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
4	$4 < \zeta < 5$	$1\frac{1}{10}x$	$13 - 2x$	$4\frac{6}{31}$	$\frac{7}{10}$	0	$\frac{7}{10}$
1	$\alpha = 2$	$-1\frac{2}{5} + 2\frac{2}{5}x$	$4 - \frac{3}{10}x$	2	-	-	$5\frac{9}{10}$
0	$\alpha = 1$	não única	$2\frac{1}{5} + \frac{3}{35}x$	1	-	$9\frac{13}{35}$	$9\frac{13}{35}$

Tabela 6.1

A curva n°0 foi acrescentada mas não é necessário pesquisar-la pois uma única reta não dará tão bom ajuste quanto duas.

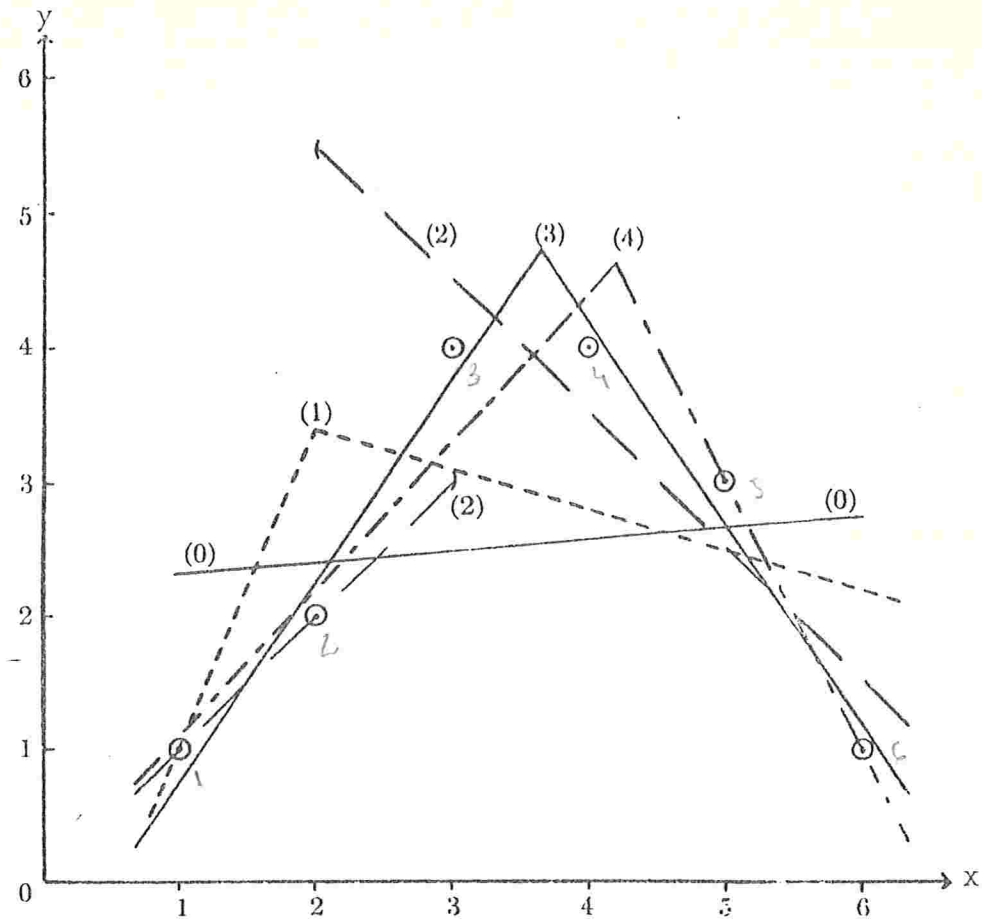
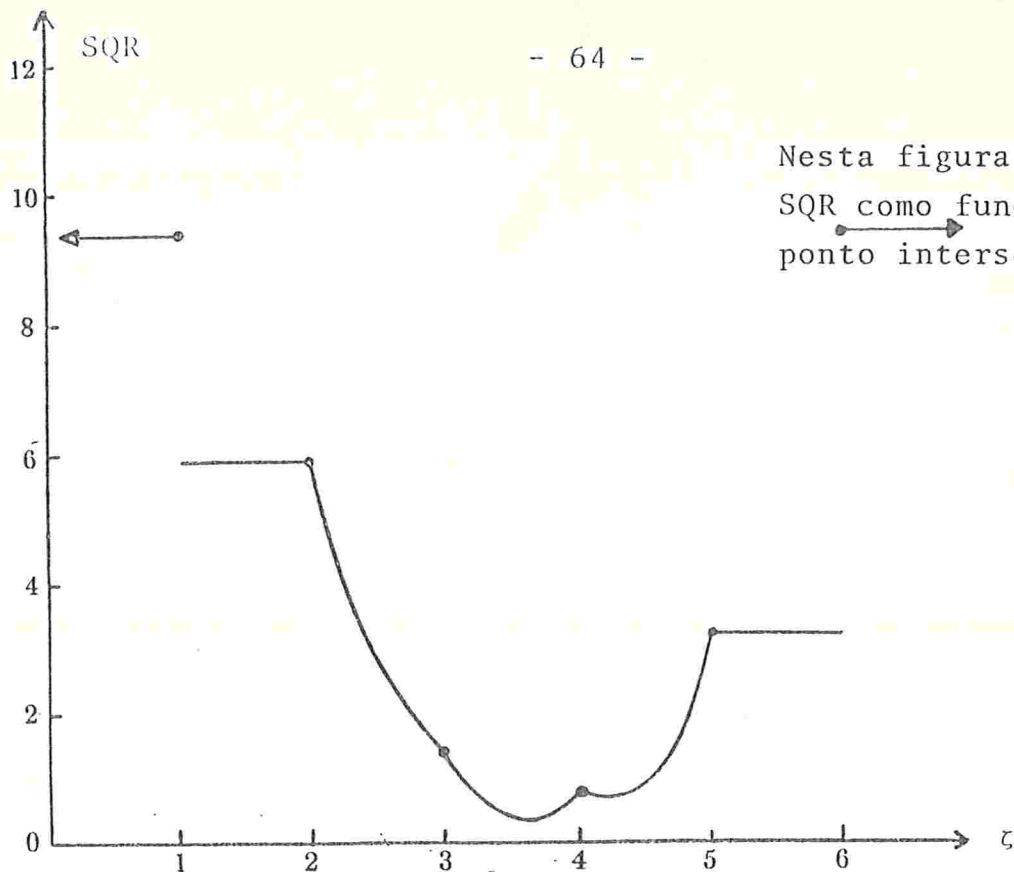


Figura 6.1

A solução global de mínimos quadrados corresponde a $R = \frac{1}{3}$, isto é, a união é do tipo I e $\hat{\beta}'_1 = (-\frac{2}{3}, 1\frac{1}{2})$; $\hat{\beta}'_2 = (10\frac{1}{6}, -1\frac{1}{2})$; $\hat{\zeta} = 3\frac{11}{18}$



Nesta figura vemos a SQR como função do ponto intersecção ζ .

Figura 6.2

6.5.2 - Exemplo 2:

Ajuste de duas retas unidas em $x=\zeta$. Veremos por este exemplo que 2º passo não é inútil. Dados

x	1	2	3	4	5	6
y	1	2	4	7	3	1

1º Passo

Os valores calculados para as curvas 2, 3 e 4 estão na tabela (6.2). Nenhuma das curvas, neste passo se interceptam no lugar certo, donde o valor ∞ atribuído às S.Q.Re síduos.

2º Passo

Escolhemos $\zeta = 4$ como primeira tentativa da forma $\zeta = x_{t_0}$ e encontramos $S(4) = 1\frac{15}{38}$.

Isto imediatamente implica que $\hat{\zeta} \neq 2$ e $\hat{\zeta} \neq 3$ pois $SQR^*(2) = 10\frac{3}{10}$ e então $S(2) \geq 10\frac{3}{10}$

Curva n°	Posição de ζ	$f_1(x)$	$f_2(x)$	Ponto Inters.	SQR_1^*	SQR_2^*	R ou S
2	$2 < \zeta < 3$	x	$9\frac{3}{5} - 1\frac{3}{10}x$	$4\frac{4}{23}$	0	$10\frac{3}{10}$	∞
3	$3 < \zeta < 4$	$-\frac{2}{3} + 1\frac{1}{2}x$	$18\frac{2}{3} - 3x$	$4\frac{8}{27}$	$\frac{1}{6}$	$\frac{2}{3}$	∞
4	$4 < \zeta < 5$	$-1\frac{1}{2} + 2x$	$13 - 2x$	$3\frac{5}{8}$	1	0	∞
3a	$\zeta = 4$	$-1\frac{7}{19} + 1\frac{35}{38}x$	$17\frac{9}{19} - 2\frac{15}{19}x$	$\zeta = 4$	-	-	$1\frac{15}{38}$
5	$\zeta = 5$	$\frac{7}{10} + \frac{9}{10}x$	$26\frac{1}{5} - 4\frac{1}{5}x$	$\zeta = 5$	-	-	$13\frac{1}{10}$

Tabela 6.2

A solução global de mínimos quadrados corresponde a $S(4) = 1\frac{15}{38}$, isto é, a união é do tipo II e

$$\hat{\beta}'_1 = (-1\frac{7}{19}, 1\frac{35}{38}); \hat{\beta}'_2 = (17\frac{9}{19}, -2\frac{15}{19});$$

$$\hat{\zeta} = 4$$

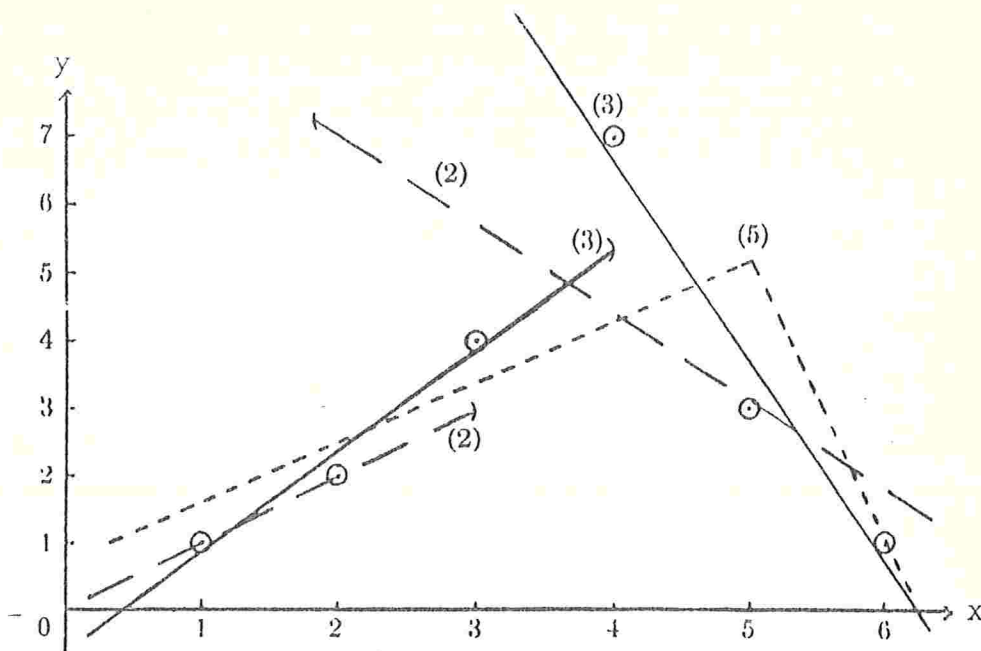


Figura 6.3

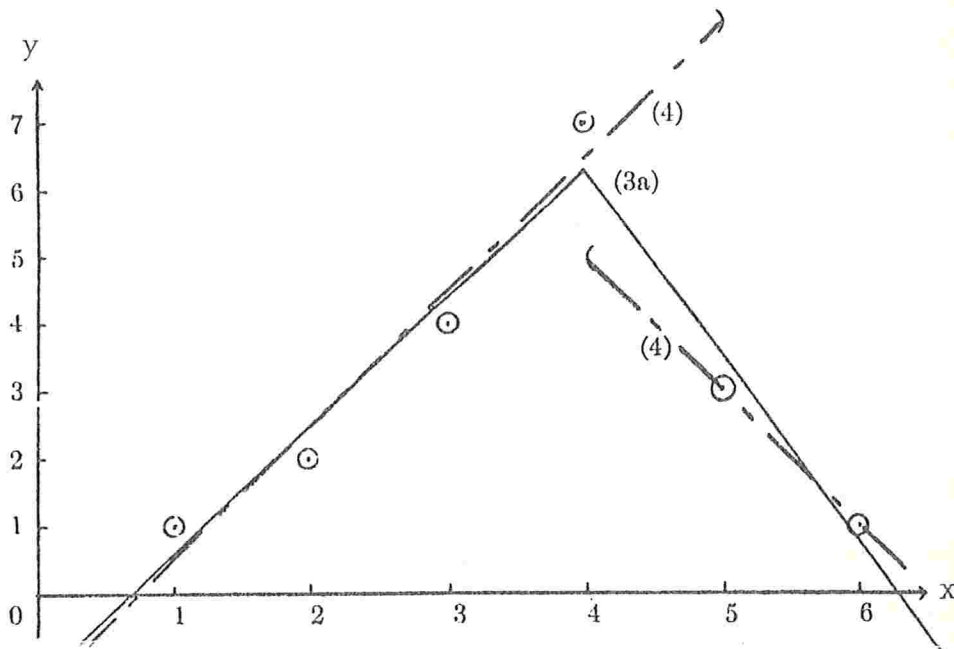


Figura 6.4

A figura 6.3 representa as curvas 2, 3 e 5 e a figura 6.4 as curvas 3(a) e 4, para evitar confusão.

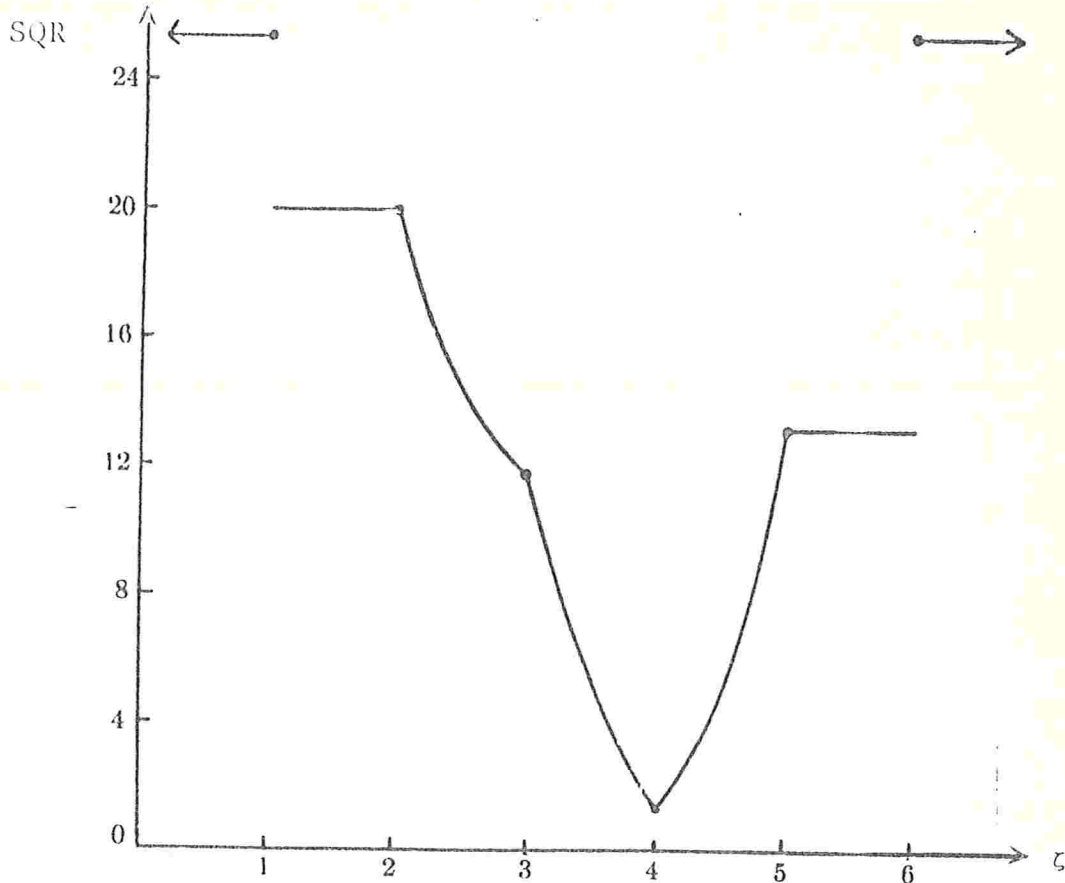


Figura 6.5

A figura 6.5 representa a SQR como função de ζ .

6.5.3 - Exemplo 3:

Neste exemplo veremos um cosseno unido a uma reta.

Dados

x	-3	-2	-1	2	3	4
y	0	4	7	7	6	5

O modelo é

$$f_1(x, \underline{\beta}_1) = \beta_{11} \cos \frac{\pi x}{6} \quad -3 \leq x \leq \zeta$$

$$f_2(x, \underline{\beta}_2) = \beta_{20} + \beta_{21} x \quad \zeta \leq x \leq 4$$

Ajustamos um cosseno aos três primeiros pontos e uma reta aos pontos restantes e vemos que as duas curvas não se encontram, o que nos sugere que a união é do tipo III, com $-1 < \hat{\zeta} < 2$ (figura 6.6). A solução é primeiramente obtida por sucessivas aproximações, depois que as duas restrições

$$f_1(\hat{\zeta}, \hat{\beta}_1) = f_2(\hat{\zeta}, \hat{\beta}_2) \quad \text{e}$$

$$f'_1(\hat{\zeta}, \hat{\beta}_1) = f'_2(\hat{\zeta}, \hat{\beta}_2)$$

foram ambas introduzidas no modelo.

Com isto os modelos ficam

$$f_1(x, \underline{\beta}) = \beta_{11} \cos \frac{\pi x}{6} \quad -3 < x < \zeta$$

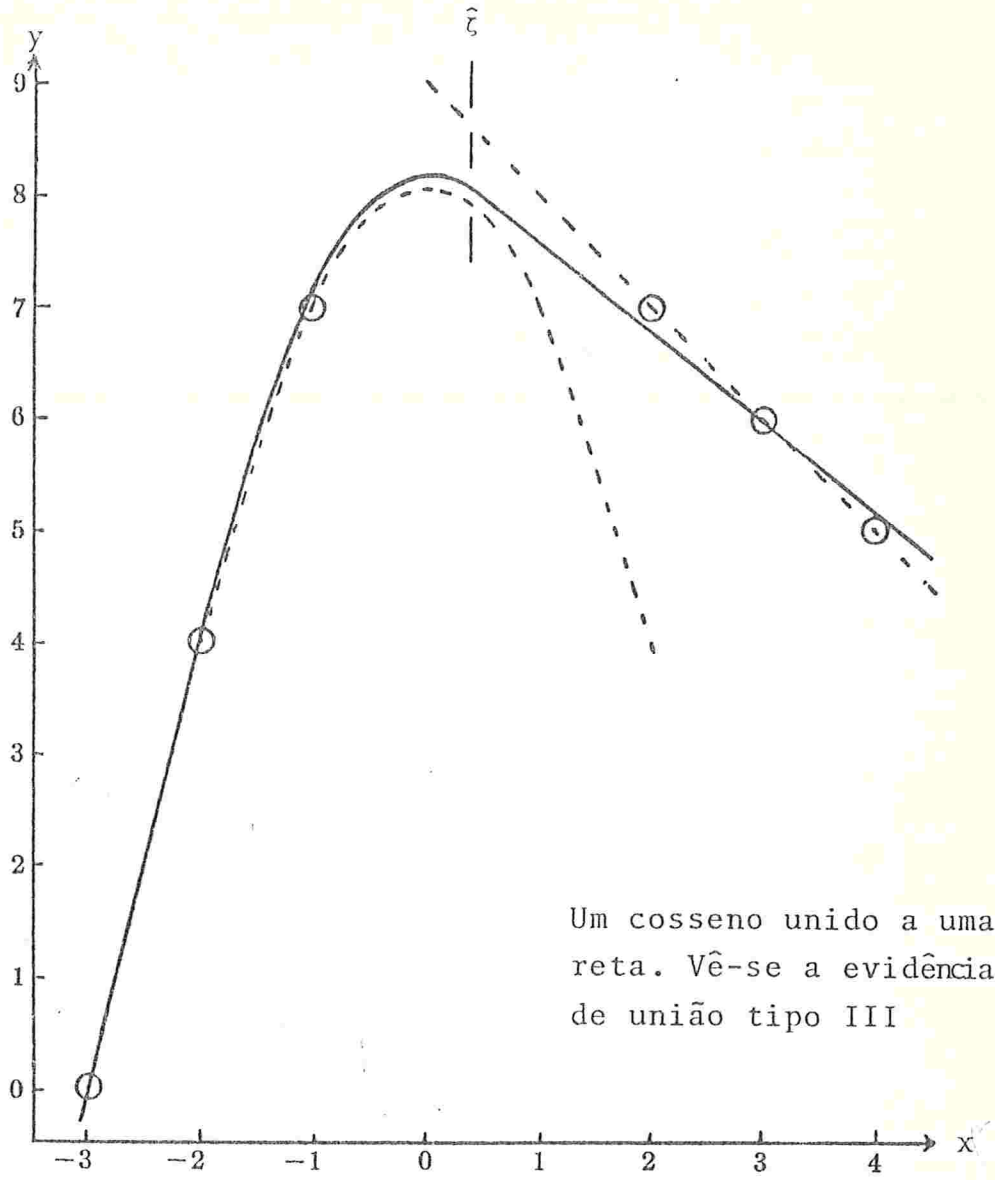
$$f_2(x, \underline{\beta}) = \beta_{11} \left[\cos \frac{\pi \zeta}{6} + \frac{\pi \zeta}{6} \operatorname{sen} \frac{\pi \zeta}{6} - \left(\frac{\pi}{6} \operatorname{sen} \frac{\pi \zeta}{6} \right) x \right] \quad \zeta \leq x \leq 4$$

A solução de mínimos quadrados é

$$\hat{\zeta} = 0,358 \quad \hat{\beta}_{11} = 8,21$$

logo

$$\hat{\beta}_{20} = 8,35 \quad \text{e} \quad \hat{\beta}_{21} = -0,80$$



Um cosseno unido a uma
reta. Vê-se a evidência
de união tipo III

Figura 6.6

A figura 6.7 nos dá a SQR como função de τ .

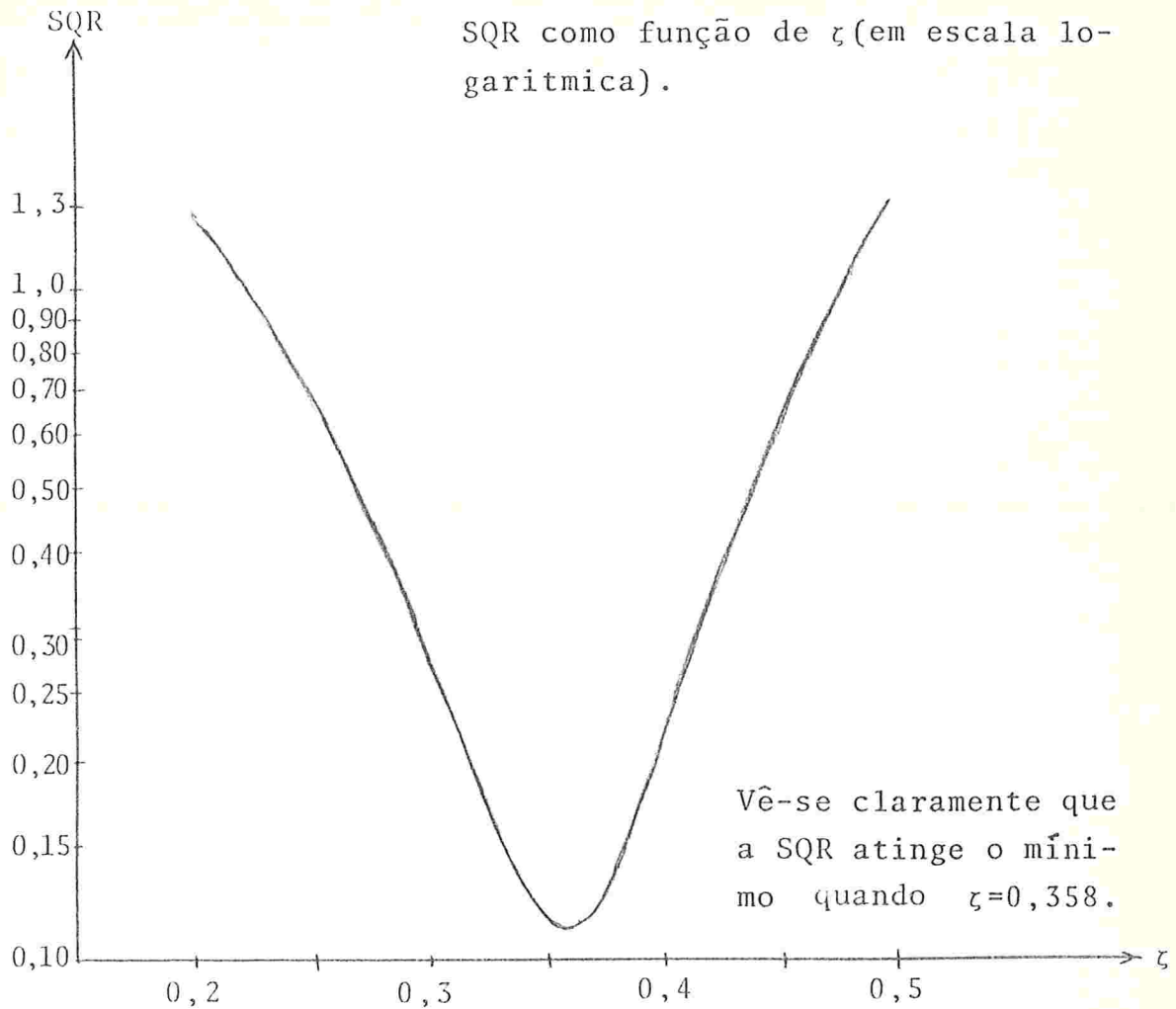


Figura 6.7

6.6 - GENERALIZAÇÃO DO METODO

Quando temos mais do que dois submodêlos procedemos de maneira análoga, tentando tôdas as possíveis combinações de tipos de união. Isto entretanto pode dificultar muito o nosso trabalho, pois apesar do número de tentativas ser finita, êle pode ser muito grande.

Porém o uso de computadores torna possível o uso do método.

CAPÍTULO VII

7.1 - CONCLUSÕES

No decorrer deste trabalho vimos algumas técnicas para detectar mudanças de modelo em um ponto desconhecido e como sempre que uma mudança ocorre devemos estimar o ponto onde isto ocorreu, vimos também técnicas para estimar este ponto de mudança.

Vamos sugerir agora uma estratégia a ser usada.

Ao tomarmos conhecimento dos dados cumpre-nos identificar primeiramente o problema. Se quisermos testar a hipótese de que os dados todos, $x_1 \dots x_T$ provêm de uma mesma população com uma distribuição $F(x, \theta)$ onde θ é (em geral) a média, suposta conhecida, porém havendo suspeitas de que as observações x_1, \dots, x_{t_0} vem desta distribuição e as observações x_{t_0+1}, \dots, x_T vem de outra distribuição $F(x, \theta')$, θ' , desconhecido e t_0 também, a estratégia a ser adotada é usar as

técnicas de Page descritas no Capítulo 2.

Cumpre-nos citar aqui alguns trabalhos mais recentes que tratam exatamente desta situação. São os trabalhos de Hinkley, datados de 1970 e 1971, onde são estudadas em particular distribuições normais onde a média muda. No seu trabalho de 1970 Hinkley deriva a distribuição assintótica do estimador de máxima verossimilhança para o índice t_0 do ponto de mudança que é dada por

$$P(\hat{t}_0 \leq t_0) \sim \{\Phi(\Delta)\}^2$$

com $\Delta = \frac{\theta_1 - \theta_0}{2\sigma}$, θ_0 e θ_1 conhecidos e estende o conceito para θ_1 desconhecido e após para θ_0 e θ_1 desconhecidos.

Ainda para a situação do estudo de mudança na média destas normais em um ponto desconhecido, Hinkley, em 1971, compara os estimadores \hat{t}_0 (estimador de máxima verossimilhança para o índice do ponto de mudança) com o \bar{t}_0 (estimador para o índice do ponto de mudança quando se usam técnicas CUSUM). A distribuição assintótica de \bar{t}_0 é aí derivada. Através de estudos teórico e empíricos, Hinkley conclui que o estimador baseado no teste CUSUM (\bar{t}_0) não é um estimador tão bom quanto \hat{t}_0 (de máxima verossimilhança) porém o último é mais difícil de computar.

Se quisermos em nosso problema detectar mudança, não em parâmetros da distribuição porém mudanças em modê

los de regressão, num ponto desconhecido, também podemos usar a técnica de Page (usaremos CUSUMS dos resíduos ordinários de mínimos quadrados) porém a técnica só se aplica se o submodelo inicial for conhecido. Nada se sabendo sobre o submodelo inicial, então devemos passar a outra estratégia, que é a desenvolvida no capítulo 3, devida a Brown et al, e que usa resíduos recursivos para detectar mudança.

E, caso o modelo, sob a hipótese nula de que não houve mudança, seja uma reta, e queiramos testar a hipótese de que o declive muda (uma só vez), em algum ponto desconhecido, e ainda caso a mudança seja pequena quando comparada com a variância, precisaremos de um teste bem sensível. Usaremos então o teste devido a Farley, que encontramos no capítulo 4, e que se mostra o teste localmente mais poderoso para este caso específico.

Neste ponto do estudo dos nossos dados, duas coisas podem ter ocorrido:

- 1) Não se detectou mudança e, ou simplesmente constatamos que todos os dados vieram de uma mesma distribuição, ou, se o caso for de ajuste de modelo, ajusta-se o modelo conveniente, único, aos dados.
- 2) Detectou-se uma mudança, num ponto desconhecido, que cumpre-nos estimar.

Foram dadas 2 técnicas: Para o caso de uma regressão linear obedecendo a 2 regimes separados podemos usar a técnica de Quant descrita no capítulo 5.

A técnica descrita no capítulo 6, devida a Hudson, mostra-se porém menos trabalhosa e não se restringe a ajuste de 2 retas. Faz porém uma outra restrição, não necessária para se aplicar a técnica de Quant. Esta restrição obriga que os 2 modelos se encontrem, isto é, não haja descontinuidade. Cabe a nós decidir o que será mais apropriado para os dados que temos em mãos. A técnica de Hudson pode ser estendida também para o caso de mais do que 2 submodelos.

Devemos ainda a Hinkley, num trabalho datado de 1969, o estudo mais aprofundado para este caso específico tratado por Hudson. Estudando um modelo de regressão em 2 fases, fez inferências sobre a intersecção, desenvolvendo um método para achar o estimador de max verossimilhança para τ , denotado por $\hat{\tau}$, no modelo

$$y_t = \begin{cases} \alpha_0 + \beta_0 x_t + \epsilon_t & t=1, \dots, t_0 \\ \alpha_1 + \beta_1 x_t + \epsilon_t & t=t_0+1, \dots, T \end{cases} \quad \text{onde}$$

$$\tau = \frac{\alpha_0 - \alpha_1}{\beta_1 - \beta_0},$$

abscissa do ponto de intersecção.

Também, como no trabalho de Hudson, requer-se a restrição de continuidade do modelo, isto é

$$x_{t_0} \leq \hat{\tau}(t_0) \leq x_{t_0+1}$$

O tratamento dado é de certa forma semelhante ao de Hudson. Neste trabalho Hinkley ainda nos deriva uma distribuição assintótica de $\hat{\tau}$ que é também uma boa aproximação para pequenas amostras, e comprova sua normalidade assintótica, já citada por vários autores.

Um dos trabalhos mais recentes de que temos notícia, sobre o estudo a que nos propuzemos, é o trabalho devido a Tore Schweder (1976) no qual é proposta uma aproximação tipo Neyman-Pearson, para o problema de detectar saltos em problemas de regressão, e que maximizem o poder médio. Tore Schweder nos demonstra que CUSUMS retroativos maximizam o poder médio. Não nos alongaremos porém mais sobre este trabalho neste nosso estudo.

Uma linha natural de continuidade deste estudo seria a procura de um "teste mais poderoso" para detectar mudanças, que se aplicasse ao caso geral de modelos de regressão e não somente ao caso específico determinado por Farley e outros. Talvez um caminho a seguir seria desenvolver a linha de Tore Schweder que já nos deu um teste que maximiza o poder médio.

Então passaríamos a procurar o "melhor" estima
dor para o(s) ponto(s) de mudança e determinar sua distri
buição, e não somente sua distribuição assintótica. Uma
vez de posse da distribuição do estimador de um ponto de
mudança, ela poderia também ser usada para testar mudan
ça, podendo surgir daí o teste mais poderoso que procura-
mos

APÊNDICE A

A.1 - Propriedades de w_r

A.1.1 Sob H_0 , w_{k+1}, \dots, w_T são v.a.i. $\sim N(0, \sigma^2)$

Prova: w_r é obviamente não viesado.

$V(w_r) = \sigma^2$ vem imediatamente de independência entre y_r e \underline{b}_{r-1} .

$$\text{Além disto } w_r = \left[\underline{\epsilon}_r - \underline{x}'_r (X'_{r-1} X_{r-1})^{-1} \sum_{j=1}^{r-1} \underline{x}_j \epsilon_j \right] \left[1 + \underline{x}'_r (X'_{r-1} X_{r-1})^{-1} \underline{x}_r \right]^{-\frac{1}{2}}$$

Visto que cada w_r é uma combinação linear de v. normais ϵ_j , os w_j 's tem distribuição normal conjunta.

Basta provar que w_{k+1}, \dots, w_T são não correlacionados que, devido à conjunto normalidade, teremos a independência. É o que veremos:

$$E \left\{ \left[\begin{array}{c} \epsilon_r - \underline{x}'_r (X'_{r-1} X_{r-1})^{-1} \sum_{j=1}^{r-1} \underline{x}_j \epsilon_j \end{array} \right] \left[\begin{array}{c} \epsilon_s - \underline{x}'_s (X'_{s-1} X_{s-1})^{-1} \sum_{j=1}^{s-1} \underline{x}_j \epsilon_j \end{array} \right] \right\} =$$

$$= \sigma^2 \left[0 - \underline{x}'_s (X'_{s-1} X_{s-1})^{-1} \underline{x}_r + \underline{x}'_r (X'_{r-1} X_{r-1})^{-1} (X_{r-1} X_{r-1}) (X'_{s-1} X_{s-1})^{-1} \underline{x}'_s \right] =$$

$$= 0 \quad (r < s)$$

$$A.1.2 \quad (X'_r X_r)^{-1} = (X'_{r-1} X_{r-1})^{-1} - \frac{(X'_{r-1} X_{r-1})^{-1} \underline{x}_r \underline{x}'_r (X'_{r-1} X_{r-1})^{-1}}{1 + \underline{x}'_r (X'_{r-1} X_{r-1})^{-1} \underline{x}_r}$$

esta relação é usada para evitar inverter a matriz $X'_r X_r$ diretamente em cada estágio dos cálculos. Prova-se multiplicando os membros da esquerda e da direita por $X'_r X_r$ e por $X'_{r-1} X_{r-1} + \underline{x}_r \underline{x}'_r = X'_r X_r$, respectivamente.

$$A.1.3 \quad \underline{b}_r = \underline{b}_{r-1} + (X'_r X_r)^{-1} \underline{x}_r (y_r - \underline{x}'_r \underline{b}_{r-1})$$

Desde que \underline{b}_r é o estimador de mínimos quadrados, ele satisfaz

$$X'_r X_r \underline{b}_r = X'_r y_r = X'_{r-1} y_{r-1} + \underline{x}_r y_r = X'_{r-1} X_{r-1} \underline{b}_{r-1} + \underline{x}_r y_r =$$

$$= X'_r X_r \underline{b}_{r-1} + \underline{x}_r (y_r - \underline{x}'_r \underline{b}_{r-1})$$

$$A.1.4 \quad S_r = S_{r-1} + w_r^2, \quad r = r+1, \dots, T$$

Prova

$$S_r = (y_r - X_r b_r)' (y_r - X_r b_r) = (y_r - X_r b_{r-1})' (y_r - X_r b_{r-1}) - (b_r - b_{r-1})'$$

$$\cdot X_r' X_r (b_r - b_{r-1}) = S_{r-1} + (y_r - x_r' b_{r-1})^2 - x_r' (X_r' X_r)^{-1} x_r (y_r - x_r' b_{r-1})^2$$

que nos dá a relação pedida, usando $(X_r' X_r)^{-1}$ de A.1.2

A.2 - A Determinação do Par de Retas

Devemos escolher o par de retas simétricas, acima e abaixo de $W_T=0$, de forma que a probabilidade de cruzar uma ou ambas as retas seja α , pré determinado. A forma de escolher as 2 linhas é decidida em 2 estágios.

O desvio padrão de Z_t é $\sqrt{t-k}$. Se quisermos encontrar uma curva tal que, sob H_0 , a probabilidade de que o caminho fique situado abaixo dela, (sem cruzá-la), para qualquer ponto entre $t=k$ e $t=T$, seja constante, temos que procurar curvas da forma $\pm \lambda \sqrt{t-k}$, com λ constante. Mas queremos nos limitar a retas e então esta probabilidade de cruzamento não pode ser constante para todo t e o procedimento adotado é escolher a família de linhas tangentes às curvas $\pm \lambda \sqrt{t-k}$ nos pontos a meio caminho entre $t=k$ e $t=T$. Isto nos leva à família de pares de retas pelos pontos $\{k; \pm a \sqrt{T-k}\}$; $\{T; \pm 3a \sqrt{T-k}\}$ onde a é o parâmetro a ser determinado, a fim de nos dar nível α .

Para qualquer linha dada nesta família a probabilidade de que o ponto (r, W_r) fique fora das linhas é máxima para r no meio do caminho entre $r=k$ e $r=T$. O que queremos é encontrar um membro desta família tal que a probabilidade de que o caminho amostral Z_t a cruze seja $\alpha/2$.

Para tal vamos tomar o trabalho de Durbin (1971).

Seja $W(t)=0$, $0 \leq t \leq \infty$ um processo de movimento Browniano, isto é, um processo normal com média zero, tendo $P(W(0)=0)=1$ e $E\{W(t_1), W(t_2)\} = \min(t_1, t_2)$. Sejam $y_1(t)$ e $y_2(t)$ os 2 limites $0 \leq t \leq T \leq \infty$.

Neste estudo de Durbin, a probabilidade de cruzar os limites é determinada somente para limites lineares. Outros casos especiais não lineares, foram tratados por Anderson e Darling (1952) Darling e Siegert (1953), Breimar (1967), Daniels (1969) e outros.

Para o caso de retas, então, vamos transcrever aqui 3 Lemas demonstrados por Durbin, os 2 primeiros levando ao 3º, que é o que nos interessa.

Seja $y=d+ct$ a reta limite onde $d>0$ e $0 \leq t \leq t_1$.

Lema 1 - A probabilidade condicional que $W(t)$ cruze a reta $y=d+ct$, dado que $W(t_1)=W_1 < d+ct_1$ é dada por

$$\psi(d, c, t_1, W_1) = \exp\{-2d(d+ct_1 - W_1)/t_1\}$$

Lema 2 - A probabilidade que $W(t)$ cruze $y=d+ct$ e $W_2 \leq W(t_2) \leq W_2 + dW_2$, onde $t_1 \leq t_2$ $dW \rightarrow 0$ é $f(d, c, t_1, t_2, W_2) dW$

onde $f(d, c, t_1, t_2, W_2) = (2\pi t_2)^{-\frac{1}{2}} \exp\{-\frac{1}{2} \frac{W_2^2}{t_2}\} \phi(z_1) + \exp\{-2dc\} (2\pi t_2)^{-\frac{1}{2}} \exp\{-\frac{1}{2} (W_2 - 2d)^2 t_2\} \phi(z_2)$, onde $\phi(z) = \int_z^\infty (2\pi)^{-\frac{1}{2}} \exp\{-\frac{1}{2} \mu^2\} d\mu$

$$z_1 = \left\{ \frac{t_2}{t_1(t_2 - t_1)} \right\}^{\frac{1}{2}} \left(d + ct_1 - \frac{t_1 W_2}{t_2} \right) e$$

$$z_2 = \left\{ \frac{t_2}{t_1(t_2 - t_1)} \right\}^{\frac{1}{2}} \left(d - ct_1 + \frac{t_1(W_2 - 2d)}{t_2} \right)$$

Lema 3 - A probabilidade que $W(t)$ cruze a reta $y=d+ct$ é da
da por

$$H(d, c, t_1) = \phi\left\{t_1^{-\frac{1}{2}}(d+ct_1)\right\} + \exp\{-2dc\} \phi\left\{t_1^{-\frac{1}{2}}(d-ct_1)\right\}.$$

Com $t_1 = T-k$; $d = a\sqrt{T-k}$ e $c = \frac{2a}{\sqrt{T-k}}$, pois o nosso par
de retas simétricas foi escrito na forma

$$\phi\{3a\} + \exp\{(-4a^2)\} \{1 - \phi(a)\} = 0$$

Fazendo $\phi(3a) + \exp(-4a^2)(1 - \phi(a)) = \frac{1}{2} \alpha$, resolve-
mos, e determinamos o valor de a conveniente, co
mo os da Tabela 3.1.

APÊNDICE B

B.1 - Vamos provar que $E_1(\underline{S}) = \delta \underline{\theta}'(I-A)\underline{\theta}$

$$\sigma^{-2}V_0(\underline{S}) = \underline{\theta}'(I-A)\underline{\theta} \quad \text{e} \quad \sigma^{-2}V_1(\underline{S}) = \underline{\theta}'(I-A)\underline{\theta} + 0\left(\frac{\delta^2}{\sigma}\right)$$

Com as relações dadas e com as relações conhecidas (Searle), temos, usando a expansão de λ em torno de δ e as definições de $\hat{\alpha}$ e $\hat{\beta}$.

$$E_1(\hat{\beta}) = \beta + \delta \frac{\underline{x}'\underline{\theta} - T\bar{x}\bar{\theta}}{\underline{x}'\underline{x} - T\bar{x}^2}$$

$$E_1(\hat{\alpha}) = \alpha + \delta \left(\bar{\theta} - \frac{\underline{x}'\underline{\theta} - T\bar{x}\bar{\theta}}{\underline{x}'\underline{x} - T\bar{x}^2} \bar{x} \right) \quad \text{onde} \quad \bar{\theta} = \frac{1}{T}(\underline{\theta}'\underline{1})$$

Então

$$E_1(\underline{y} - \hat{\alpha}\underline{1} - \hat{\beta}\underline{x}) = \delta(I-A)\underline{\theta}$$

A matriz de covariância dos resíduos é

$$\sigma^{-2} \text{Cov}_0 (\underline{y} - \hat{\alpha} \underline{1} - \hat{\beta} \underline{x}) = \underline{I} - \underline{A} \quad \text{e}$$

$$\sigma^{-2} \text{Cov}_1 (\underline{y} - \hat{\alpha} \underline{1} - \hat{\beta} \underline{x}) = \underline{I} - \underline{A} + \left(\frac{\delta}{\sigma}\right)^2 (\underline{I} - \underline{A}) \underline{B} (\underline{I} - \underline{A})$$

onde

$$\underline{B} = \frac{1}{T} \sum_{s=1}^T \underline{z}(s) \underline{z}(s)' - \underline{\theta} \underline{\theta}'$$

Aplicando na fórmula de \bar{s}

$$E_1(\bar{S}) = \delta \underline{\theta}' (\underline{I} - \underline{A}) \underline{\theta}$$

$$\sigma^{-2} V_0(\bar{S}) = \underline{\theta}' (\underline{I} - \underline{A}) \underline{\theta}$$

$$\sigma^{-2} V_1(\bar{S}) = \underline{\theta}' (\underline{I} - \underline{A}) \underline{\theta} + \left(\frac{\delta}{\sigma}\right)^2 \left\{ \frac{1}{T} \sum_{s=1}^T \left[\underline{\theta}' (\underline{I} - \underline{A}) \underline{z}(s) \right]^2 - \left[\underline{\theta}' (\underline{I} - \underline{A}) \underline{\theta} \right]^2 \right\} = \sigma^{-2} V_0(\bar{S}) + O\left(\frac{\delta^2}{\sigma^2}\right)$$

APÊNDICE C

C.1 - 1º Teorema

Se i) $x_{t_0} < \hat{\zeta} < x_{t_0+1}$

ii) $g'(\hat{\zeta}) \neq 0$

então iii) $\hat{\beta}_j = \beta_j^*(t_0)$ e $\hat{f}_j(x) = f_j^*(x) \quad j=1,2$

iv) $\hat{\zeta} = \zeta^*(t_0)$

Prova:

Seja y_j o vetor das variáveis dependentes e F_j a matriz planejamento (designação) para $j=1,2$. Façamos

$$S = (y_1 - F_1 \beta_1)' (y_1 - F_1 \beta_1) + (y_2 - F_2 \beta_2)' (y_2 - F_2 \beta_2) + \\ + 2\lambda [f_1(\zeta, \beta_1) - f_2(\zeta, \beta_2)]$$

onde λ é um multiplicador indeterminado. Se minimizarmos S com respeito a β_1 , β_2 e ζ obteremos o ajuste de mínimos qua

drados para o modelo global sujeito a $f_1(\hat{\zeta}, \hat{\beta}_1) = f_2(\hat{\zeta}, \hat{\beta}_2)$.
(O que fizemos foi usar a técnica dos multiplicadores de Lagrange).

Para minimizar, derivamos parcialmente em relação a β_1 , β_2 , ζ e λ obtendo as equações que seguem:

$$\left. \begin{aligned} \frac{1}{2} \frac{\partial S}{\partial \beta_1} &= (F_1' F_1) \hat{\beta}_1 - F_1' Y_1 + \lambda \frac{\partial f_1(\hat{\zeta}, \hat{\beta}_1)}{\partial \beta_1} = 0 \\ \frac{1}{2} \frac{\partial S}{\partial \beta_2} &= (F_2' F_2) \hat{\beta}_2 - F_2' Y_2 - \lambda \frac{\partial f_2(\hat{\zeta}, \hat{\beta}_2)}{\partial \beta_2} = 0 \end{aligned} \right\} \begin{array}{l} \text{equações} \\ \text{vetoriais} \end{array}$$

e

$$\left. \begin{aligned} \frac{1}{2} \frac{\partial S}{\partial \zeta} &= \lambda g'(\hat{\zeta}) = 0 \quad (\text{C.1}) \quad \text{desde que } \frac{\partial S}{\partial \zeta} \text{ em } \hat{\zeta} \\ \frac{1}{2} \frac{\partial S}{\partial \lambda} &= f_1(\hat{\zeta}, \hat{\beta}_1) - f_2(\hat{\zeta}, \hat{\beta}_2) = 0 \end{aligned} \right\} \begin{array}{l} \text{equações} \\ \text{escalares} \end{array}$$

ora $\frac{\partial S}{\partial \zeta}$ existe em ζ pois da condição i) a variação de S com respeito a ζ depende somente do 3º termo em S, na vizinhança de ζ . Desde que tanto f_1 quanto f_2 são diferenciáveis com respeito a ζ , concluímos que $\frac{\partial S}{\partial \zeta}$ existe em $\zeta = \hat{\zeta}$. De ii) e de C.1 concluímos que $\lambda = 0$. Então

$$\hat{\beta}_j = (F_j' F_j)^{-1} F_j' Y_j \quad (j=1,2)$$

que é o conhecido estimador de mínimos quadrados.

A prova falha se a união é do tipo III pois $g'(\hat{\zeta}) = 0$ e $\lambda \neq 0$, e também se a união é do tipo II, pois S será então uma função não diferenciável de ζ , no ponto $\hat{\zeta}$.

O teorema apresenta um corolário. Se $\exists \zeta^*(t_0)$ tal que $x_{t_0} < \zeta^*(t_0) < x_{t_0+1}$ então não é possível que i e ii) do Teorema 1º sejam ambos verdadeiros.

Prova-se por redução ao absurdo.

C.2 - Ajuste de duas curvas condicionados a se encontrar no ponto $\bar{x} = \zeta(\text{dado } \zeta)$

O modelo é

$$f(x) = \begin{cases} f_1(x, \beta_1), & x_1 \leq x \leq \zeta \\ f_2(x, \beta_2) & \zeta < x < x_T \end{cases}$$

Usando a notação tradicional, os estimadores de β_1 e β_2 , não condicionados são

$$\beta_1^* = (X_1' X_1)^{-1} X_1' Y_1 = C_1^{-1} X_1' Y_1$$

$$\beta_2^* = (X_2' X_2)^{-1} X_2' Y_2 = C_2^{-1} X_2' Y_2$$

A SQR não condicionada é $SQR_1^* + SQR_2^*$ onde

$$SQR_j^* = Y_j' Y_j - Y_j' X_j \beta_j^* ; \quad j=1,2$$

A condição de que as duas curvas se unem em ζ é

$$f_1(\zeta; \underline{\beta}_1) = f_2(\zeta; \underline{\beta}_2)$$

Como f_1 e f_2 são lineares em $\underline{\beta}_1$ e $\underline{\beta}_2$ respectivamente a condição é também linear nêles e pode ser escrita

$$(\underline{\beta}'_1, \underline{\beta}'_2) \cdot \underline{q} = 0.$$

Como os estimadores não condicionados não satisfazem esta equação, temos

$$(\underline{\beta}^{*'}_1, \underline{\beta}^{*'}_2) \underline{q} = s$$

Precisamos ainda de $C^{-1}q$ e $q'C^{-1}q = p$ onde

$$C^{-1} = \begin{pmatrix} C_1^{-1} & 0 \\ 0 & C_2^{-1} \end{pmatrix}$$

Segundo Plackett (1950), encontraremos, usando multiplicadores de Lagrange, que os estimadores de mínimos quadrados condicionados são

$$\begin{pmatrix} \hat{\underline{\beta}}_1 \\ \hat{\underline{\beta}}_2 \end{pmatrix} = \begin{pmatrix} \beta_1^* \\ \beta_2^* \end{pmatrix} - \frac{s}{p} C^{-1} q$$

A nova SQR é dada por

$$(SQR_1^* + SQR_2^* + \frac{s^2}{p})$$

Com um exemplo esta dedução se tornará clara. Sejam duas retas que se unem em $\zeta = x_{t_0}$

$$f_1(x, \underline{\beta}_1) = \beta_{10} + \beta_{11}x \quad x_1 < x < x_{t_0}$$

$$f_2(x, \underline{\beta}_2) = \beta_{20} + \beta_{21}x \quad x_{t_0} < x < x_T$$

Então

$$X_1 = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & x_{t_0} \end{pmatrix} \quad \text{e} \quad X_2 = \begin{pmatrix} 1 & x_{t_0+1} \\ 1 & x_{t_0+2} \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & x_T \end{pmatrix} \quad \begin{array}{l} \text{A colocação de } x_{t_0} \\ \text{na partição esquer} \\ \text{da e totalmente ar} \\ \text{bitrária} \end{array}$$

$$C_1 = \begin{pmatrix} t_0 & \sum_{j=1}^{t_0} x_j \\ t_0 & \sum_{j=1}^{t_0} x_j^2 \end{pmatrix} \quad C_2 = \begin{pmatrix} T-t_0 & \sum_{j=t_0+1}^T x_j \\ T & \sum_{j=t_0+1}^T x_j^2 \end{pmatrix}$$

Então

$$C_1^{-1} = \frac{\begin{pmatrix} \sum_{j=1}^{t_0} x_j^2 - \frac{t_0}{j=1} \sum_{j=1}^{t_0} x_j & \\ - \frac{t_0}{j=1} \sum_{j=1}^{t_0} x_j & t_0 \end{pmatrix}}{t_0 \sum_{j=1}^{t_0} x_j^2 - \left(\sum_{j=1}^{t_0} x_j \right)^2}$$

$$C_2^{-1} = \frac{\begin{pmatrix} T & \sum_{j=t_0+1}^T x_j^2 - \frac{T}{j=t_0+1} \sum_{j=t_0+1}^T x_j \\ - \sum_{j=t_0+1}^T x_j & T-t_0 \end{pmatrix}}{(T-t_0) \sum_{j=t_0+1}^T x_j^2 - \left(\sum_{j=t_0+1}^T x_j \right)^2}$$

Também

$$X_1' Y_1 = \begin{pmatrix} t_0 \\ \sum_{j=1}^{t_0} y_j \\ t_0 \\ \sum_{j=1}^{t_0} x_j y_j \end{pmatrix} \quad e \quad X_2' Y_2 = \begin{pmatrix} T \\ \sum_{j=t_0+1}^T y_j \\ T \\ \sum_{j=t_0+1}^T x_j y_j \end{pmatrix}$$

Finalmente, a restrição é

$$\beta_{10} + \beta_{11} \zeta = \beta_{20} + \beta_{21} \zeta$$

então $q' = (1, \zeta, -1, -\zeta)$

Os cálculos delineados anteriormente podem ser efetuados então.

C.3 - 2º Teorema

a) Se $x_{t_0} < \zeta^*(t_0) < x_{t_0+1}$ isto é $R(t_0) < \infty$ (encontram-se no lugar certo)

Então $S(t_0) \geq R(t_0)$ e $S(t_0+1) \geq R(t_0)$

Demonstração

O ajuste de mínimos quadrados não condicionados tem SQR não maior que a dos mínimos quadrados condicionados. Então, se se condiciona $\hat{\zeta}$ a ser $\hat{\zeta} = x_{t_0}$ ou $\hat{\zeta} = x_{t_0+1}$ encontraremos

$$S(t_0) \geq SQR^*(t_0) \quad e \quad S(t_0+1) \geq SQR^*(t_0)$$

(lembramos que $SQR^* = SQR_1^* + SQR_2^*$)

Já vimos que se $R(t_0) < \infty$ tínhamos $R(t_0) = SQR^*(t_0)$.

Caso $R(t_0) = \infty \neq SQR^*(t_0) = SQR_1^*(t_0) + SQR_2^*(t_0)$

Vimos no Apêndice C.C2 que a SQR condicionada é dado por $SQR_1^* + SQR_2^* + \frac{S^2}{p}$

b) Se

$$SQR^*(t_0) > \underset{\text{todo } j}{\text{Min}} \{R(j)\}$$

então

$$S(t_0) > \underset{\text{todo } j}{\text{Min}} \{R(j)\} \quad \text{e} \quad S(t_{0+1}) > \underset{\text{todo } j}{\text{Min}} \{R(j)\}$$

Demonstração

$$\text{Como no ítem a } S(\ell) \geq SQR^*(\ell) \geq \underset{\text{todo } j}{\text{Min}} \{R(j)\}$$

$$\ell = t_0, t_{0+1}$$

c) Se i) $S(k)$ foi determinado para certo k

$$\text{ii) } SQR^*(t_0) \geq \underset{\text{aqueles } k}{\text{Min}} \{S(k)\}$$

então

$$S(t_0) > \underset{\text{aqueles } k}{\text{Min}} \{S(k)\} \quad \text{e} \quad S(t_{0+1}) > \underset{\text{aqueles } k}{\text{Min}} \{S(k)\}$$

Demonstração

De maneira totalmente análoga

$$S(\ell) \geq SQR^*(\ell) \geq \underset{\text{aqueles } k}{\text{Min}} \{S(k)\} \quad \text{para } \ell = t_0, t_{0+1}$$

sempre baseado no fato de que a SQR não condicionada é não maior que a SQR condicionado.

C.4 - 3º Teorema

a) Se i) $x_{t_0} \leq \zeta^*(t_0) \leq x_{t_0+1}$

ii) $g'[\zeta^*(t_0)] \neq 0$

então não é possível que

iii) $x_{t_0} < \hat{\zeta} < x_{t_0+1}$

iv) $g'(\hat{\zeta}) = 0$

b) Se

$$SQR^*(t_0) \geq \underset{j}{\text{Min}} \{R(j), S(j)\}$$

aplicam-se as mesmas conclusões anteriores.

Prova - Aplicando-se as restrições

$$f_1(\hat{\zeta}, \hat{\beta}_1) = f_2(\hat{\zeta}, \hat{\beta}_2) \quad e$$

$$f'_1(\hat{\zeta}, \hat{\beta}_1) = f'_2(\hat{\zeta}, \hat{\beta}_2)$$

não diminuirá a SQR obtida nos passos anteriores

REFERÊNCIAS BIBLIOGRÁFICAS

1. Brown, R.L. and Durbin (1968) - Methods of Investigation whether a regression relationship is constant over time. *Ann. Meth. Statist*, 39, pg 1731-1743.
2. Brown, R.L., Durbin J. and Evans J.M.(1975) - Techniques for testing the constancy of Regression Relationships over Time, *J.R. Statist. Soc.*, 37, pg 149-163.
3. Capon J. (1961) - "On the Asymptotic Efficiency of Locally Optimum Detectors" - *IEEE Transaction on Information Theory* I T, 7, pg 67-71.
4. Draper N.R. (1966) - Applied Regression Analysis.
5. Durbin J. (1969) - Tests for serial correlation in regression analysis based on the periodogram of least squares residuals - *Biometrika*, 56, pg 1-15.
6. Durbin J. (1971) - Boundary - Crossing Probabilities for the Brownian Motion and Poisson Processes and techniques for computing the Power of the Kolmogorov-Smirnov test - *J. Appl. Prob.*, 8, pg 431-453.
7. Farley, J.U. and Hinich, M. J. (1970) - A test for a Shifting Slope Coefficient in a Linear Model - *J. Amer. Statist. Ass.*, 65 n° 331 - 1320-1329.
8. Hinkley D.V. (1969) Inference about the Intersection in two phase Regression - *Biometrika*, 56, 3 - pg 495-504.

9. Hinkley D.V. (1970) - Inference about the Change-point in a sequence of random variables - *Biometrika*, 57,1 pg 1, 17.
10. Hinkley D.V. (1971) - Inference about the change-point from cumulative sum testes - *Biometrika*, 58,3 - pg 509-523.
11. Hudson D.J. (1960) - Fitting segmented curves whose join points have to be estimated - *J. Am. Statist. Ass.*, 61, pg 1097-1129.
12. Page E.S. (1954) - Continuous inspection schemes - *Biometrika*, 41, pg 100-114.
13. Page E.S. (1955) - A test for a change in a parameter occurring at an unknown point - *Biometrika*, 42, pg 523-527.
14. Page E.S. (1957) - On problems in which a change in a Parameter occurs at an unknown point - *Biometrika* , 44 - pg 248-252.
15. Plackett R.L. (1950) - Some theorems in least squares *Biometrika*, 37 - pg 149-157.
16. Plackett R.L. (1960) - Principles of Regression Analysis *Oxford - at the Charendon Press.*
17. Quant R.E. (1958) - The estimation of the parameters of a linear regression system obeying two separate re-rimes - *J. Amer. Statist. Ass.*, 53, pg 873-880.

18. Quant R.E. (1960) - Tests of the Hypotesis that a Linear Regression System obeys two separate Regimes - *J.Amer Statist. Ass.*, 55 - pg 324-330.
19. Searle, Linear Models.
20. Theil, H. (1965) - The analysis of disturbances in Regression Analysis - *J. Amer. Statist. Ass.*, 60 - pg 1067-1079.
21. Tore Schweder (1976) - Some "Optimal" Methods to Detect - Structural Shift or Outliers in Regression - *J. Amer. Statist. Ass.*, 71 - pg 491-501.

ERRATA

Pgs	linha	Correção
14	5º	$m < h$ em substituição a $1 \leq s \leq r$ para $1 \leq s < r$
14	14º	$P_{-r+1} = P P_r$ acrescente-se os " ~ "
15	2º, 6º, 15º	$p_0 ; p_r = P^r p_0 ; p_T = P_2^{T-k} P_1^k p_0$ acrescente-se os " ~ "
38	2º	$\underline{\mu} \underline{z}(s)$ acrescente-se os " ~ "
40	16º	após o 4º ponto em substituição a 3º ponto
41	6º	$\underline{\theta}' (I-A) \underline{\theta} = 0,951$ em substituição aos resultados
41	8º	$\underline{S} = \underline{\theta}' (I-A) \underline{\gamma} = -1,461$
41	10º	$z_{\alpha/2} \sigma \sqrt{\underline{\theta}' (I-A) \underline{\theta}} = 1,645 \cdot 0,03 \cdot \sqrt{0,951} = 0,048$
43	6º	$\sqrt{2\pi} \sigma_1$ e $\sqrt{2\pi} \sigma_2$ em substituição a $\sqrt{2\pi\sigma_1}$ e $\sqrt{2\pi\sigma_2}$
51	13º	$f_2(x_t, \beta_2)]^2$ em substituição a $f_2(x_x, \beta_2)]^2$
62	15º	$\zeta = 2$ $\zeta = 1$ em substituição a $\alpha = 2$ e $\alpha = 1$
65	14º	$\hat{\beta}_2^*$ acrescente-se o " ~ "
68	11º $\pi/6$ $\text{sen } \pi\zeta/6$ em substituição a $\pi\zeta/6$ $\text{sen } \pi\zeta/6$
78	2º	... $(X'_{r-1} X_{r-1}) (X'_{s-1} X_{s-1})^{-1} \underline{x}'_s$ acrescente-se o (')
79	1º	... $-(\underline{b}'_r - \underline{b}'_{r-1})'$ "
79	2º	... = $S_{r-1} + (y_r - \underline{x}'_r \underline{b}_{r-1})^2$ - ... acrescente-se o índice (r)
82	7º	após: foi escrito na forma acrescente-se $y = \pm a \sqrt{T-k} \pm \frac{2a}{\sqrt{T-k}}(t-k)$ temos que esta probabilidade é
82	8º	$\phi(3a) + \exp\{-4a^2\}\{1 - \phi(a)\}$ corrigir os () e suprimir = 0
84	4º	\underline{S} em substituição a \tilde{s}
87	18º	$SQR_j^* = \underline{y}'_j \underline{y}_j - \underline{y}'_j X_j \beta_j^*$; $j = 1, 2$ em subst. aos índices "1"
88	8º	$C^{-1}q$ e $q'C^{-1}q$ acrescente-se os " ~ "
90	5º	q' acrescente-se os " ~ "
93	2º	European Meeting, 1968. Selected Statistical Papers, I pg 37 - 45. Amsterdam. Mathematisch Centrum em subst. a Ann. Math. Statist.
95	4º	Searle (1971) Linear Models.