

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Curso de Estatística

Rafael Rocha de Oliveira Garcia

**Inferência sobre modelos de regressão linear com pontos de mudança e erros
normais autoregressivos**

Juiz de Fora
2018

Rafael Rocha de Oliveira Garcia

Inferência sobre modelos de regressão linear com pontos de mudança e erros normais autoregressivos

Monografia apresentada ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do grau de Bacharel em Estatística.

Orientadora: Camila Borelli Zeller

Juiz de Fora

2018

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Rocha de Oliveira Garcia, Rafael.

Inferência sobre modelos de regressão linear com pontos de mudança e
erros normais autoregressivos / Rafael Rocha de Oliveira Garcia. – 2018.
81 f. : il.

Orientadora: Camila Borelli Zeller

Trabalho de Conclusão de Curso – Universidade Federal de Juiz de Fora,
Instituto de Ciências Exatas. Curso de Estatística, 2018.

1. Modelos de regressão. 2. Modelos autoregressivos. 3. Pontos de
mudança. I. Borelli Zeller, Camila, orient. II. Título.

Rafael Rocha de Oliveira Garcia

Inferência sobre modelos de regressão linear com pontos de mudança e erros normais autoregressivos

Monografia apresentada ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do grau de Bacharel em Estatística.

Aprovada em:

BANCA EXAMINADORA

Prof^a. Dra. Camila Borelli Zeller - Orientadora
Universidade Federal de Juiz de Fora

Professor Dr. Lupércio França Bessegato
Universidade Federal de Juiz de Fora

Professor Dr. Michel Helcias Montoril
Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Meus agradecimentos são dirigidos a todos aqueles que me apoiaram ao longo desta jornada.

Agradeço ao meu pai e minha mãe, Alen e Meirilane, já que sem o suporte, carinho e ajuda deles nada dos últimos 5 anos seria possível. Agradeço também à minha irmã, Bárbara, responsável por vários momentos de descontração que ajudaram a respirar entre tantas provas.

Um agradecimento especial aos meus amigos que estiveram ao meu lado desde o início e até mesmo àqueles que se juntaram a mim no meio do caminho. Mas um agradecimento ainda maior à Marcela, que me acompanha desde o Ensino Médio e sempre esteve ao meu lado (mesmo durante aquele ano fora) e ao Robson pela grande amizade (e extensas ajudas no R).

Gostaria, ainda, de agradecer à minha orientadora Camila. Uma professora de qualidade ímpar que não se encontra em qualquer lugar. Um grande obrigado por me aceitar como orientando. Agradeço também aos professores Lupércio e Michel pelos conselhos nos corredores do departamento e por aceitarem o convite para uma última avaliação.

Agradeço, também, ao Programa Institucional de Bolsas de Iniciação Científica (PIBIC) da UFJF e ao CNPq pelo financiamento do projeto.

“Nem todos os que vagueiam estão perdidos.”
(J.R.R. Tolkien, O Senhor dos Anéis, 1954)

“Inductive inference is the only process known to us by which essentially new knowledge
comes into the world.”
(Sir Ronald A. Fisher, The Design of Experiments, 1935)

RESUMO

Análise de modelos de regressão linear é uma das mais importantes técnicas estatísticas. Com ela, é possível identificar relações entre variáveis que técnicas mais visuais, às vezes, não evidenciam. Contudo, para proceder com uma análise de regressão linear, muitos pressupostos devem ser satisfeitos. Um dos mais importantes, a independência dos erros, nem sempre pode ser verificada e isso pode gerar resíduos autocorrelacionados no modelo proposto após o ajuste. São várias as técnicas para tentar eliminar essa correlação, mas pode ser de interesse do pesquisador modelar a dependência dos erros e isso pode melhorar o ajuste dos parâmetros do modelo de regressão linear. Em certas situações, pode ser de interesse do pesquisador identificar mudanças na média e/ou variância de um determinado processo que suspeita-se não estar funcionando corretamente. É neste contexto que a identificação de pontos de mudança mostra-se extremamente útil. O objetivo deste trabalho é apresentar a situação em que todos estes cenários são identificados: modelos de regressão linear, processos autoregressivos e pontos de mudança; e também apresentar estimadores para esse modelo.

Palavras-chave: Regressão linear. Autoregressivos. Ponto de mudança.

ABSTRACT

Linear regression analysis models are some of the most important techniques in Statistics. With it, it's possible to identify relationships between variables that visual techniques, sometimes, do not show. However, to proceed with linear regression analysis, too many assumptions must be satisfied. A very important one, error independence, cannot always be satisfied, which may return residuals with autocorrelation after fitting the proposed model. There are many techniques to deal with this correlation, but it can be interesting to the researcher to model the dependence of the errors which can improve the model fitting. Sometimes, the researcher is interested in identifying structural breaks in the mean and/or variance of some process which is suspected to be not working correctly. In this situation, change-point identification is extremely useful. The aim of this monograph is to present the situation in which all of these scenarios are present: linear regression models, auto-regressive process and change-point; also to present estimates to this model.

Key-words: Linear regression. Auto-regressive. Change-point.

LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama de dispersão da Força de cisalhamento versus idade do combustível.	16
Figura 2 – Diagrama de dispersão da Força de cisalhamento versus idade do combustível com reta estimada.	16
Figura 3 – Série temporal com diferentes níveis.	17
Figura 4 – Série temporal com diferença na variância.	18
Figura 5 – Diagrama de dispersão da variável BSE versus NYAMSE. Em preto a reta estimada pelo método de mínimos quadrados que não considera o ponto de mudança; em vermelho tracejado a técnica proposta em [7] que considera o ponto de mudança.	19
Figura 6 – Diagrama de dispersão da Força de cisalhamento versus idade do combustível. As linhas verticais em vermelho são as distâncias das observações x_i à reta estimada.	23
Figura 7 – Realização de uma série temporal.	29
Figura 8 – Realização de um processo AR(1).	32
Figura 9 – Função de autocorrelação teórica com $\phi_1 = 0.8897$	32
Figura 10 – Função de autocorrelação empírica com $\phi_1 = 0.8897$	33
Figura 11 – Realização de um processo AR(2).	33
Figura 12 – Função de autocorrelação teórica com $\phi_1 = 0.8897$ e $\phi_2 = -0.4858$	34
Figura 13 – Função de autocorrelação empírica com $\phi_1 = 0.8897$ e $\phi_2 = -0.4858$	35
Figura 14 – Função de autocorrelação parcial teórica com $\phi_1 = 0.8897$	36
Figura 15 – Função de autocorrelação parcial teórica com $\phi_1 = 0.8897$ e $\phi_2 = -0.4858$	37
Figura 16 – Função de autocorrelação parcial empírica com $\phi_1 = 0.8897$	37
Figura 17 – Função de autocorrelação parcial empírica com $\phi_1 = 0.8897$ e $\phi_2 = -0.4858$	38
Figura 18 – Diagrama de dispersão de <i>comsales</i> versus <i>indsales</i>	42
Figura 19 – Diagrama de dispersão de <i>comsales</i> versus <i>indsales</i> com reta estimada via MQO.	42
Figura 20 – Função de autocorrelação dos resíduos da regressão.	43
Figura 21 – Função de autocorrelação parcial dos resíduos da regressão.	43
Figura 22 – Diagrama de dispersão de <i>comsales</i> versus <i>indsales</i> com retas estimadas considerando independência dos erros e autocorrelação.	45
Figura 23 – Diagrama de dispersão da variável BSE.	47
Figura 24 – Diagrama de dispersão da variável BSE.	48
Figura 25 – Fluxograma representando o processo de decisão.	50
Figura 26 – Diagrama de dispersão da variável BSE versus NYAMSE com a reta estimada por mínimos quadrados.	51

Figura 27 – Diagrama de dispersão da variável BSE versus NYAMSE. Triângulo representa os pontos <i>antes</i> do ponto de mudança e quadrado representa os pontos <i>depois</i> da mudança.	52
Figura 28 – Função de autocorrelação dos resíduos da regressão <i>antes</i> do ponto de mudança.	53
Figura 29 – Função de autocorrelação parcial dos resíduos da regressão <i>antes</i> do ponto de mudança.	54
Figura 30 – Função de autocorrelação dos resíduos da regressão <i>depois</i> do ponto de mudança.	55
Figura 31 – Função de autocorrelação parcial dos resíduos da regressão <i>depois</i> do ponto de mudança.	55
Figura 32 – Diagrama de dispersão da variável BSE versus NYAMSE. Triângulo representa os pontos <i>antes</i> do ponto de mudança e quadrado representa os pontos <i>depois</i> da mudança. As linhas tracejadas em verde representam as retas estimadas considerando o erro autoregressivos e as linhas contínuas em vermelho representam as retas estimadas considerando o erro normal.	61
Figura 33 – Diagrama de dispersão da variável dos dados simulados com a reta real evidenciada.	62
Figura 34 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança na mediana dos dados, mudança em todos os parâmetros da regressão e nos coeficientes autoregressivos, $n = 100$	62
Figura 35 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança na mediana dos dados, mudança em todos os parâmetros da regressão e nos coeficientes autoregressivos, $n = 400$	63
Figura 36 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança no primeiro quartil dos dados, mudança em todos os parâmetros da regressão e nos coeficientes autoregressivos, $n = 100$	64
Figura 37 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança no primeiro quartil dos dados, mudança em todos os parâmetros da regressão e nos coeficientes autoregressivos, $n = 400$	65
Figura 38 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança na mediana dos dados, mesmo intercepto e mudança nos coeficientes autoregressivos, $n = 100$	65
Figura 39 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança na mediana dos dados, mesmo intercepto e mudança nos coeficientes autoregressivos, $n = 400$	66

Figura 40 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança no primeiro quartil dos dados, mesmo intercepto e mudança nos coeficientes autoregressivos, $n = 100$	66
Figura 41 – Diagrama de dispersão da variável dos dados simulados com a reta real evidenciada para o caso interceptos iguais e ponto de mudança no primeiro quartil.	69
Figura 42 – Diagrama de dispersão da variável dos dados simulados com a reta real evidenciada para o caso de retas paralelas e ponto de mudança na mediana.	69
Figura 43 – Diagrama de dispersão da variável dos dados simulados com a reta real evidenciada para o caso de retas concorrentes e ponto de mudança na mediana.	70

LISTA DE TABELAS

Tabela 1	– Dados sobre combustíveis de foguete.	15
Tabela 2	– Valores das variáveis BSE e NYAMSE.	20
Tabela 3	– Coeficientes estimados para os modelos considerando independência dos erros e autocorrelação nos dados.	44
Tabela 4	– Estimativas dos coeficientes para os modelos sem e com ponto de mudança.	51
Tabela 5	– Resultado do teste de Durbin-Watson para presença de autocorrelação nos resíduos antes e depois do ponto de mudança, respectivamente. . .	56
Tabela 6	– Estimativas dos coeficientes para os modelos sem ponto de mudança, com ponto de mudança e com ponto de mudança e erro AR(1).	60
Tabela 7	– Comparação entre simulações de modelos de regressão linear simples com ponto de mudança e erro autoregressivo AR(2). Pontos de mudança na mediana dos dados e no primeiro quartil, com variação no tamanho da amostra. Em parêntesis, o desvio padrão amostral das simulações. .	63
Tabela 8	– Comparação entre simulações de modelos de regressão linear simples com ponto de mudança e erro autoregressivo AR(2). Em parêntesis, o desvio padrão amostral das simulações. Para este caso, consideramos mesmo intercepto antes e depois do ponto de mudança.	67
Tabela 9	– Resultados para simulação com 1000 replicações de cada um dos tamanhos amostrais abaixo. Em parêntesis, os desvios padrão amostrais. . .	67
Tabela 10	– Resultados para modelos com mesmo intercepto, retas paralelas e retas concorrentes. Em parêntesis, os desvios-padrão amostrais.	68
Tabela 11	– Posições do ponto de mudança identificadas e porcentagem de ocorrências para cada um dos três casos.	70

SUMÁRIO

1	INTRODUÇÃO	13
1.1	CONTEXTUALIZAÇÃO E OBJETIVOS	13
1.2	ANÁLISE DE REGRESSÃO LINEAR	14
1.3	MODELOS AUTOREGRESSIVOS	14
1.4	PONTOS DE MUDANÇA	15
2	MODELO DE REGRESSÃO LINEAR SIMPLES E MODE- LOS AUTOREGRESSIVOS	21
2.1	MODELO DE REGRESSÃO LINEAR	21
2.1.1	Estimação do modelo de regressão linear simples	21
2.1.1.1	<i>Método de mínimos quadrados ordinários</i>	<i>22</i>
2.1.1.2	<i>Método de máxima verossimilhança</i>	<i>26</i>
2.2	MODELOS AUTOREGRESSIVOS	28
2.2.1	Modelo AR(1)	31
2.2.2	Modelo AR(2)	33
2.2.3	Função de autocorrelação parcial	34
3	MODELO DE REGRESSÃO LINEAR SIMPLES COM ER- ROS AUTOREGRESSIVOS DE ORDEM p	39
3.1	ESTIMAÇÃO	40
3.2	APLICAÇÃO	41
4	MODELO DE REGRESSÃO LINEAR SIMPLES COM PONTO DE MUDANÇA E ERRO NORMAL AUTOREGRESSIVO .	46
4.1	MODELO DE REGRESSÃO LINEAR SIMPLES COM PONTO DE MUDANÇA E ERROS INDEPENDENTES	46
4.1.1	Identificação do ponto de mudança via critério de seleção . . .	46
4.2	MODELO DE REGRESSÃO LINEAR SIMPLES COM PONTO DE MUDANÇA E ERRO NORMAL AUTOREGRESSIVO DE ORDEM (p)	52
4.2.1	Estimação do modelo de regressão linear simples com ponto de mudança e erro normal autoregressivos de ordem p	56
4.3	ESTUDO DE SIMULAÇÃO	60
4.3.1	Ponto de mudança conhecido	60
4.3.2	Ponto de mudança desconhecido	68
5	Considerações Finais	71

REFERÊNCIAS	72
APÊNDICE A – Rotinas em R para o modelo proposto . . .	74
APÊNDICE B – Rotinas para simulação	80

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO E OBJETIVOS

Modelos com pontos de mudança são de grande interesse estatístico. Uma mudança, abrupta ou suave, e sua identificação em um determinado processo pode impactar severamente nos lucros de uma empresa ou ganhos em ações na Bolsa de Valores, por exemplo. É nesse contexto de identificação que surgem os métodos para lidar com os pontos de mudança. Tais mudanças podem ocorrer em vários tipos de modelos; os exemplos mais comuns são em séries temporais e em modelos de regressão.

No contexto de séries temporais, [2] apresentam uma abordagem utilizando teste da razão de verossimilhança empírica, além de mostrar algumas propriedades assintóticas. Ainda em séries temporais, [20] traz uma análise sobre mudança de variância em processos autoregressivos de primeira ordem.

Podemos estar interessados, contudo, em identificar uma quebra estrutural em médias, mais especificamente, no contexto de análise de regressão linear, na média de uma variável aleatória Y dado uma variável não estocástica X . Isto é, estamos interessados em estudar o comportamento de Y *antes* de uma certa posição k e *depois* desta posição k . Estamos interessados em estudar essa quebra em modelos de regressão linear simples, porém, para o contexto de regressão linear múltipla, podemos encontrar em [4] abordagens para lidar com esse caso.

O objetivo deste trabalho é propor o modelo de regressão linear simples com ponto de mudança e erro autoregressivos, além de encontrar expressões para os estimadores dos parâmetros desse modelo. Com relação ao ponto de mudança, seguiremos por uma abordagem que considera critérios de informação - veja [1] e [19] - para poder identificar a posição do ponto de mudança e o melhor ajuste possível com base no princípio de critério de informação [8].

Neste capítulo, faremos uma breve apresentação dos conceitos que serão abordados durante o texto, tais como análise de regressão linear, estimação de um modelo de regressão linear simples, séries temporais e modelos autoregressivos, assim como pontos de mudança em séries temporais e em modelos de regressão.

No Capítulo 2, faremos uma apresentação formal do Modelo de Regressão Linear Simples (ou MRLS) com erro normal e dos Modelos Autoregressivos. Já no Capítulo 3, introduziremos o Modelo de Regressão Linear Simples com Erros Autoregressivos (ou MRLS-AR).

No Capítulo 4, introduziremos o Modelo de Regressão Linear Simples com Ponto de Mudança e Erro autoregressivos (ou MRLS-PM-AR), além de apresentar uma técnica

de como lidar com pontos de mudança. Mostraremos também estudos de simulação e possíveis aplicações. No Capítulo 5, faremos algumas considerações finais sobre o trabalho, além de possíveis trabalhos futuros.

Todos os cálculos, resultados, gráficos e tabelas ao longo desse texto foram obtidos com o apoio do software R[17]. Nos Apêndices A e B encontram-se as funções desenvolvidas para esse trabalho.

1.2 ANÁLISE DE REGRESSÃO LINEAR

Regressão linear é uma técnica estatística para investigar e modelar a relação entre variáveis. São muitas as aplicações: biologia, engenharias, ciências físicas e químicas, para listar algumas. De forma a introduzir o conceito de regressão linear, considere o seguinte exemplo que trata de combustível (ou propelente) de foguetes encontrado em [16].

Um motor de foguete é fabricado ao unir um combustível de ignição com um combustível de sustentação, e a força de cisalhamento entre esses combustíveis é uma importante característica de qualidade. Foram observadas vinte medidas de **força de cisalhamento** (em psi, libra-força por polegada quadrada) e de **idade do propelente** (em semanas). Suspeita-se de que a força de cisalhamento é relacionada de alguma forma com a idade do combustível do foguete. Os dados encontram-se na Tabela 1.

A Figura 1 sugere uma relação linear entre a força de cisalhamento e a idade do combustível. Poderíamos pensar em traçar uma reta para evidenciar essa relação. A Figura 2 apresenta essa reta, obtida pelos métodos de estimação que serão apresentados no Capítulo 2.

1.3 MODELOS AUTOREGRESSIVOS

Uma série temporal é um conjunto de observações gerada sequencialmente no tempo [5]. Se o conjunto é contínuo, a série é dita *contínua*; se o conjunto é discreto, a série é dita *discreta*. Suponha que temos N observações nos tempos t_1, \dots, t_N da série $z(t_1), \dots, z(t_N)$. Estas observações podem ser feitas em momentos equidistantes ou não; quando feitas em instantes igualmente espaçados podemos denotar a série simplesmente por z_1, \dots, z_N .

Uma série temporal pode ser descrita por sua média, variância e funções de autocorrelação e autocorrelação parcial. Um processo autoregressivo de ordem p é um caso particular de série temporal, sendo da forma $z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + \eta_t$, em que $(\phi_1, \dots, \phi_p)^\top$ são os parâmetros autoregressivos e $\eta_t \sim (0, \sigma^2)$ o ruído com valor esperado zero e variância σ^2 . Todos esses conceitos serão abordados no Capítulo 2.

Tabela 1 – Dados sobre combustíveis de foguete.

Observação	Força de Cisalhamento (psi)	Idade do combustível (semanas)
1	2158,70	15,50
2	1678,15	23,75
3	2316,00	8,00
4	2061,30	17,00
5	2207,50	5,50
6	1708,30	19,00
7	1784,70	24,00
8	2575,00	2,50
9	2357,90	7,50
10	2256,70	11,00
11	2165,20	13,00
12	2399,55	3,75
13	1779,80	25,00
14	2336,75	9,75
15	1765,30	22,00
16	2053,50	18,00
17	2414,40	6,00
18	2200,50	12,50
19	2654,20	2,00
20	1753,70	21,50

Fonte: Encontrada em [16].

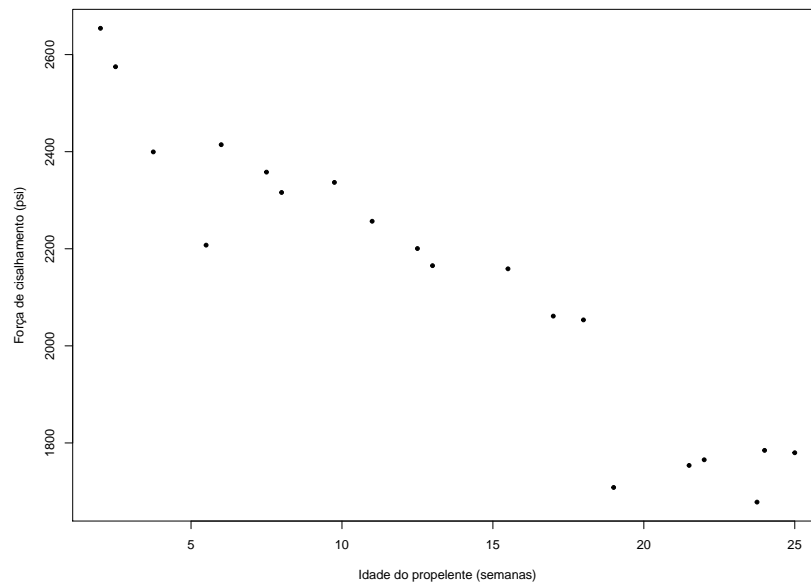
No Capítulo 3, estudaremos o caso em que o Modelo de Regressão Linear Simples é estendido para compreender Erros Autoregressivos.

1.4 PONTOS DE MUDANÇA

Pontos de mudança surgiram da necessidade de estudar mudanças, abruptas ou não, em algum tipo de processo; podem ser mudanças na média, variância ou em ambos. Em controle de qualidade de processos, por exemplo, um certo processo de produção contínua é esperado que se mantenha estável. Contudo, por diversas razões, o processo pode falhar em produzir itens com a mesma qualidade. É de interesse do fabricante, e do pesquisador, encontrar o ponto em que a qualidade dos produtos começa a se deteriorar [8]. Em [8], [7] e em [11] há uma revisão de outros trabalhos que lidaram com pontos de mudança nos mais variados contextos.

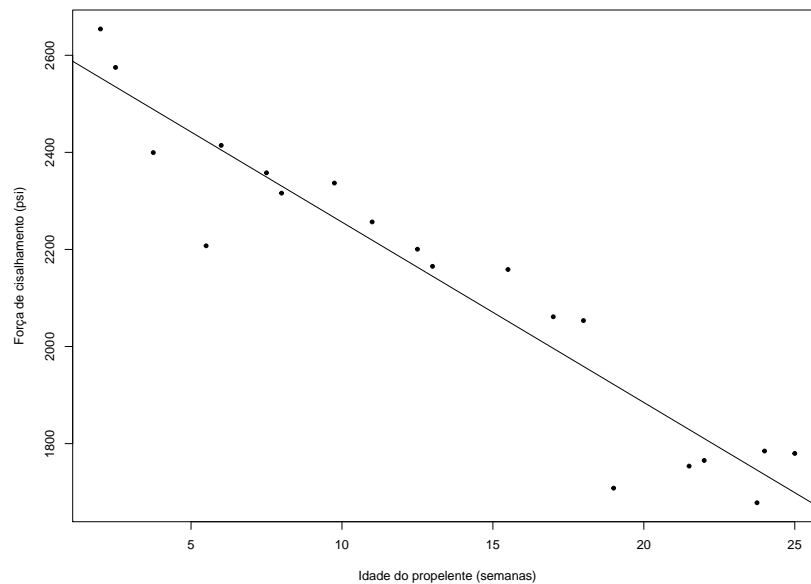
No contexto de pontos de mudança em séries temporais, temos [11] que descrevem técnicas para a identificação do ponto de mudança. Os principais métodos utilizados são o Teste da Razão de Verossimilhanças (clássico), Bayesiano (que visa a maximização da

Figura 1 – Diagrama de dispersão da Força de cisalhamento versus idade do combustível.



Fonte: Elaborado pelo autor.

Figura 2 – Diagrama de dispersão da Força de cisalhamento versus idade do combustível com reta estimada.

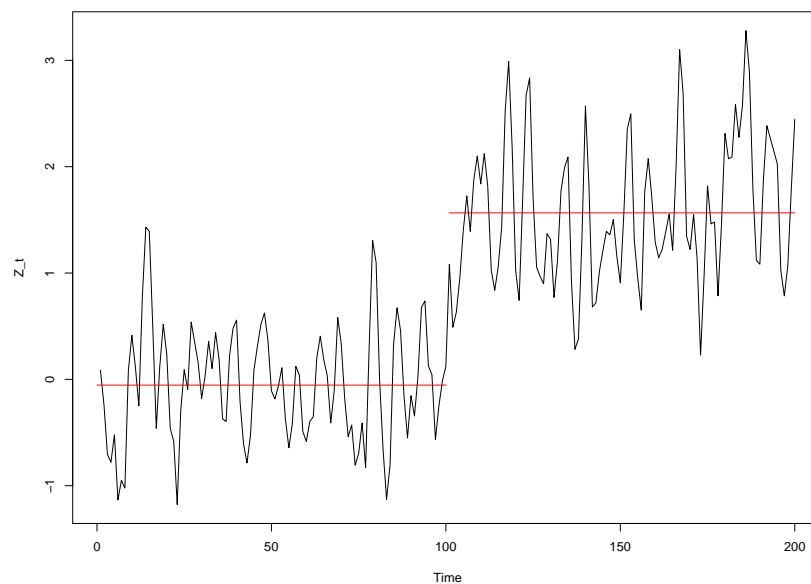


Fonte: Elaborado pelo autor.

densidade *a posteriori*) e abordagem por critérios de informação. Ao longo deste texto, iremos considerar o contexto de mudança abrupta, assim, *identificaremos* o ponto de

mudança; [8] apresentam o caso de mudança suave, em que deve ser *estimado*, além de apresentarem outras referências para lidar com este tipo de situação. Ainda em séries temporais, as mudanças podem ser no nível (ou média) da série, na variância do ruído ou mudança nos parâmetros que geram a série (modelos autoregressivos, por exemplo, que também causam mudanças na variância do processo). As Figuras 3 e 4 ilustram o caso de mudança no nível e na variância da série, respectivamente.

Figura 3 – Série temporal com diferentes níveis.

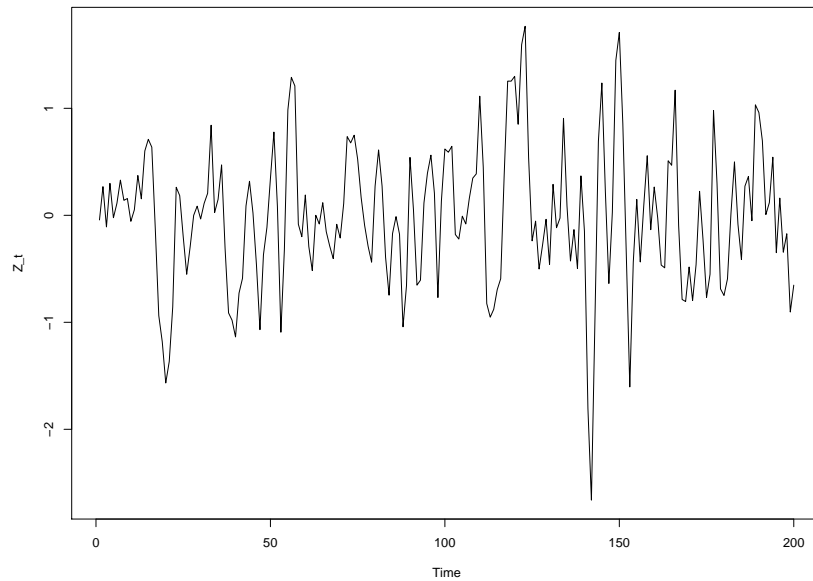


Fonte: Elaborado pelo autor.

Já no contexto de análise de regressão, [7] e [8] apresentam abordagens que utilizam critérios de seleção para a detecção do ponto de mudança. Um exemplo muito trabalhado na literatura é com os dados encontrados em [13]. Esse conjunto de dados apresenta o volume de vendas mensal (em milhões de dólares) da bolsa de valores de Boston (BSE) e o volume combinado da bolsa de valores de Nova York com a bolsa americana (NYAMSE). A variável resposta é BSE e a variável explicativa, NYAMSE. O autor estava interessado em detectar o ponto de mudança devido à abolição da prática de desistência¹ em troca de ações na bolsa de valores. Para tal, utilizou uma abordagem bayesiana em que considerou como ponto de mudança o ponto que maximizava a densidade *a posteriori*. Por outro lado, Chen, [7] e [8], estava interessado em estudar a relação entre as duas variáveis, e ajustou

¹ Procedimento em negociação de valores mobiliários ou commodities, em que o corretor executor coloca uma negociação em nome de uma segunda corretora como se a segunda corretora tivesse efetivamente executado a operação. <https://www.investopedia.com/terms/g/giveup.asp>

Figura 4 – Série temporal com diferença na variância.



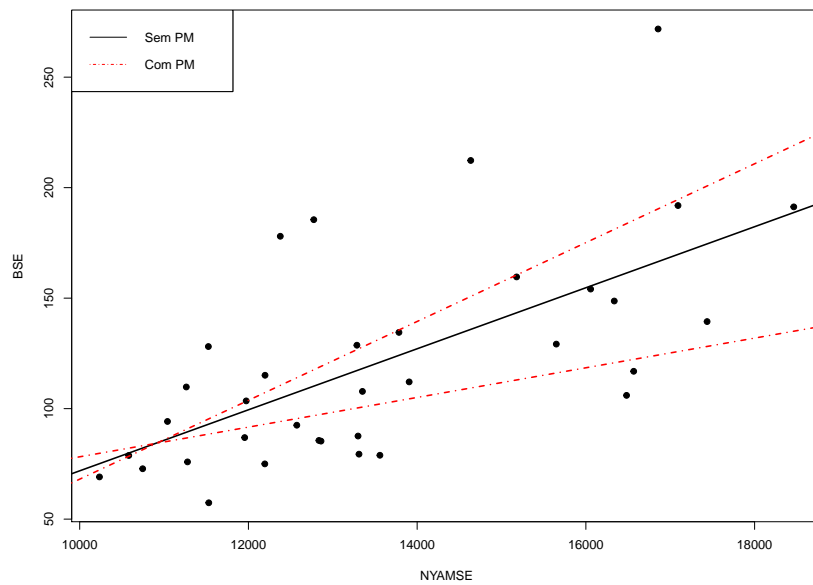
Fonte: Elaborado pelo autor.

um modelo de regressão que explicava o comportamento da variável BSE antes e depois da mudança na lei estadunidense.

A Tabela 2 apresenta os dados e a Figura 5 apresenta as retas estimada sob a hipótese de não haver ponto de mudança e sob a hipótese de ponto de mudança. Para a estimação das retas e da posição do ponto de mudança, [7] utiliza o SIC (Critério de Informação de Schwarz), definido da seguinte forma $SIC = -2\ell(\hat{\theta}) + s \log(n)$, em que $\ell(\hat{\theta})$ denota a função de log-verossimilhança para o modelo, s o número de parâmetros do modelo e n é o tamanho da amostra.

Note que nesta abordagem os autores trabalharam com a hipótese de independência dos erros para a estimação dos parâmetros do modelo de regressão. No presente trabalho, desejamos tratar do caso em que esta suposição pode não estar satisfeita. Além disso, desejamos encontrar a posição do ponto de mudança através de algum critério de seleção, no caso, o SIC.

Figura 5 – Diagrama de dispersão da variável BSE versus NYAMSE. Em preto a reta estimada pelo método de mínimos quadrados que não considera o ponto de mudança; em vermelho tracejado a técnica proposta em [7] que considera o ponto de mudança.



Fonte: Elaborado pelo autor.

Tabela 2 – Valores das variáveis BSE e NYAMSE.

Observação	BSE (milhões de dólares)	NYAMSE (milhões de dólares)
1	78,80	10581,60
2	69,10	10234,30
3	87,60	13299,50
4	72,80	10746,50
5	79,40	13310,70
6	85,60	12835,50
7	75,00	12194,20
8	85,30	12860,40
9	86,90	11955,60
10	107,80	13351,50
11	128,70	13285,90
12	134,50	13784,40
13	148,70	16336,70
14	94,20	11040,50
15	128,10	11525,30
16	154,10	16056,40
17	191,30	18464,30
18	191,90	17092,20
19	159,60	15178,80
20	185,50	12774,80
21	178,00	12377,80
22	271,80	16856,30
23	212,30	14635,30
24	139,40	17436,90
25	106,00	16482,20
26	112,10	13905,40
27	103,50	11973,70
28	92,50	12573,60
29	116,90	16566,80
30	78,90	13558,70
31	57,40	11530,00
32	75,90	11278,00
33	109,80	11263,70
34	129,20	15649,50
35	115,10	12197,10

Fonte: Encontrada em [7] e [13].

2 MODELO DE REGRESSÃO LINEAR SIMPLES E MODELOS AUTO-REGRESSIVOS

2.1 MODELO DE REGRESSÃO LINEAR

Suponha que a média de Y , uma variável aleatória de interesse, possa ser explicada por uma variável auxiliar X através de um relação linear. Suponha, também, que Y possa ser descrito como sendo a soma de uma quantidade determinística com uma quantidade aleatória, independente de X , denotado por ϵ , cujo valor esperado é zero e, para $i \neq j$, $Cor(\epsilon_i, \epsilon_j) = 0$. Esta quantidade aleatória será chamada de *erro*. Esta seção utiliza [6] e [16] como referências principais.

Sendo β_0 e β_1 os coeficientes da reta, intercepto e inclinação, respectivamente, e σ^2 a variância do erro, denotamos o modelo de regressão linear simples como:

$$Y|X = x = \beta_0 + \beta_1 x + \epsilon.$$

Um resultado imediato segue desta formulação.

Resultado 1.

$$\begin{aligned} E[Y|X = x] &= \beta_0 + \beta_1 x \\ Var[Y|X = x] &= \sigma^2. \end{aligned}$$

Demonstração. Pela suposição inicial, Y pode ser descrito como uma soma através da relação linear com X , $\beta_0 + \beta_1 x$ com o erro aleatório ϵ . Como a parte dada por uma função de x é constante e a partir das suposições iniciais do valor esperado do erro e da variância do erro temos que:

$$\begin{aligned} E[Y|X = x] &= E[\beta_0 + \beta_1 x + \epsilon|X = x] \\ &= E[\beta_0 + \beta_1 x|X = x] + E[\epsilon] \\ &= \beta_0 + \beta_1 x + 0 \\ &= \beta_0 + \beta_1 x \\ Var[Y|X = x] &= Var[\beta_0 + \beta_1 x + \epsilon] \\ &= Var[\epsilon] \\ &= \sigma^2. \end{aligned}$$

□

2.1.1 Estimação do modelo de regressão linear simples

A estimação de um modelo de regressão linear simples será aqui abordada por dois métodos: método dos mínimos quadrados ordinários e o método de máxima verossimilhança.

O primeiro método não requer suposições sobre a distribuição adjacente dos erros, enquanto o segundo método pressupõe uma distribuição para os erros.

O caso aqui apresentado será para a classe de distribuições simétricas, as quais englobam a distribuição normal.

2.1.1.1 Método de mínimos quadrados ordinários

Suponha que $Y = \beta_0 + \beta_1 X + \epsilon$ satisfaça as pressuposições do modelo de regressão linear simples. Além disso, tome uma amostra de tamanho n , isto é, $i = 1, \dots, n$ de Y e de X , tal que

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i.$$

Os erros ϵ_i representam a parte aleatória que não é explicada pelo modelo e β_0 e β_1 os coeficientes da reta desconhecidos que serão estimados. Para a estimação, gostaríamos de definir uma forma de medir *quanto* o modelo ajustado *não* explica dos dados. Os erros podem ser reescritos como diferenças entre a variável resposta e a parte sistemática do modelo $y_i - \beta_0 - \beta_1 x_i$. Gostaríamos que, para todo $i = 1, \dots, n$, essa diferença fosse mínima. Suponha que tomemos $\sum_{i=1}^n y_i - \beta_0 - \beta_1 x_i$ como uma função de β_0 e β_1 .

Ao procedermos à minimização desta soma em relação a β_0 e β_1 , encontraríamos que os parâmetros são iguais a zero. Note que se tomarmos o valor esperado da mesma soma,

$$E \left[\sum_{i=1}^n y_i - \beta_0 - \beta_1 x_i \right] = E \left[\sum_{i=1}^n \epsilon_i \right] = 0.$$

Para uma segunda abordagem, poderíamos tentar a minimização do *módulo* da mesma soma. Apesar de garantirmos que a soma será diferente de zero, perderíamos propriedades de continuidade, que são necessárias para a solução analítica que estamos buscando.

Sendo assim, iremos definir a soma de quadrados de erros como a medida para o quanto nosso modelo está tendo de informação perdida.

Definição 1. *Suponha que os erros de um modelo de regressão linear simples satisfaçam as pressuposições do modelo. Definimos, então, a soma de quadrados dos erros como sendo:*

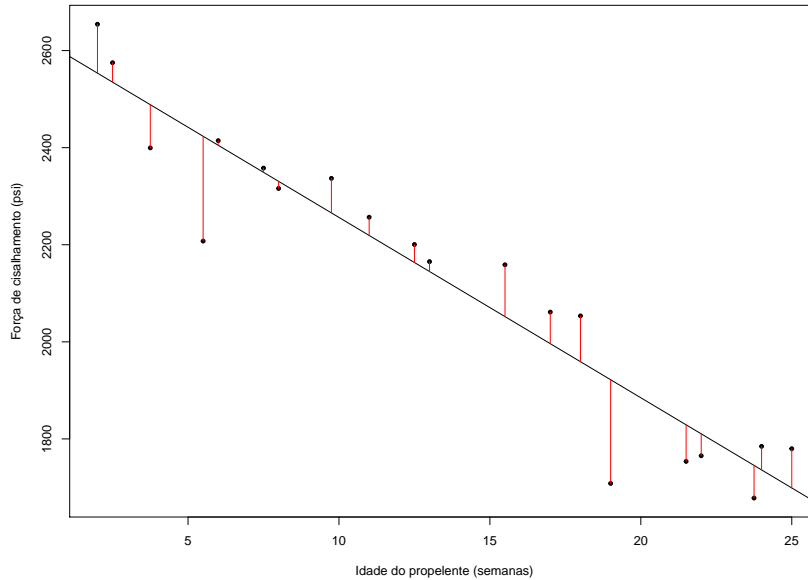
$$SQE = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n [y_i - \beta_0 - \beta_1 x_i]^2.$$

Para simplificar a notação, utilizaremos $SQE = S(\boldsymbol{\theta})$, em que $\boldsymbol{\theta} = (\beta_0, \beta_1)^T$. Note que temos a *soma de quadrados dos erros* no modelo teórico, enquanto que em uma amostra aleatória teremos *soma de quadrados dos resíduos*, denotada por

$$SQRes = \sum_{i=1}^n \hat{\epsilon}_i^2;$$

a soma de quadrados dos resíduos é um estimador para a soma de quadrados dos erros. Na Figura 6, temos nas linhas verticais em vermelho os resíduos que gostaríamos de minimizar.

Figura 6 – Diagrama de dispersão da Força de cisalhamento versus idade do combustível. As linhas verticais em vermelho são as distâncias das observações x_i à reta estimada.



Fonte: Elaborado pelo autor.

Desejamos encontrar $\hat{\theta}$ tal que

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} S(\theta).$$

Para tal, basta obtermos as derivadas parciais em relação a cada β_j do modelo de regressão e montarmos o *sistema de equações normais*.

Definição 2. Chamamos de *sistema de equações normais* o sistema:

$$\frac{\partial S(\theta)}{\partial \theta} = \mathbf{0}, \quad (2.1)$$

em que $\theta = (\beta_0, \beta_1)^\top$ no MRLS.

Resolvendo (2.1) para β_0 e β_1 , obtemos:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (2.2a)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n Y_i (X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} := \frac{S_{XY}}{S_{XX}} \quad (2.2b)$$

Um fato importante para os estimadores (2.2a) e (2.2b) é que são estimadores *não viesados* para β_0 e β_1 , respectivamente. Para tanto, basta tomarmos os valores esperados

de cada estimador:

$$\begin{aligned}
 E[\hat{\beta}_1|X] &= E\left[\frac{\sum_i Y_i(X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2}\right] = \frac{\sum_i E[Y_i](X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2} \\
 &= \frac{\sum_i (\beta_0 + \beta_1 X_i)(X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2} \\
 &= \frac{\sum_i \beta_0 (X_i - \bar{X}) + \sum_i \beta_1 X_i (X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2} \\
 &= \frac{\sum_i \beta_1 (X_i^2 - X_i \bar{X})}{\sum_i (X_i - \bar{X})^2}, \left(pois \sum_i (X_i - \bar{X}) = 0 \text{ e } \sum_i (X_i - \bar{X})^2 = \sum_i X_i (X_i - \bar{X}) \right) \\
 &= \beta_1.
 \end{aligned}$$

Para o estimador de β_0 , o argumento é semelhante:

$$\begin{aligned}
 E[\hat{\beta}_0|X] &= E[\bar{Y} - \hat{\beta}_1 \bar{X}|X] \\
 &= \frac{1}{n}n(\beta_0 + \beta_1 \bar{X}) - \beta_1 \bar{X} \\
 &= \beta_0.
 \end{aligned}$$

Retornando ao exemplo da **força de cisalhamento** em combustíveis de foguetes, temos as seguintes estimativas para os parâmetros do modelo:

$$\begin{aligned}
 \hat{\beta}_0 &= 2627.822 \text{ (44.184)} \\
 \hat{\beta}_1 &= -37.154 \text{ (2.189)}.
 \end{aligned}$$

Assim, o modelo estimado será denotado por:

$$\hat{y} = 2627.822 - 37.154x. \quad (2.3)$$

Como interpretação dos parâmetros, β_0 indica a força de cisalhamento na semana zero, ou quando o combustível acaba de ser produzido; β_1 indica o quanto a força de cisalhamento varia a cada semana, isto é, a cada semana a força de cisalhamento diminui em 37.154 unidades.

Com relação à variância dos erros, σ^2 , temos que o estimador não viesado pode ser obtido a partir da soma de quadrados dos resíduos, satisfazendo

$$\widehat{\sigma^2} = \frac{SQRes}{n - 2}$$

Para justificar, note que a soma de quadrados dos resíduos pode ser reescrita como:

$$\begin{aligned}
 SQRes &= \sum_i \hat{\epsilon}_i^2 = \sum_i \hat{\epsilon}_i(Y_i - \hat{Y}_i) = \sum_i \hat{\epsilon}_i Y_i - \overbrace{\sum_i \hat{\epsilon}_i \hat{Y}_i}^{=0} \\
 &= \sum_i (Y_i - \hat{Y}_i) Y_i \\
 &= \sum_i Y_i^2 - \sum_i (\hat{\beta}_0 + \hat{\beta}_1 X_i) Y_i && \text{(de 2.2a e 2.2b)} \\
 &= \sum_i Y_i^2 - n\bar{Y}^2 + n\hat{\beta}_1 \bar{X} \bar{Y} - \hat{\beta}_1 \sum_i X_i Y_i \\
 &= \sum_i Y_i^2 - n\bar{Y}^2 - \hat{\beta}_1 [\sum_i Y_i X_i - n\bar{X} \bar{Y}] \\
 &= \sum_i Y_i^2 - n\bar{Y}^2 - \hat{\beta}_1 \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2} \sum_i (X_i - \bar{X})^2 \\
 &= \sum_i Y_i^2 - n\bar{Y}^2 - \hat{\beta}_1^2 S_{XX}. && (2.4)
 \end{aligned}$$

Antes de tomarmos o valor esperado de (2.4), devemos notar que, fixadas as observações X_1, X_2, \dots, X_n ,

$$E[\sum_i Y_i^2] = n\sigma^2 + n\beta_0^2 + 2n\beta_0\beta_1\bar{X} + \beta_1^2 \sum_i X_i^2 \quad (2.5a)$$

$$E[n\bar{Y}^2] = \sigma^2 + n\beta_0^2 + n\beta_1^2 + 2n\beta_0\beta_1\bar{X} \quad (2.5b)$$

$$S_{XX}E[\hat{\beta}_1^2] = \sigma^2 + S_{XX}\beta_1^2. \quad (2.5c)$$

Assim, a partir de (2.5a), (2.5b) e (2.5c) podemos calcular o valor esperado de (2.4):

$$E[\sum_i \hat{\epsilon}_i^2] = (n - 2)\sigma^2.$$

Logo, um estimador não viesado para a variância dos erros σ_ϵ^2 é

$$\widehat{\sigma^2} = \frac{SQRes}{n - 2}.$$

Novamente em relação ao exemplo de **força de cisalhamento** em combustíveis de foguetes, a estimativa da variância será

$$\begin{aligned}
 \widehat{\sigma^2} &= \frac{SQRes}{n - 2} \\
 \widehat{\sigma^2} &= \frac{166254.858}{18} \\
 \widehat{\sigma^2} &= 9236.381.
 \end{aligned}$$

Além disso, podemos calcular as variâncias de $\widehat{\beta}_0$ e $\widehat{\beta}_1$

$$\begin{aligned}
 \text{Var}(\widehat{\beta}_1|X) &= \text{Var}\left(\frac{\sum_i Y_i(X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2} \middle| X\right) \\
 &= \frac{\sum_i (X_i - \bar{X})^2 \text{Var}(Y_i|X)}{\left[\sum_i (X_i - \bar{X})^2\right]^2} \\
 &= \frac{\sigma^2}{\sum_i (X_i - \bar{X})^2} \\
 \text{Var}(\widehat{\beta}_0|X) &= \text{Var}(\bar{Y} - \widehat{\beta}_1 \bar{X} | X) \\
 &= \text{Var}(\bar{Y}|X) + \text{Var}(\widehat{\beta}_1 \bar{X} | X) - 2\text{Cov}(\bar{Y}, \widehat{\beta}_1 \bar{X} | X) \\
 &= \frac{\sigma^2}{n} + \bar{X}^2 \frac{\sigma^2}{\sum_i (X_i - \bar{X})^2} - 2\bar{X} \text{Cov}(\bar{Y}, \widehat{\beta}_1) \quad (\text{se } \text{Cov}(\bar{Y}, \widehat{\beta}_1) = 0) \\
 &= \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_i (X_i - \bar{X})^2} \right]
 \end{aligned}$$

Para verificarmos que $\text{Cov}(\bar{Y}, \widehat{\beta}_1) = 0$, temos que

$$\begin{aligned}
 \text{Cov}(\bar{Y}, \widehat{\beta}_1 | X) &= \text{Cov}\left[\frac{\sum_i Y_i}{n}, \frac{\sum_i Y_i(X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2}\right] \\
 &= \frac{\sum_i (X_i - \bar{X}) \text{Cov}(Y_i, Y_i)}{n \sum_i (X_i - \bar{X})^2} \\
 &= 0 \frac{\sigma^2}{n \sum_i (X_i - \bar{X})^2} \\
 &= 0.
 \end{aligned}$$

Se supormos normalidade para os erros, ou seja, $\epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$, teremos o seguinte resultado.

Resultado 2. (*Distribuições de $\widehat{\beta}_0$ e $\widehat{\beta}_1$*)

$$\begin{aligned}
 \widehat{\beta}_0 &\sim N\left(\beta_0, \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_i (X_i - \bar{X})^2}\right]\right) \\
 \widehat{\beta}_1 &\sim N\left(\beta_1, \frac{\sigma^2}{\sum_i (X_i - \bar{X})^2}\right)
 \end{aligned}$$

A prova do Resultado 2 pode ser encontrada em [6].

2.1.1.2 Método de máxima verossimilhança

Uma outra abordagem para estimação é a do método de máxima verossimilhança.

Definição 3. *Seja θ um vetor de parâmetros de dimensão $s \times 1$. A função de verossimilhança de n variáveis aleatórias Y_1, \dots, Y_n é definida como sendo a densidade conjunta*

das n variáveis aleatórias, digamos $f_{Y_1, \dots, Y_n}(y_1, \dots, y_n | \boldsymbol{\theta})$, que é considerada uma função de $\boldsymbol{\theta}$ fixados os valores observados. Em particular, se Y_1, \dots, Y_n for uma amostra aleatória da densidade $f(y | \boldsymbol{\theta})$, então a função de verossimilhança será $f(y_1 | \boldsymbol{\theta}) f(y_2 | \boldsymbol{\theta}) \cdots f(y_n | \boldsymbol{\theta})$. Denotamos a função de verossimilhança por

$$L(\boldsymbol{\theta} | \mathbf{y}) = \prod_{i=1}^n f_{Y_i}(y_i | \boldsymbol{\theta}).$$

A função de log-verossimilhança será o logaritmo de $L(\boldsymbol{\theta} | \mathbf{y})$:

$$\ell(\boldsymbol{\theta} | \mathbf{y}) = \ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta} | \mathbf{y}) = \sum_{i=1}^n \log f_{Y_i}(y_i | \boldsymbol{\theta}).$$

Definição 4. Chamamos de função escore a quantidade:

$$U(\boldsymbol{\theta}) = \frac{\partial \log f_{\mathbf{Y}}(\mathbf{y} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}. \quad (2.6)$$

(2.6) satisfaz

$$E \left[\frac{\partial \log f_{\mathbf{Y}}(\mathbf{y} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right] = \mathbf{0}.$$

Para o modelo de regressão linear simples normal, a função de log-verossimilhança será dada por:

$$\ell(\beta_0, \beta_1, \sigma^2 | \mathbf{y}) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2.$$

Para os estimadores de máxima verossimilhança, basta determinarmos a função escore para o caso normal e igualar a zero, obtendo, assim, um *sistema de equações normais*. As raízes da função escore serão os estimadores de máxima verossimilhança.

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_0} = 0 &\implies \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_1} = 0 &\implies \hat{\beta}_1 = \frac{\sum_i Y_i (X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2} = \frac{S_{XY}}{S_{XX}}, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma^2} = 0 &\implies \hat{\sigma}^2 = \frac{1}{n} \sum_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \frac{SQRes}{n}. \end{aligned}$$

Para garantir que estamos em um ponto de máximo, temos que avaliar a matriz Hessiana da função de verossimilhança. A condição é que a matriz

$$H = \frac{\partial^2 L(\boldsymbol{\theta} | \mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$$

seja negativa definida, isto é, $\mathbf{z}^\top H \mathbf{z} < 0, \forall \mathbf{z} \neq \mathbf{0}$.

Assim como no processo de maximização da função de verossimilhança, observaremos a matriz Hessiana para a log-verossimilhança, devido à monotonicidade da mesma. Assim, a matriz H pode ser reescrita como

$$H = \frac{\partial^2 \ell(\boldsymbol{\theta} | \mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}.$$

As entradas da matriz H para o modelo de regressão simples serão:

$$\begin{aligned}\frac{\partial^2 \ell(\boldsymbol{\theta}|\mathbf{y})}{\partial \widehat{\beta}_0^2} &= -\frac{n}{\widehat{\sigma}^2} \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\mathbf{y})}{\partial \widehat{\beta}_1^2} &= -\frac{\sum_i X_i^2}{\widehat{\sigma}^2} \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\mathbf{y})}{\partial (\widehat{\sigma}^2)^2} &= \frac{n}{2(\widehat{\sigma}^2)^2} - \frac{\sum_i (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i)^2}{\widehat{\sigma}^2} \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\mathbf{y})}{\partial \widehat{\beta}_0 \partial \widehat{\beta}_1} &= -\frac{n\bar{X}}{\widehat{\sigma}^2} \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\mathbf{y})}{\partial \widehat{\beta}_0 \partial \widehat{\sigma}^2} &= -\frac{\sum_i (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i)}{\widehat{\sigma}^2} \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\mathbf{y})}{\partial \widehat{\beta}_1 \partial \widehat{\sigma}^2} &= -\frac{\sum_i (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i) X_i}{\widehat{\sigma}^2}.\end{aligned}$$

Note que a matriz Hessiana é simétrica, isto é, $H^\top = H$. Para verificarmos que H é negativa definida, tomaremos os determinantes das submatrizes principais (menores principais): $|H_1|, |H_2|, \dots, |H_p|$, em que $|H_j|$ denota o determinante da j -ésima submatriz quadrada.

Proposição 1. *Uma matriz A é dita negativa definida se:*

- para todo $\mathbf{z} \neq \mathbf{0}$, $\mathbf{z}^\top A \mathbf{z} < 0$ ou
- se os menores principais de A tiverem sinais alternados, isto é: $|A_1| < 0, |A_2| > 0, \dots$ ou
- se todos os autovalores de A , digamos λ_i , forem negativos.

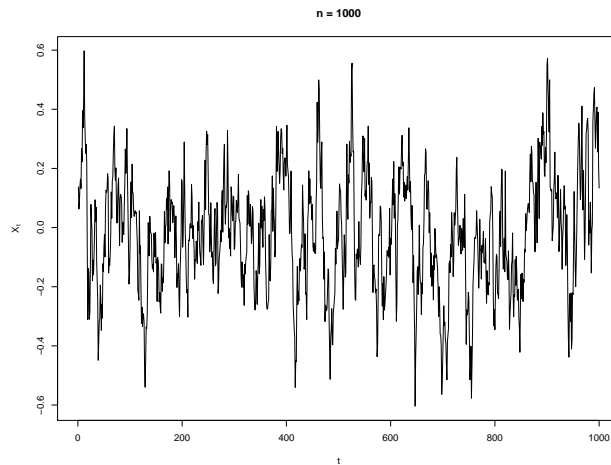
Para a prova da Proposição 1, veja [14].

2.2 MODELOS AUTOREGRESSIVOS

Para começarmos o estudo sobre modelos autoregressivos, devemos entender o que é uma série temporal. Uma *série temporal* “é um conjunto de observações de uma variável que evolui ao longo do tempo, realizadas em instantes diferentes” [5]. Os instantes podem ser equidistante.

As séries temporais podem ser utilizadas em diferentes aplicações, como em economia ou demanda e aquisição de recursos. Um exemplo de uma série temporal simulada pode ser visto na Figura 7.

Figura 7 – Realização de uma série temporal.



Fonte: Elaborado pelo autor.

Definição 5. Sejam Z_t e Z_{t-k} observações distintas de uma variável aleatória Z e sejam τ_{Z_t} e $\tau_{Z_{t-k}}$ seus valores esperados (também conhecidos como nível da série), respectivamente. A covariância para um defasamento de k instantes, aqui denominada autocovariância, é definida por:

$$\gamma_k = E[(Z_t - \tau_{Z_t})(Z_{t-k} - \tau_{Z_{t-k}})].$$

Para uma série estacionária, a Definição 5 se reduz ao caso em que $\tau_{Z_t} = \tau_{Z_{t-k}} = \tau$, o nível da série. Observe que a autocovariância para $k = 0$ nos dá a variância do processo

$$\gamma_0 = E[(Z_t - \tau_{Z_t})(Z_t - \tau_{Z_t})] = \text{Var}[Z_t] = \sigma_Z^2.$$

Definimos, também, a autocorrelação de um processo em defasagem k como sendo a autocovariância na defasagem k padronizada pela variância do processo, isto é

$$\rho_k = \frac{\gamma_k}{\sigma_Z^2}$$

e para um processo estacionário a autocorrelação se torna

$$\rho_k = \frac{\gamma_k}{\gamma_0}. \quad (2.8)$$

Definição 6. Um modelo autoregressivo de ordem p , denotado por $AR(p)$, é definido como

$$Z_t = \tau + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \epsilon_t,$$

em que Z_t é uma série temporal, ϕ_1, \dots, ϕ_p são os parâmetros, τ é o nível da série (constante) e ϵ_t é o ruído branco (aleatório).

É possível centrar os valores da série ao subtrairmos o nível da mesma $\bar{Z}_t = Z_t - \tau$.

Definimos o operador de defasagem B como sendo uma operação sobre um elemento da série temporal que produz o elemento anterior da série

$$BZ_t = Z_{t-1}. \quad (2.9)$$

Caso o operador seja aplicado k vezes

$$B^k Z_t = Z_{t-k}. \quad (2.10)$$

Munido do operador de defasagem, um modelo autoregressivo pode ser reescrito como

$$\begin{aligned} \bar{Z}_t &= \phi_1 \bar{Z}_{t-1} + \dots + \phi_p \bar{Z}_{t-p} + \epsilon_t \\ \bar{Z}_t &= \phi_1 B \bar{Z}_t + \dots + B^p \bar{Z}_t + \epsilon_t \\ (1 - \phi_1 B + \dots + \phi_p B^p) \bar{Z}_t &= \epsilon_t \\ \Phi(B) \bar{Z}_t &= \epsilon_t, \end{aligned}$$

se definirmos como operador autoregressivos de ordem p o polinômio aplicado em B dado por

$$\Phi_p(B) = 1 - \phi_1 B + \dots + \phi_p B^p. \quad (2.11)$$

O polinômio (2.11) é chamado de *polinômio característico do processo* e para que ele descreva um modelo autoregressivo, é necessário que verifique uma condição de estacionariedade.

Condição de estacionariedade. *As raízes do polinômio característico devem estar fora do círculo de raio unitário [5]*

$$|B| > 1.$$

Algumas propriedades importantes do modelo AR(p):

P.1 $E[Z_t] = \frac{\tau}{1 - \phi_1 - \dots - \phi_p};$

P.2 $\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p};$

P.3 $\sigma_Z^2 = \frac{\sigma_\epsilon^2}{1 - \phi_1 \rho_{k-1} - \dots - \phi_p \rho_{k-p}}.$

Para as provas e mais propriedades acerca do modelo autoregressivo, veja [5, Capítulo 3].

A Propriedade P.2 define a *função de autocorrelação* (ou FAC) para um modelo AR(p). Note também que se multiplicarmos ambos os lados por γ_0 obtemos a *função de autocovariância*.

Observe que, para cada $k = 1, \dots, n$ a P.2 define um sistema de equações denominado *equações de Yule-Walker* [5]. Os coeficientes ϕ_1, \dots, ϕ_p e ρ_1, \dots, ρ_n satisfazem as equações de Yule-Walker e podem ser escritos como

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}. \quad (2.12)$$

Ou em notação matricial

$$\boldsymbol{\rho}_p = \mathbf{P}_p \boldsymbol{\phi}$$

em que o índice p denota a ordem do processo. Com isso, os estimadores de Yule-Walker [5, Subseção 3.2.2] para os parâmetros $\boldsymbol{\phi}$, ao substituirmos as correlações ρ_k teóricas pelas correlações r_k observadas, será

$$\hat{\boldsymbol{\phi}} = \mathbf{P}_p^{-1} \boldsymbol{\rho}_p.$$

A seguir apresentamos algumas propriedades de modelos autoregressivos de ordens 1 e 2.

2.2.1 Modelo AR(1)

Um modelo $AR(1)$ pode ser descrito da seguinte forma

$$Z_t = \tau + \phi_1 Z_{t-1} + \epsilon_t, \quad (2.13)$$

e quando centrado

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + \epsilon_t.$$

A Figura 8 é um exemplo de um processo autoregressivos de ordem 1.

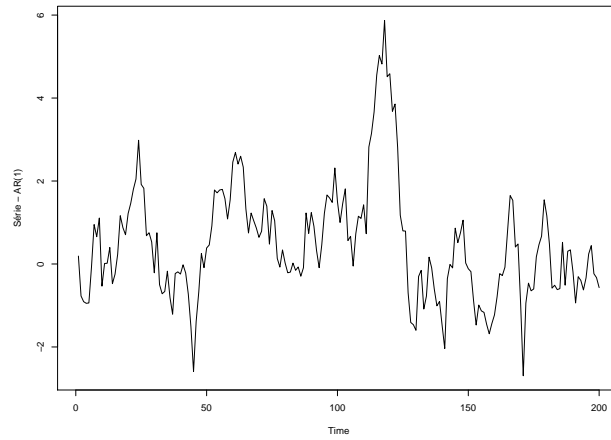
Considerando $\tau = 0$ em (2.13), aplicaremos o operador de defasagem B :

$$\begin{aligned} Z_t &= \phi_1 B Z_t + \epsilon_t \\ Z_t - \phi_1 B Z_t &= \epsilon_t \\ (1 - \phi_1 B) Z_t &= \epsilon_t \\ \Phi_1(B) Z_t &= \epsilon_t, \end{aligned} \quad (2.14)$$

note que $\Phi_1(B)$ em (2.14) é o polinômio característico do processo $AR(1)$ e gostaríamos que sua raiz esteja fora do círculo de raio unitário, isto é

$$\begin{aligned} \Phi(B) &= 0 \\ 1 - \phi_1 B &= 0 \\ B &= \phi_1^{-1} > 1 \\ \implies \phi_1 &< 1. \end{aligned}$$

Figura 8 – Realização de um processo AR(1).



Fonte: Elaborado pelo autor.

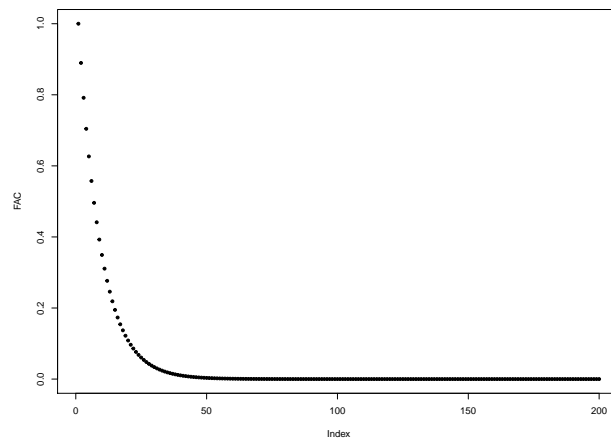
A função de autocorrelação, definida em P.2, fixando $\rho_0 = 1$, é da forma

$$\rho_k = \phi_1 \rho_{k-1}, \quad k > 0$$

$$\rho_k = \phi_1^k, \quad k \geq 0.$$

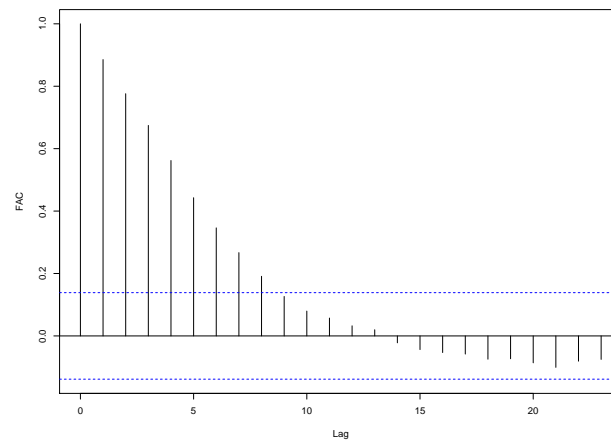
A Figura 9 ilustra um exemplo de função de autocorrelação teórica, enquanto a Figura 10 ilustra um exemplo de função de autocorrelação empírica, obtido a partir da função `acf` no software R[17].

Figura 9 – Função de autocorrelação teórica com $\phi_1 = 0.8897$.



Fonte: Elaborado pelo autor.

Figura 10 – Função de autocorrelação empírica com $\phi_1 = 0.8897$.



Fonte: Elaborado pelo autor.

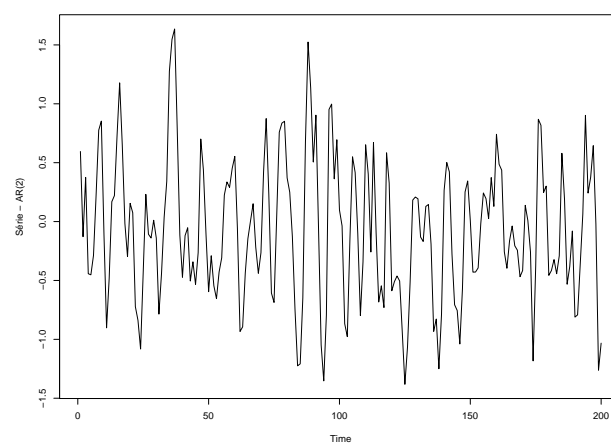
2.2.2 Modelo AR(2)

Um modelo $AR(2)$, já sendo considerado centrado, pode ser escrito da seguinte forma:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \epsilon_t.$$

A Figura 11 é um exemplo de um processo autoregressivos de ordem 2.

Figura 11 – Realização de um processo AR(2).



Fonte: Elaborado pelo autor.

Ao aplicarmos o operador de defasagem B , obtemos:

$$\begin{aligned}
 Z_t &= \phi_1 B Z_t + \phi_2 B^2 Z_t + \epsilon_t \\
 Z_t - \phi_1 B Z_t - \phi_2 B^2 Z_t &= \epsilon_t \\
 (1 - \phi_1 B - \phi_2 B^2) Z_t &= \epsilon_t \\
 \Phi_2(B) Z_t &= \epsilon_t,
 \end{aligned} \tag{2.15}$$

em que $\Phi_2(B)$ em 2.15 é o polinômio característico do processo $AR(2)$ e gostaríamos que suas raízes também estejam fora do círculo de raio unitário, o que implica que os parâmetros ϕ_1 e ϕ_2 estejam fora da região triangular:

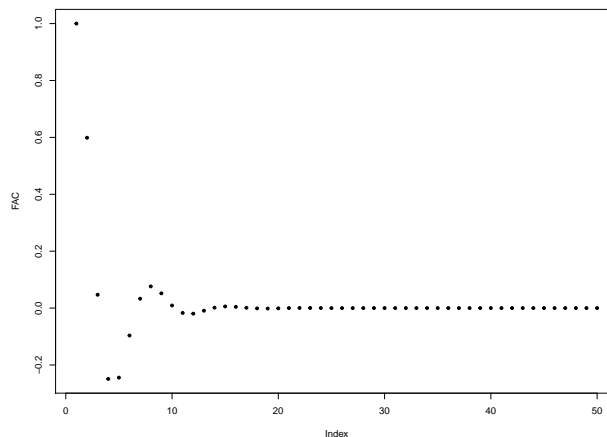
$$\begin{aligned}
 \phi_2 + \phi_1 &< 0 \\
 \phi_2 - \phi_1 &< 1 \\
 -1 &< \phi_2 < 1.
 \end{aligned}$$

Podemos também definir a função de autocorrelação para o processo $AR(2)$. Fixando $\rho_0 = 1$ e $\rho_1 = \phi_1 / (1 - \phi_2)$ [5]

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k > 0.$$

A Figura 12 ilustra um exemplo de função de autocorrelação teórica, enquanto a Figura 13 ilustra um exemplo de função de autocorrelação empírica.

Figura 12 – Função de autocorrelação teórica com $\phi_1 = 0.8897$ e $\phi_2 = -0.4858$.

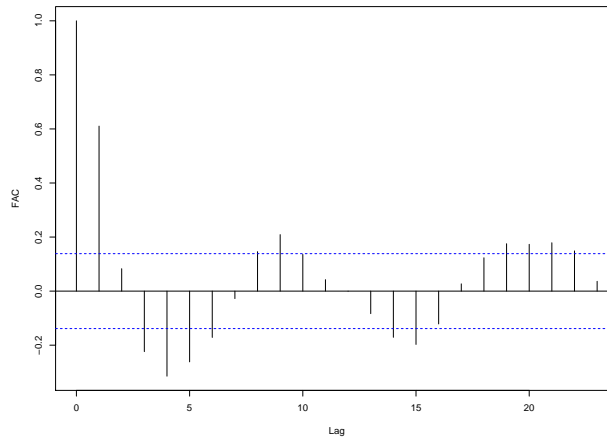


Fonte: Elaborado pelo autor.

2.2.3 Função de autocorrelação parcial

A *função de autocorrelação parcial* (ou FACP) é um mecanismo que explora o fato de podermos descrever a autocorrelação de um processo autoregressivo por p funções não

Figura 13 – Função de autocorrelação empírica com $\phi_1 = 0.8897$ e $\phi_2 = -0.4858$.



Fonte: Elaborado pelo autor.

nulas. Além disso, quando nos deparamos com o problema de não termos a informação sobre a ordem do processo, a FACP nos dá um indicativo sobre o possível valor da mesma.

Seja ϕ_{kj} o j -ésimo coeficiente na representação de um processo autoregressivo de ordem k de forma que ϕ_{kk} seja o último coeficiente. A partir da Propriedade P.2, ϕ_{kj} satisfaz o conjunto de equações

$$\rho_j = \phi_{k1}\rho_{j-1} + \dots + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k} \quad j = 1, 2, \dots, k, \quad (2.16)$$

isto é, as equações de Yule-Walker em (2.12) podem ser reescritas da seguinte forma

$$\begin{bmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}, \quad (2.17)$$

ou em notação matricial,

$$\mathbf{P}_k \boldsymbol{\phi}_k = \boldsymbol{\rho}_k.$$

Resolvendo (2.17) para $k = 1, 2$ obtemos

$$\phi_{11} = \rho_1 \quad (2.18a)$$

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}. \quad (2.18b)$$

(2.18a) é a função de autocorrelação parcial teórica para o modelo AR(1) e (2.18a) e (2.18a) juntas são a função de autocorrelação teórica para o modelo AR(2). Para o caso $k \geq 3$, ver [5, Seção 3.2.5].

A quantidade ϕ_{kk} , enquanto função da defasagem k , é chamada *função de autocorrelação parcial*. Para um processo autoregressivo de ordem p , a função de autocorrelação parcial ϕ_{kk} é *diferente de zero* para k menor ou igual a p e *zero* para k maior do que p . Em outras palavras, a FACP de um processo $AR(p)$ tem um *corte* após a defasagem p . Sendo assim, a FACP pode ser usada como uma estimativa para a ordem.

Podemos obter estimativas de mínimos quadrados para os coeficientes de (2.16) ao substituímos as autocorrelações teóricas ρ_j pelas estimativas r_j . O que nos leva às seguintes equações

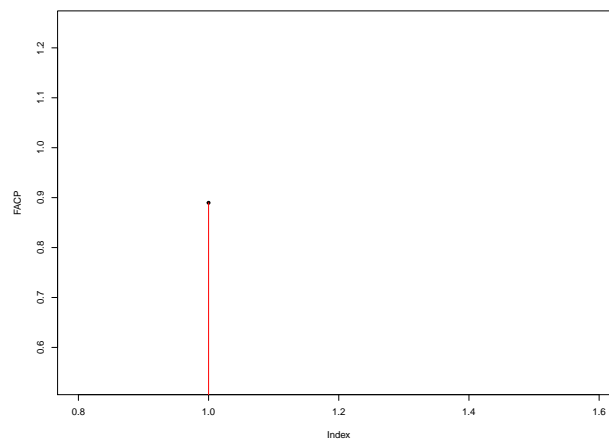
$$r_j = \hat{\phi}_{k1}r_{j-1} + \hat{\phi}_{k2}r_{j-2} + \dots + \hat{\phi}_{kk}r_{j-k}, \quad j = 1, 2, \dots, k. \quad (2.19)$$

Em [5], há uma proposta para o erro padrão dos coeficientes estimados em (2.19). O erro padrão para a autocorrelação parcial estimada $\hat{\phi}_{kk}$ é

$$EP[\hat{\phi}_{kk}] = \hat{\sigma}[\hat{\phi}_{kk}] \simeq \frac{1}{\sqrt{n}} \quad k \geq p + 1.$$

As Figuras 14 e 15 ilustram as funções de autocorrelação parcial teóricas para exemplos de $AR(1)$ e $AR(2)$, respectivamente. A linha pontilhada na Figura 15 é o erro padrão estimado; na Figura 14 o erro padrão não é exibido por ter ficado muito próximo de zero.

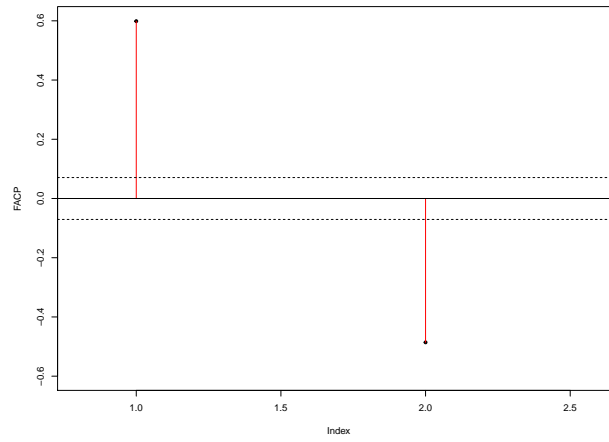
Figura 14 – Função de autocorrelação parcial teórica com $\phi_1 = 0.8897$.



Fonte: Elaborado pelo autor.

As Figuras 16 e 17 são as funções de autocorrelação empíricas para as séries nas Figuras 8 e 11, respectivamente. A linha azul pontilhada nas Figuras indicam a significância dos coeficientes estimados. No caso da Figura 16, há apenas um coeficiente significativo, o que sugere que a ordem desta série seja 1. Já na Figura 17, há dois coeficientes significativos, que indica ordem 2 para o processo.

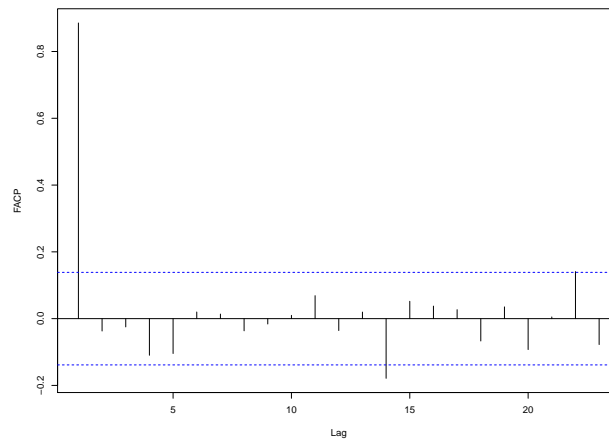
Figura 15 – Função de autocorrelação parcial teórica com $\phi_1 = 0.8897$ e $\phi_2 = -0.4858$.



Fonte: Elaborado pelo autor.

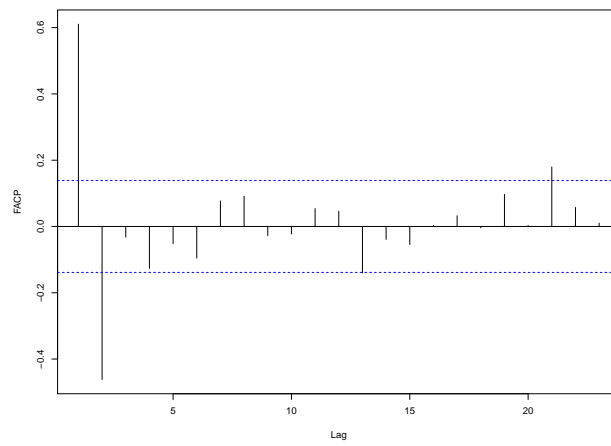
No Capítulo 3, apresentamos o modelo de regressão que possui erro autoregressivos. Mostraremos a estimação dos parâmetros da regressão, assim como dos coeficientes autoregressivos.

Figura 16 – Função de autocorrelação parcial empírica com $\phi_1 = 0.8897$.



Fonte: Elaborado pelo autor.

Figura 17 – Função de autocorrelação parcial empírica com $\phi_1 = 0.8897$ e $\phi_2 = -0.4858$.



Fonte: Elaborado pelo autor.

3 MODELO DE REGRESSÃO LINEAR SIMPLES COM ERROS AUTO-REGRESSIVOS DE ORDEM p

Na Seção 2.1, apresentamos as suposições que um modelo deve satisfazer para procedermos em uma análise considerando-o como um Modelo de Regressão Linear Simples. Aqui, a suposição de *independência dos erros* não pode ser satisfeita. Gostaríamos, então, de modelar essa dependência nos erros. Este capítulo, assim como a Seção 2.2, baseia-se em [5] e é complementado por [18].

Considere o modelo de regressão

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad (3.1a)$$

$$\epsilon_i = \phi_1 \epsilon_{i-1} + \dots + \phi_p \epsilon_{i-p} + \eta_i, \quad \eta_i \sim N(0, \sigma^2), \quad \forall i = 1, \dots, n, \quad (3.1b)$$

em que Y_i é a variável resposta do modelo, $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ é o vetor de parâmetros da regressão, X_i é a variável explicativa do modelo, ϵ_i o termo de erro autoregressivo com ruído normal η_i e $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^\top$ o vetor dos coeficientes autoregressivos. O modelo formado por (3.1a) e (3.1b) será denotado por MRLS-AR(p).

A estacionariedade do modelo (3.1b) pode ser assegurada desde que as raízes de

$$1 - \phi_1 B - \dots - \phi_p B^p = 0$$

estejam fora do círculo de raio unitário [18], em que B denota o operador de retardo como definido em (2.9) e (2.10).

Em notação matricial, o modelo MRLS-AR(p) pode ser reescrito como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (3.2)$$

em que $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ é o vetor com as respostas do modelo, \mathbf{X} é uma matriz $n \times l$ ($l = 2$ no caso do modelo de regressão linear simples) e $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$, em que $\boldsymbol{\Sigma} = \sigma^2 M_n(\boldsymbol{\phi})$, tal que

$$M_n(\boldsymbol{\phi}) = \frac{1}{\sigma^2} \begin{bmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{n-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \dots & \gamma_0 \end{bmatrix}, \quad (3.3)$$

em que $\gamma_0, \dots, \gamma_{n-1}$ são as autocovariâncias teóricas do processo e as autocorrelações teóricas $\rho_k = \gamma_k / \gamma_0$, $k = 1, 2, \dots$, como definidas em (2.8). Além disso, σ_2 é a variância do ruído η_i e os coeficientes ϕ_1, \dots, ϕ_p e ρ_1, \dots, ρ_n satisfazem as equações de Yule-Walker apresentadas em (2.12).

3.1 ESTIMAÇÃO

Antes de iniciarmos o processo de estimação, devemos garantir a admissibilidade (devido à complexidade do ϕ -espaço) dos parâmetros autoregressivos ϕ e estabilização do procedimento. Para isso, [3] apresenta uma reparametrização de ϕ como

$$\begin{aligned}\phi_p^{(p)} &= \pi_p, \\ \phi_\nu^{(p)} &= \phi_\nu^{(p-1)} - \pi_p \phi_{p-\nu}^{(p-1)},\end{aligned}\tag{3.4}$$

em que $\phi_\nu^{(p)}$ denota o ν -ésimo parâmetro AR no modelo em (3.1b), e $\pi_\nu = \phi_\nu^{(\nu)}$ é a autocorrelação parcial na defasagem ν , para $\nu = 1, \dots, p-1$. Esta recursão é conhecida como recursão de Durbin-Levinson. Com (3.4) definimos a seguinte transformação

$$\mathcal{B} : (\pi_1, \dots, \pi_p) \rightarrow (\phi_1, \dots, \phi_p).$$

A transformação \mathcal{B} é 1 a 1, contínua e diferenciável na região admissível. A vantagem desta reparametrização é que dentro do π -espaço a região admissível é simplesmente um cubo p -dimensional com as superfícies de fronteiras correspondendo a ± 1 , enquanto no ϕ -espaço é muito mais complicado [15].

Prossigamos à estimação dos parâmetros do modelo (3.1). A função de log-verossimilhança exata, em notação matricial, é

$$\ell(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top M_n^{-1}(\boldsymbol{\phi}) (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - \log |M_n^{-1}(\boldsymbol{\phi})|,\tag{3.5}$$

em que $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \sigma^2)^\top$ é o vetor de parâmetros do modelo e \mathbf{Y} e \mathbf{X} são definidas como em (3.2). Reparametrizando com (3.4), a log-verossimilhança é reescrita da seguinte forma:

$$\ell(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} S(\boldsymbol{\pi}, \boldsymbol{\beta}) - \frac{1}{2} \log g_p,\tag{3.6}$$

em que $\boldsymbol{\pi} = (\pi_1, \dots, \pi_p)^\top$, $g_p = |M_n(\boldsymbol{\phi})| = |M_n(\boldsymbol{\pi})| = \prod_{j=1}^p (1 - \pi_j^2)^{-j}$ e

$$S(\boldsymbol{\pi}, \boldsymbol{\beta}) = \boldsymbol{\lambda}^\top D(\mathbf{y}, \boldsymbol{\beta}) \boldsymbol{\lambda},\tag{3.7}$$

com $D(\mathbf{y}, \boldsymbol{\beta})$ sendo uma matriz $(p+1) \times (p+1)$ em que a (i, j) -ésima entrada é a soma dos $n - (i-1) - (j-1)$ produtos quadrados e defasados, definidos por:

$$D_{i,j} = D_{j,i} = (Y_i - \mathbf{x}_i^\top \boldsymbol{\beta})(Y_j - \mathbf{x}_j^\top \boldsymbol{\beta}) + \dots + (Y_{n+1-j} - \mathbf{x}_{n+1-j}^\top \boldsymbol{\beta})(Y_{n+1-j} - \mathbf{x}_{n+1-j}^\top \boldsymbol{\beta})\tag{3.8}$$

e $\boldsymbol{\lambda}^\top = (-1, \boldsymbol{\phi}^\top) = (-1, \mathcal{B}(\boldsymbol{\pi})^\top)$.

Os parâmetros desconhecidos são estimados por máxima verossimilhança (3.6). Como $\partial \ell(\boldsymbol{\theta}|\mathbf{y})/\partial \boldsymbol{\pi} = 0$ não tem solução analítica para $\boldsymbol{\pi}$, para obtermos o estimador de máxima verossimilhança (EMV) para $\boldsymbol{\theta}$ precisamos de uma aproximação numérica, isto é,

uma estimação perfilada. Segundo [18], dado $\boldsymbol{\pi}$, a log-verossimilhança (3.6) é maximizada em:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\pi}) = \left(\mathbf{X}^\top M_n^{-1}(\mathcal{B}(\boldsymbol{\pi})) \mathbf{X} \right)^\top \mathbf{X}^\top M_n^{-1}(\mathcal{B}(\boldsymbol{\pi})) \mathbf{y} \quad (3.9a)$$

$$\widehat{\sigma}^2(\boldsymbol{\pi}) = \frac{1}{n} S(\boldsymbol{\pi}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\pi})). \quad (3.9b)$$

Assim, substituindo (3.9a) e (3.9b) em (3.6), eliminando termos constantes, temos a função de log-verossimilhança concentrada:

$$\ell_c(\boldsymbol{\pi}|\mathbf{y}) = -\frac{n}{2} \log S(\boldsymbol{\pi}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\pi})) - \frac{1}{2} \log g_p, \quad (3.10)$$

que deve ser maximizada numericamente em relação à $\boldsymbol{\pi} \in (-1, 1)^p$, o cubo p -dimensional com superfícies de fronteira correspondendo a ± 1 , obtendo $\widehat{\boldsymbol{\pi}}$. Finalmente, os EMVs de $\boldsymbol{\beta}$ e σ^2 são calculados como $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}})$ e $\widehat{\sigma}^2 = \widehat{\sigma}^2(\widehat{\boldsymbol{\pi}})$. Podemos também obter estimativas para ϕ como sendo $\widehat{\phi} = \mathcal{B}(\widehat{\boldsymbol{\pi}})$.

3.2 APLICAÇÃO

Procederemos, nesta seção, com o ajuste de um modelo de regressão linear simples em um banco de dados ¹ descrito em [12]. Os dados são da Companhia Blaisdell, e a empresa gostaria de analisar a relação das vendas da companhia (variável *comsales*, em milhões de dólares) com as da indústria (variável *indsales*, em milhões de dólares).

A Figura 18 mostra o gráfico de dispersão da variável *comsales* versus *indsales*. A relação linear é bem clara e é evidenciada na Figura 19 ao traçarmos a reta de mínimos quadrados. Os coeficientes foram estimados utilizando a função `lm` no software R[17] apresentando os valores $\widehat{\beta}_0 = -1.455$ e $\widehat{\beta}_1 = 0.176$.

Prosseguindo a análise de resíduos, temos nas Figuras 20 e 21 as funções de autocorrelação e autocorrelação parcial para os resíduos $\widehat{\epsilon}_t = y_t - \widehat{y}_t$, $i = 1, \dots, 20$, em que \widehat{y}_t denota os valores ajustados. Essas figuras apontam para a existência de autocorrelação. Uma alternativa para verificar a presença de autocorrelação é o Teste de Durbin-Watson [9]. Este teste assume que a série segue um processo $AR(1)$ observado em períodos de tempo igualmente espaçados. As hipóteses consideradas no teste são:

$$H_0 : \phi = 0$$

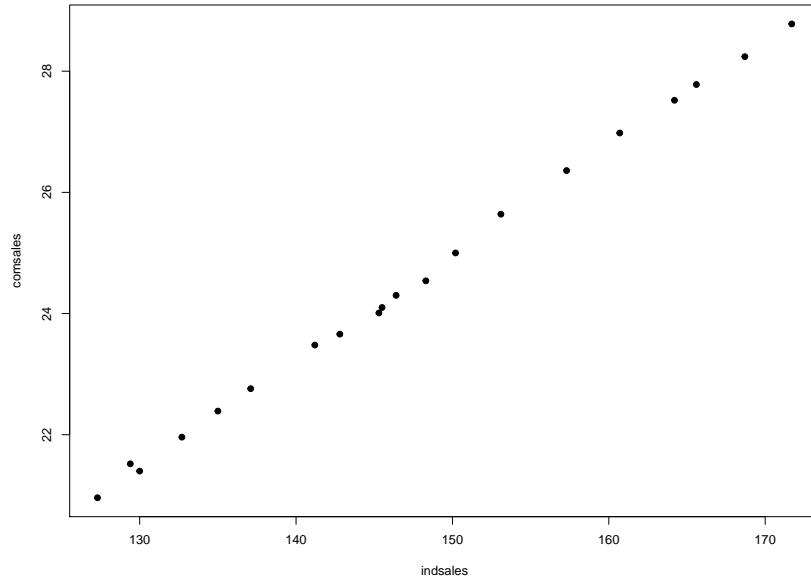
$$H_1 : \phi > 0.$$

A estatística de teste de Durbin-Watson é:

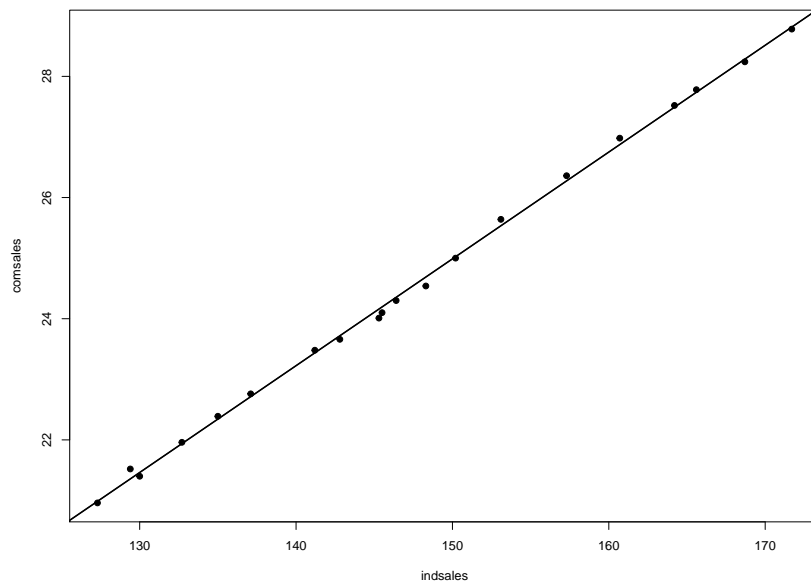
$$d = \frac{\sum_{t=2}^T (\widehat{\epsilon}_t - \widehat{\epsilon}_{t-1})^2}{\sum_{t=2}^T \widehat{\epsilon}_t^2}, \quad (3.11)$$

em que $\widehat{\epsilon}_t$ denota o resíduo de uma regressão por mínimos quadrados ordinários (MQO).

¹ <https://newonlinecourses.science.psu.edu/stat462/sites/onlinecourses.science.psu.edu.stat462/files/data/blaisdell/index.txt>

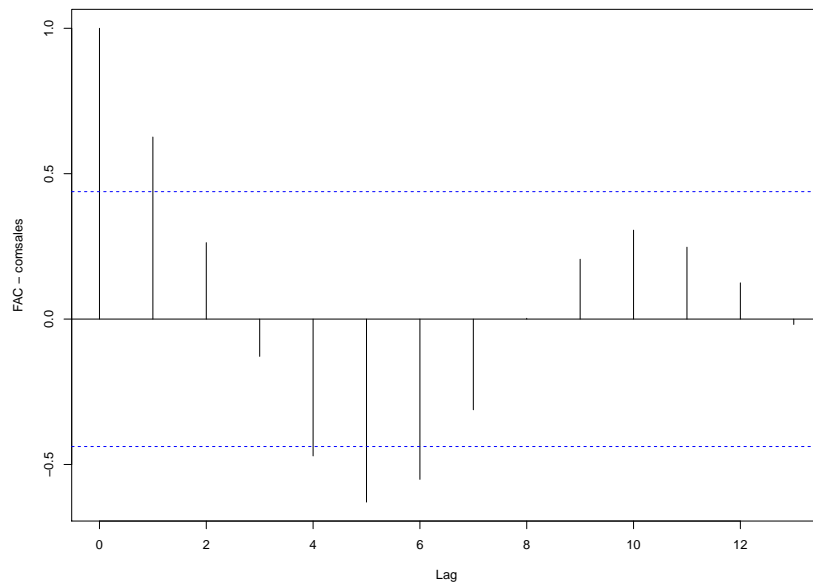
Figura 18 – Diagrama de dispersão de *comsales* versus *indsales*.

Fonte: Elaborado pelo autor.

Figura 19 – Diagrama de dispersão de *comsales* versus *indsales* com reta estimada via MQO.

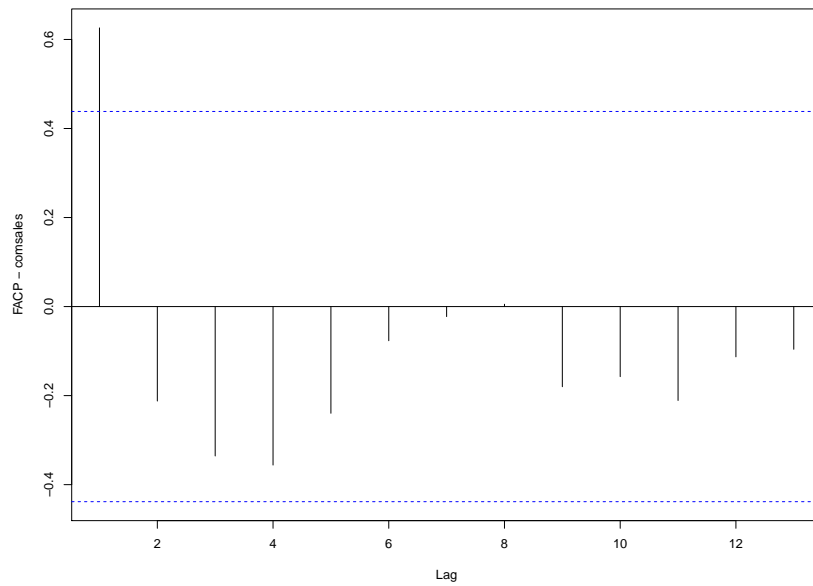
Fonte: Elaborado pelo autor.

Figura 20 – Função de autocorrelação dos resíduos da regressão.



Fonte: Elaborado pelo autor.

Figura 21 – Função de autocorrelação parcial dos resíduos da regressão.



Fonte: Elaborado pelo autor.

A estatística d está entre dois valores extremos d_I e d_S , inferior e superior, respectivamente. Em [10], foram construídas tabelas para os valores extremos com níveis de significância $\alpha = 0.05$, $\alpha = 0.025$ e $\alpha = 0.01$. As tabelas também podem ser encontradas em [16]. O processo de decisão é:

Se $d < d_I$, rejeitamos H_0 ,

Se $d > d_S$, não rejeitamos H_0 ,

Se $d_I \leq d \leq d_S$, o teste é inconclusivo.

Caso desejássemos investigar autocorrelação negativa, uma sugestão é utilizar a estatística $4 - d$, em que d é definida em (3.11). O processo de decisão é o mesmo e tem significância 2α .

O pacote `lmtest` [21], encontrado no R, apresenta a função `dwtest`, a qual realiza o teste de Durbin-Watson. Basta entrarmos com os dados que gostaríamos de investigar a presença de autocorrelação da seguinte forma:

`dwtest(formula = y ~ x).`

em que \mathbf{y} é o vetor com as observações da variável resposta e \mathbf{x} o vetor com as observações da variável explicativa. Por padrão, a função realiza o teste bilateral, mas é possível realizar testes unilaterais.

Utilizando, então a função `dwtest` da forma

`dwtest(comsales ~ indsales).`

A estatística d calculada tem valor 0.73473 e p -valor = 0.0001748. Com isso, rejeitamos a hipótese nula de autocorrelação igual a zero, coincidindo com o resultado que havíamos encontrado ao analisarmos a FACP. A Figura 22 e a Tabela 3 apresentam os resultados para o ajuste dos dados da Companhia Blaisdell.

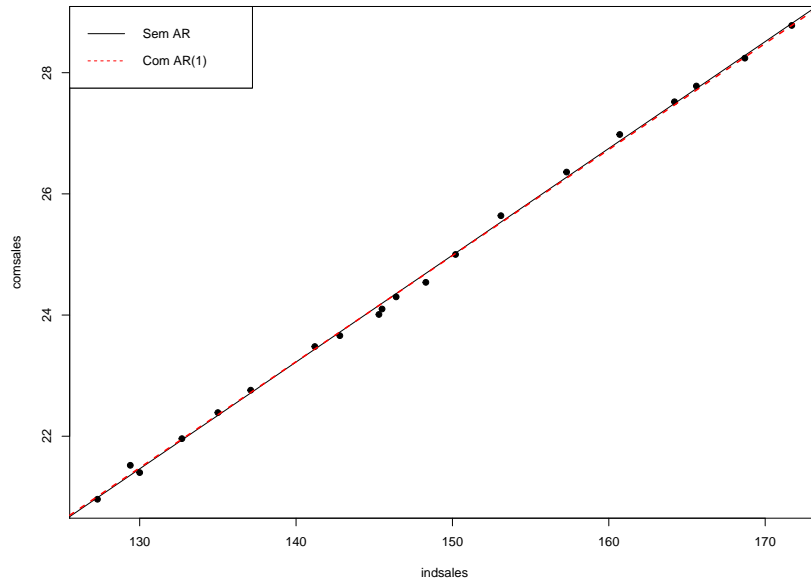
Tabela 3 – Coeficientes estimados para os modelos considerando independência dos erros e autocorrelação nos dados.

	Sem AR	Com AR
$\widehat{\beta}_0$	-1,455	-1,288
$\widehat{\beta}_1$	0,176	0,175
$\widehat{\sigma}^2$	0,007	0,004
$\widehat{\phi}_1$		0,629

Fonte: Elaborado pelo autor.

No Capítulo 4, introduziremos o modelo de regressão com ponto de mudança e erros normais autoregressivos. Assim como fizemos neste capítulo e nos anteriores, apresen-

Figura 22 – Diagrama de dispersão de *comsales* versus *indsales* com retas estimadas considerando independência dos erros e autocorrelação.



Fonte: Elaborado pelo autor.

taremos os estimadores para os parâmetros da regressão, dos coeficientes autoregressivos, além de identificarmos a posição do ponto de mudança.

4 MODELO DE REGRESSÃO LINEAR SIMPLES COM PONTO DE MUDANÇA E ERRO NORMAL AUTOREGRESSIVO

Na Seção 1.4, apresentamos uma breve introdução sobre pontos de mudança. Aqui, faremos uma apresentação formal, bem como introduziremos o Modelo de Regressão Linear Simples com Ponto de Mudança e Erro Normal autoregressivos de ordem p (ou MRLS-PM-AR(p)).

4.1 MODELO DE REGRESSÃO LINEAR SIMPLES COM PONTO DE MUDANÇA E ERROS INDEPENDENTES

Seja $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ uma sequência de observações obtidas em uma situação prática. Estamos interessados em ajustar um modelo de regressão linear simples aos dados

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \dots, n,$$

em que x_i , $i = 1, \dots, n$, é uma variável não-estocástica, β_0 e β_1 são parâmetros da regressão (desconhecidos), ϵ_i , $i = 1, \dots, n$, é um erro aleatório com distribuição $N(0, \sigma^2)$, σ^2 desconhecido, e ϵ_i e ϵ_j são não correlacionados $\forall i \neq j$. Isto é, y_i , $i = 1, \dots, n$ é uma variável aleatória com distribuição $N(\beta_0 + \beta_1 x_i, \sigma^2)$. Devido à natureza dos dados, suspeitamos que os coeficientes da regressão mudaram após um certo ponto, digamos k . Assim, podemos realizar um teste de hipóteses tal que:

$$\begin{aligned} H_0 : y_i &= \beta_0 + \beta_1 x_i, \quad \text{para } i = 1, \dots, n \\ H_1 : y_i &= \begin{cases} \beta_0^1 + \beta_1^1 x_i, & \text{para } i = 1, \dots, k, \\ \beta_0^2 + \beta_1^2 x_i, & \text{para } i = k + 1, \dots, n, \end{cases} \end{aligned} \quad (4.1)$$

em que $k = 2, \dots, n - 2$ é a posição do ponto de mudança, $\beta_0, \beta_1, \beta_0^1, \beta_1^1, \beta_0^2$ e β_1^2 são os parâmetros da regressão desconhecidos e σ^2 também é desconhecido. Gostaríamos de encontrar um método para identificar a posição k do ponto de mudança.

4.1.1 Identificação do ponto de mudança via critério de seleção

Critérios de seleção estão comumente associados à seleção de variáveis para um determinado modelo, seja para corrigir um sobreajuste ou um subajuste do modelo de interesse. Dessa forma, critérios de seleção nos ajudam a encontrar um modelo parcimonioso com boa habilidade de predição. [16, Capítulo 10] traz algumas abordagens que se adequam ao interesse do pesquisador.

Um dos critérios mais utilizados, é o Critério de Informação de Akaike, ou AIC [1], que se baseia em maximizar a *entropia* de um modelo. Entropia é uma medida da

informação esperada de um modelo [16]. Em sua essência, o AIC é uma penalização na *log-verossimilhança* de um modelo. Seja ℓ a *log-verossimilhança* de um modelo. O AIC é definido por:

$$AIC = -2\ell(\boldsymbol{\theta}) + 2s, \quad (4.2)$$

em que $\boldsymbol{\theta}$ é o vetor de parâmetros do modelo e s é a dimensão desse vetor, isto é, o número de parâmetros do modelo.

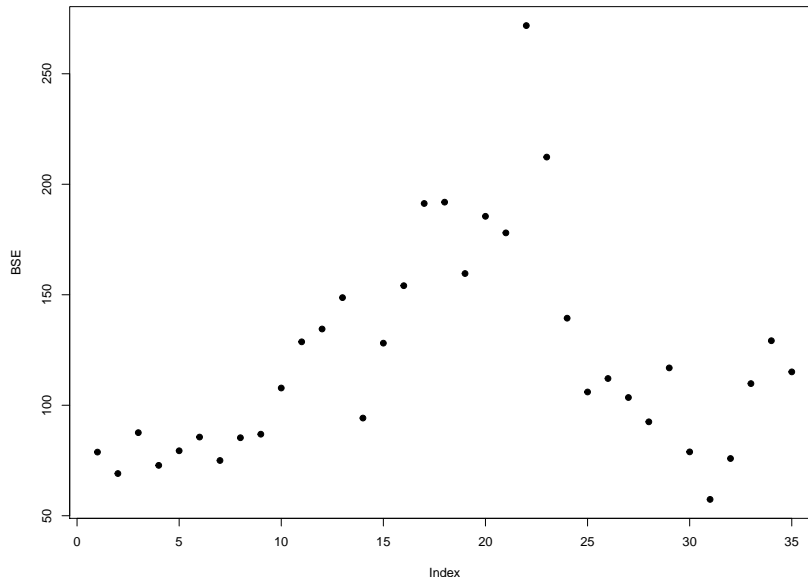
Já em [19] é proposta uma modificação do AIC, no qual a penalização na função de *log-verossimilhança* é maior,

$$SIC = -2\ell(\boldsymbol{\theta}) + s \log(n), \quad (4.3)$$

em que SIC é o Critério de Informação de Schwarz (também conhecido como Critério de Informação Bayesiano).

Retornemos ao exemplo apresentado na Seção 1.4. A Figura 23 apresenta o diagrama de dispersão da variável BSE (volume de vendas mensal da Bolsa de valores de Boston em milhões de dólares). Observando apenas a dispersão da variável reposta do conjunto de dados, podemos observar uma provável quebra no processo que gerou estas observações.

Figura 23 – Diagrama de dispersão da variável BSE.

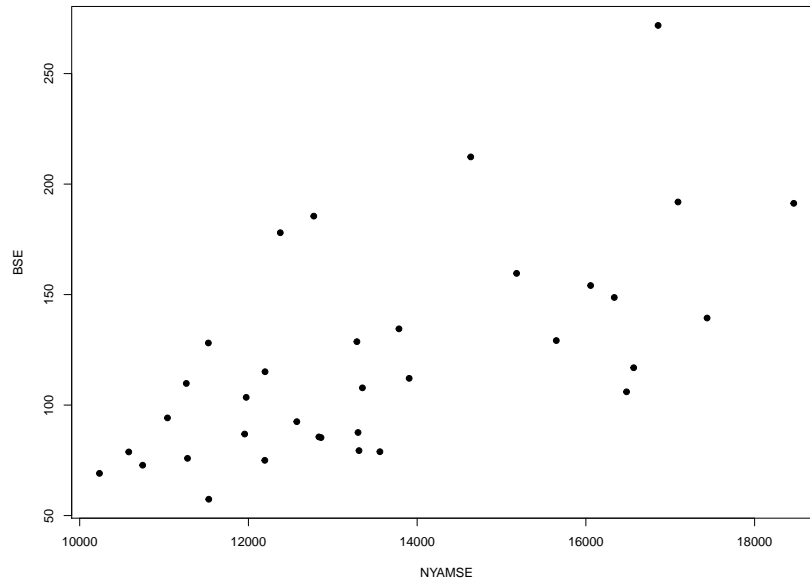


Fonte: Elaborado pelo autor com os dados em [13].

Contudo, sabemos que a variável BSE se relaciona com a variável NYAMSE de tal forma que podemos estimar um modelo regressão linear para explicar essa relação. A Figura 24 apresenta o diagrama de dispersão da variável BSE versus a variável NYAMSE.

Nela observamos uma relação linear entre as duas variáveis. Gostaríamos de investigar se esta relação é afetada pela mudança de comportamento das variáveis.

Figura 24 – Diagrama de dispersão da variável BSE.



Fonte: Elaborado pelo autor com os dados em [13].

Para tal, realizaremos o teste de hipóteses proposto em (4.1) e com isso precisamos de uma estatística de teste para que possamos tomar nossa decisão. Porém, a construção de tal estatística não é simples e, por isso, [7] propõe a utilização de critérios de informação. Pelo princípio de critério de informação [8, Subseção 2.2.2], encontramos o ponto de mudança k por \hat{k} tal que minimiza $SIC(\hat{k})$. Sendo assim, sob H_0 , a função log-verossimilhança é da forma:

$$\begin{aligned} \ell(\beta_0, \beta_1, \sigma^2 | \mathbf{Y}) &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \\ &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \end{aligned}$$

em que os EMVs do MRLS serão

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_0} = 0 &\implies \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_1} = 0 &\implies \hat{\beta}_1 = \frac{\sum_i Y_i (X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2} = \frac{S_{XY}}{S_{XX}}, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma^2} = 0 &\implies \hat{\sigma}^2 = \frac{1}{n} \sum_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = \frac{SQRes}{n}. \end{aligned}$$

O SIC sob H_0 , denotado por $SIC(n)$, será

$$\begin{aligned} SIC(n) &= -2\ell(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\sigma}^2) + s \log n \\ &= n \log 2\pi + n \log \widehat{\sigma}^2 + \frac{1}{\widehat{\sigma}^2} (\mathbf{Y} - \mathbf{X}\widehat{\beta})^\top (\mathbf{Y} - \mathbf{X}\widehat{\beta}) + s \log n \\ &= n \left[\log(\mathbf{Y} - \mathbf{X}\widehat{\beta})^\top (\mathbf{Y} - \mathbf{X}\widehat{\beta}) + 1 + \log 2\pi \right] + (s + 1 - n) \log n. \end{aligned}$$

Sob H_1 , supondo o ponto de mudança na posição k desconhecido, a log-verossimilhança é da forma:

$$\begin{aligned} \ell(\beta_0^1, \beta_1^1, \beta_0^2, \beta_1^2, \sigma^2 | \mathbf{Y}) &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^k (Y_i - \beta_0^1 - \beta_1^1 X_i)^2 \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=k+1}^n (Y_i - \beta_0^2 - \beta_1^2 X_i)^2 \quad (4.5) \end{aligned}$$

e os EMVs são:

$$\begin{aligned} \widehat{\beta}_0^1 &= \bar{Y}_k - \widehat{\beta}_1^1 \bar{X}_k, \\ \widehat{\beta}_1^1 &= \frac{\sum_{i=1}^k (Y_i - \bar{Y}_k)(X_i - \bar{X}_k)}{\sum_{i=1}^k (X_i - \bar{X}_k)^2}, \\ \widehat{\beta}_0^2 &= \bar{Y}_{n-k} - \widehat{\beta}_1^2 \bar{X}_{n-k}, \\ \widehat{\beta}_1^2 &= \frac{\sum_{i=k+1}^n (Y_i - \bar{Y}_{n-k})(X_i - \bar{X}_{n-k})}{\sum_{i=k+1}^n (X_i - \bar{X}_{n-k})^2}, \\ \widehat{\sigma}^2 &= \frac{1}{n} \left[\sum_{i=1}^k (Y_i - \widehat{\beta}_0^1 - \widehat{\beta}_1^1 X_i)^2 + \sum_{i=k+1}^n (Y_i - \widehat{\beta}_0^2 - \widehat{\beta}_1^2 X_i)^2 \right], \end{aligned}$$

em que

$$\bar{X}_k = \frac{1}{k} \sum_{i=1}^k X_i, \quad \bar{Y}_k = \frac{1}{k} \sum_{i=1}^k Y_i, \quad \bar{X}_{n-k} = \frac{1}{n-k} \sum_{i=k+1}^n X_i, \quad \bar{Y}_{n-k} = \frac{1}{n-k} \sum_{i=k+1}^n Y_i.$$

Para reescrever (4.5) em notação matricial, precisamos reescrever \mathbf{Y} e \mathbf{X} de forma que seu número de linhas dependa de k , isto é:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \quad \text{e} \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \quad (4.6)$$

tais que \mathbf{Y}_1 recebe as k primeiras linhas de \mathbf{Y} e \mathbf{Y}_2 recebe as $n - k$ linhas restantes; analogamente, \mathbf{X}_1 recebe as k primeiras linhas de \mathbf{X} , enquanto \mathbf{X}_2 recebe as $n - k$ linhas restantes. Com isso, a log-verossimilhança se torna:

$$\begin{aligned} \ell(\beta_0^1, \beta_1^1, \beta_0^2, \beta_1^2, \sigma^2 | \mathbf{Y}) &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} [(\mathbf{Y}_1 - \mathbf{X}_1 \beta_1)^{\top} (\mathbf{Y}_1 - \mathbf{X}_1 \beta_1) + \\ &\quad (\mathbf{Y}_2 - \mathbf{X}_2 \beta_2)^{\top} (\mathbf{Y}_2 - \mathbf{X}_2 \beta_2)] \quad (4.7) \end{aligned}$$

e os EMVs são:

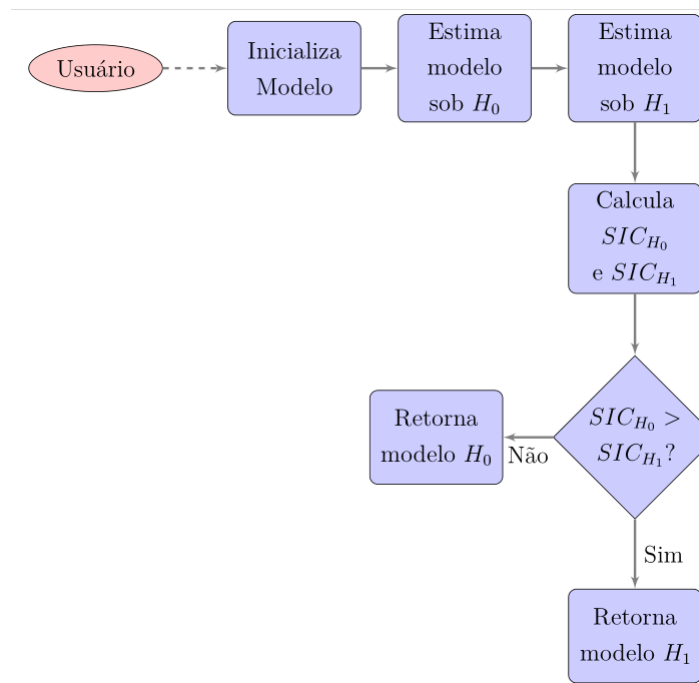
$$\begin{aligned}\widehat{\beta}_1 &= (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{Y}_1 \\ \widehat{\beta}_2 &= (\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{Y}_2 \\ \widehat{\sigma}^2 &= \frac{1}{n} [(\mathbf{Y}_1 - \mathbf{X}_1 \widehat{\beta}_1)^\top (\mathbf{Y}_1 - \mathbf{X}_1 \widehat{\beta}_1) + (\mathbf{Y}_2 - \mathbf{X}_2 \widehat{\beta}_2)^\top (\mathbf{Y}_2 - \mathbf{X}_2 \widehat{\beta}_2)].\end{aligned}$$

O SIC, sob H_1 , se torna:

$$SIC(k) = n \{ \log [(\mathbf{Y}_1 - \mathbf{X}_1 \widehat{\beta}_1)^\top (\mathbf{Y}_1 - \mathbf{X}_1 \widehat{\beta}_1) + (\mathbf{Y}_2 - \mathbf{X}_2 \widehat{\beta}_2)^\top (\mathbf{Y}_2 - \mathbf{X}_2 \widehat{\beta}_2)] + (\log 2\pi + 1) \} + (2s + 1 - n) \log n.$$

Com a implementação do SIC, o teste (4.1) se torna uma seleção de modelos. Podemos, então, determinar uma regra de decisão para o teste (4.1): selecionar o modelo sem ponto de mudança (aceitar H_0) se $SIC(n) < SIC(k) \forall k = 2, \dots, n-2$; ou selecionar o modelo com ponto de mudança na posição \hat{k} se $SIC(\hat{k}) = \min_{2 \leq k \leq n-2} SIC(k) < SIC(n)$, isto é, se o menor dos valores de SIC calculados para todos os k entre $2, \dots, n-2$ for menor que $SIC(n)$, selecionamos o modelo com ponto de mudança na posição \hat{k} . O fluxograma na Figura 25 ilustra o processo de decisão.

Figura 25 – Fluxograma representando o processo de decisão.

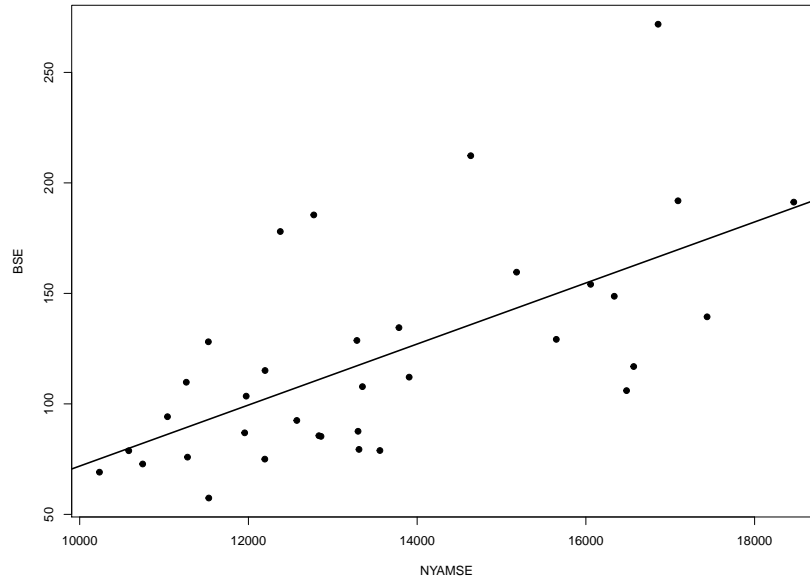


Fonte: Elaborado pelo autor.

Com as estimativas obtidas, podemos proceder ao processo de seleção de modelos para o conjunto de dados das bolsas de Boston e Nova York. Sem a suposição de ponto

de mudança, temos na Figura 26 a reta estimada pelo método de mínimos quadrados, enquanto na Figura 27 temos as retas estimadas supondo um ponto de mudança. Na mesma figura, temos evidenciada a observação 23, a qual foi identificada como sendo o ponto de mudança.

Figura 26 – Diagrama de dispersão da variável BSE versus NYAMSE com a reta estimada por mínimos quadrados.



Fonte: Elaborado pelo autor.

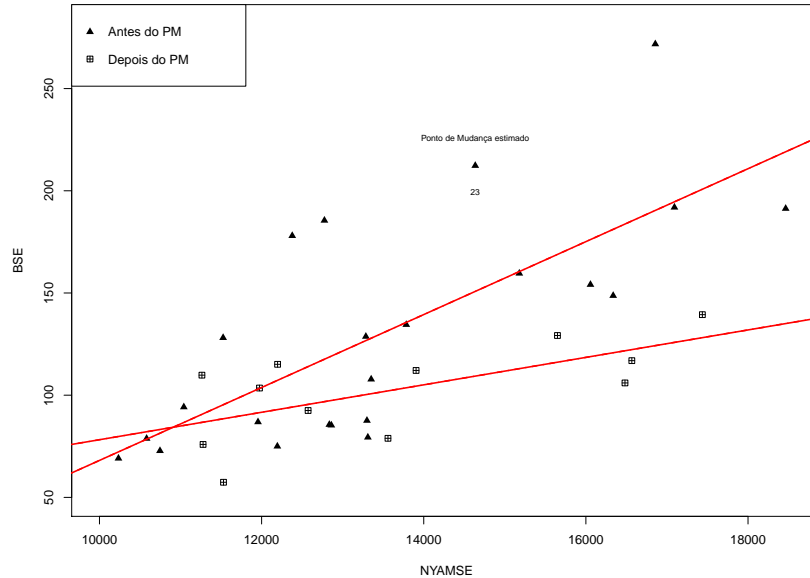
Tabela 4 – Estimativas dos coeficientes para os modelos sem e com ponto de mudança.

<i>Sem Ponto de Mudança</i>		<i>Com Ponto de Mudança</i>	
Estimativas		Estimativas	
$\widehat{\beta}_0$	-66,22	$\widehat{\beta}_0^1$	-110,31
$\widehat{\beta}_1$	0,01	$\widehat{\beta}_1^1$	0,02
$\widehat{\sigma}^2$	1400,59	$\widehat{\beta}_0^2$	11,07
		$\widehat{\beta}_1^2$	0,01
		$\widehat{\sigma}^2$	980,49
<i>SIC</i>	357,94	<i>SIC</i>	351,07

Fonte: Elaborado pelo autor.

A Tabela 4 apresenta os coeficientes estimados para o conjunto de dados das bolsas de valores de Boston e Nova York. Podemos notar que o SIC calculado para o modelo considerando ponto de mudança foi menor do que para o modelo que não considera a

Figura 27 – Diagrama de dispersão da variável BSE versus NYAMSE. Triângulo representa os pontos *antes* do ponto de mudança e quadrado representa os pontos *depois* da mudança.



Fonte: Elaborado pelo autor.

mudança. Assim, é preferível escolhermos o modelo que considera o ponto de mudança do que o modelo que não o faz.

Os resultados que encontramos são coerentes com outros trabalhos que exploraram este conjunto de dados, tais como [7] e [13].

Ajustado o modelo, poderíamos pensar em realizar uma análise de resíduos para detectar possível autocorrelação. Tal análise será feita seção seguinte.

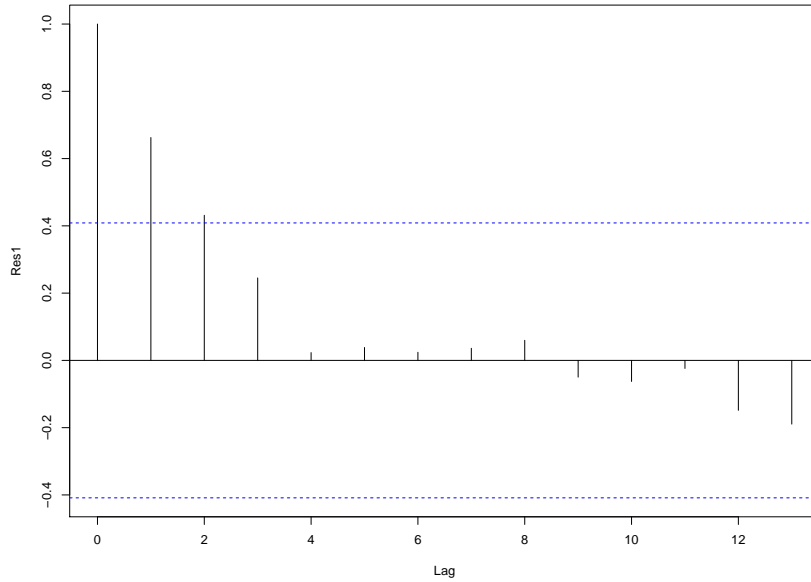
4.2 MODELO DE REGRESSÃO LINEAR SIMPLES COM PONTO DE MUDANÇA E ERRO NORMAL AUTOREGRESSIVO DE ORDEM (p)

Dados como os das bolsas de valores de Boston e Nova York, devido à sua natureza econômica, usualmente são modelados como sendo séries temporais, isto é, possuem algum tipo de estrutura de autocorrelação adjacente. Logo, é natural suspeitarmos que os resíduos do ajuste realizado na Seção 4.1 sejam autocorrelacionados. As Figuras 28 e 29 apresentam, respectivamente, as funções de autocorrelação e autocorrelação parcial para os resíduos *antes* do ponto de mudança identificado na seção anterior. Ambas indicam uma possível autocorrelação nos resíduos e a FACP indica um processo AR(1).

Já as Figuras 30 e 31 apresentam, respectivamente, as funções de autocorrelação e

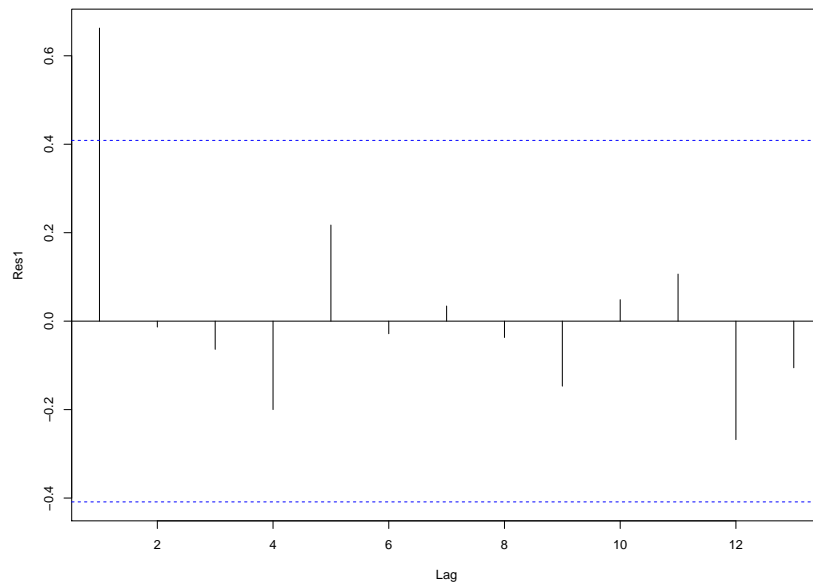
autocorrelação parcial para os resíduos *depois* do ponto de mudança identificado. Também indicam um processo autoregressivos, porém a FACP não nos dá certeza sobre a ordem do modelo.

Figura 28 – Função de autocorrelação dos resíduos da regressão *antes* do ponto de mudança.

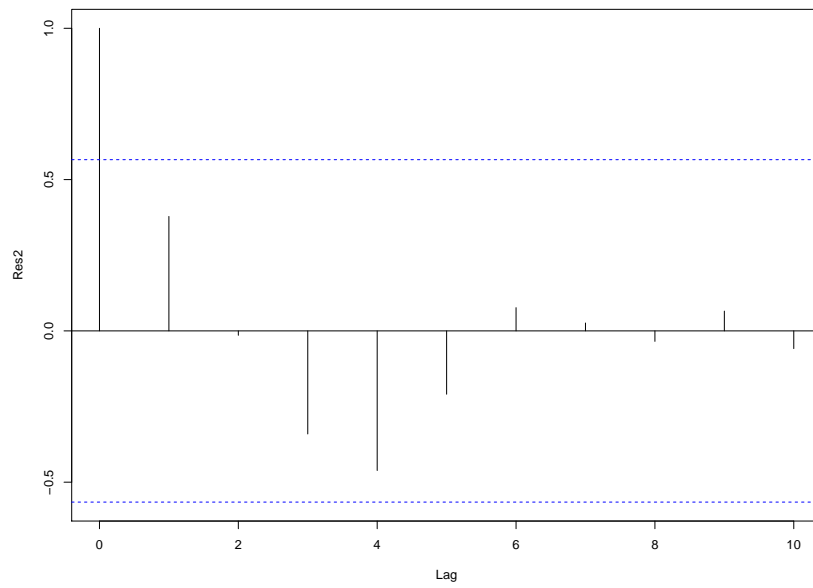


Fonte: Elaborado pelo autor.

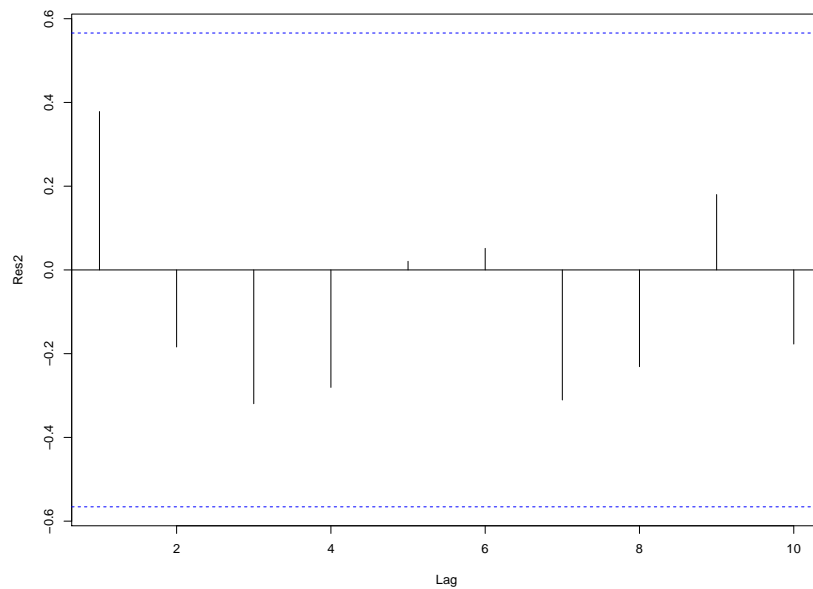
Figura 29 – Função de autocorrelação parcial dos resíduos da regressão *antes* do ponto de mudança.



Fonte: Elaborado pelo autor.

Figura 30 – Função de autocorrelação dos resíduos da regressão *depois* do ponto de mudança.

Fonte: Elaborado pelo autor.

Figura 31 – Função de autocorrelação parcial dos resíduos da regressão *depois* do ponto de mudança.

Fonte: Elaborado pelo autor.

Assim como fizemos no Capítulo 3, podemos aplicar o teste de Durbin-Watson aos resíduos. Caso o teste seja significativo, é possível que os resíduos sigam um processo AR(1). Para os resíduos antes e depois do ponto de mudança, realizamos o teste utilizando, novamente, a função `lmtest`, que nos dá o resultado apresentado na Tabela 5.

Tabela 5 – Resultado do teste de Durbin-Watson para presença de autocorrelação nos resíduos antes e depois do ponto de mudança, respectivamente.

	<i>Antes</i> do ponto de mudança	<i>Depois</i> do ponto de mudança
d	0,55223	1,0638
$p - \text{valor}$	$< 10^{-16}$	$< 10^{-16}$

Fonte: Elaborado pelo autor.

Confirmada a presença de autocorrelação, devemos proceder com a reestimação do modelo de regressão linear simples com ponto de mudança, considerando agora a presença de um processo autoregressivo. Apresentaremos a estimação do modelo para um processo AR(p) e depois iremos particularizar para o nosso exemplo das bolsas de valores de Boston e Nova York.

4.2.1 Estimação do modelo de regressão linear simples com ponto de mudança e erro normal autoregressivos de ordem p

Considere uma amostra aleatória de tamanho n , isto é, y_i , $i = 1, \dots, n$, de um modelo de regressão linear simples com ponto de mudança tal que

$$y_i = \begin{cases} \beta_0^1 + \beta_1^1 x_i + \epsilon_i, & \text{para } i = 1, \dots, k, \\ \beta_0^2 + \beta_1^2 x_i + \epsilon_i, & \text{para } i = k + 1, \dots, n, \end{cases} \quad (4.8)$$

em que o ponto de mudança k , $k = 2, \dots, n - 2$, é desconhecido, x_i , $i = 1, \dots, n$, é uma variável não estocástica, β_0^1 , β_1^1 , β_0^2 , β_1^2 , são os parâmetros da regressão antes (sobrescrito 1) e depois (sobrescrito 2) do ponto de mudança e ϵ_i é o erro tal que

$$\epsilon_i = \phi_1 \epsilon_{i-1} + \dots + \phi_p \epsilon_{i-p} + \eta_i, \quad \eta_i \sim N(0, \sigma^2),$$

isto é, o erro ϵ_i é proveniente de um processo autoregressivos de ordem p e η_i é um ruído branco que segue uma distribuição normal com valor esperado zero e variância σ^2 desconhecida.

Em notação matricial, o modelo (4.8) se torna

$$\mathbf{Y} = \begin{cases} \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_1, \\ \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2, \end{cases} \quad (4.9)$$

em que $\mathbf{Y} = [\mathbf{Y}_1^\top, \mathbf{Y}_2^\top]^\top$, $\mathbf{X}_1, \mathbf{X}_2$ são definidos como em (4.6), além disso, $\boldsymbol{\epsilon}_1 = (\epsilon_1, \dots, \epsilon_k)^\top$ e $\boldsymbol{\epsilon}_2 = (\epsilon_{k+1}, \dots, \epsilon_n)^\top$.

Para a estimação dos parâmetros EM (4.9), iremos novamente proceder com o teste de hipóteses realizado em (4.1), isto é, procederemos por uma abordagem de critérios de seleção, considerando a autocorrelação, e o modelo escolhido será aquele que minimizar o Critério de Informação de Schwarz, o SIC.

Sendo assim, trabalharemos com dois modelos, um sob a hipótese nula H_0 , que não considera o ponto de mudança, e um outro modelo sob a hipótese alternativa H_1 , que considera a mudança. Para o modelo sob H_0 , a estimação será semelhante àquela apresentada no Capítulo 3, já que o modelo apresentado não possuía ponto de mudança. Já para o modelo sob H_1 , procederemos com uma combinação das técnicas apresentadas no Capítulo 3 com as técnicas da Seção 4.1.

Sob H_0 , o modelo tem a seguinte log-verossimilhança exata

$$\ell(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top M_n^{-1}(\boldsymbol{\phi}) (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - \log |M_n^{-1}(\boldsymbol{\phi})|, \quad (4.10)$$

em que $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \sigma^2)^\top$ é o vetor de parâmetros do modelo e \mathbf{Y} e \mathbf{X} são definidas como em (3.2). Reparametrizando com (3.4), a log-verossimilhança é reescrita da seguinte forma:

$$\ell(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} S(\boldsymbol{\pi}, \boldsymbol{\beta}) - \frac{1}{2} \log g_p, \quad (4.11)$$

em que $\boldsymbol{\pi} = (\pi_1, \dots, \pi_p)^\top$, $g_p = |M_n(\boldsymbol{\phi})| = |M_n(\boldsymbol{\pi})| = \prod_{j=1}^p (1 - \pi_j^2)^{-j}$ e

$$S(\boldsymbol{\pi}, \boldsymbol{\beta}) = \boldsymbol{\lambda}^\top D(\mathbf{y}, \boldsymbol{\beta}) \boldsymbol{\lambda},$$

e $\boldsymbol{\lambda}^\top = (-1, \boldsymbol{\phi}^\top) = (-1, \mathcal{B}(\boldsymbol{\pi})^\top)$.

Os EMVs de $\boldsymbol{\beta}$ e σ^2 , dado $\boldsymbol{\pi}$, serão

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\boldsymbol{\pi}) &= (\mathbf{X}^\top M_n^{-1}(\mathcal{B}(\boldsymbol{\pi})) \mathbf{X})^{-1} \mathbf{X}^\top M_n^{-1}(\mathcal{B}(\boldsymbol{\pi})) \mathbf{y} \\ \hat{\sigma}^2(\boldsymbol{\pi}) &= \frac{1}{n} S(\boldsymbol{\pi}, \hat{\boldsymbol{\beta}}(\boldsymbol{\pi})). \end{aligned}$$

Substituindo os estimadores $\hat{\boldsymbol{\beta}}(\boldsymbol{\pi})$ e $\hat{\sigma}^2$ e eliminando termos constantes, temos a função de log-verossimilhança concentrada

$$\ell_c(\boldsymbol{\pi}|\mathbf{y}) = -\frac{n}{2} \log S(\boldsymbol{\pi}, \hat{\boldsymbol{\beta}}(\boldsymbol{\pi})) - \frac{1}{2} \log g_p,$$

que deve ser maximizada numericamente em relação à $\boldsymbol{\pi} \in (-1, 1)^p$, o cubo p -dimensional com superfícies de fronteira correspondendo a ± 1 , obtendo $\hat{\boldsymbol{\pi}}$. Finalmente, os EMVs de $\boldsymbol{\beta}$ e σ^2 são calculados como $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\pi}})$ e $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{\boldsymbol{\pi}})$. Podemos também obter estimativas para $\boldsymbol{\phi}$ como sendo $\hat{\boldsymbol{\phi}} = \mathcal{B}(\hat{\boldsymbol{\pi}})$.

O SIC sob H_0 será

$$SIC(n) = n \left[\log(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top M_n^{-1}(\hat{\boldsymbol{\phi}})(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + 1 + \log 2\pi \right] + (s + p + 1 - n) \log n + 2 \log |M_n(\hat{\boldsymbol{\phi}})|, \quad (4.12)$$

em que s é o número de parâmetros da regressão (para o modelo de regressão linear simples são 3 parâmetros no caso sem ponto de mudança, 2 devidos aos coeficientes da reta e 1 devido à variância do erro), p é a ordem do processo AR. Contudo, devido à singularidade de $M_n(\hat{\boldsymbol{\phi}})$, devemos reparametrizar (4.12) como sendo

$$SIC(n) = n \left[\log(\boldsymbol{\lambda}^\top D(\mathbf{Y}, \hat{\boldsymbol{\beta}})\boldsymbol{\lambda}) + 1 + \log 2\pi \right] + (s + p + 1 - n) \log n + 2 \log |M_n(\hat{\boldsymbol{\phi}})|. \quad (4.13)$$

Já sob H_1 , isto é, sob o contexto de ponto de mudança, temos a seguinte log-verossimilhança

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{Y}) &= -\frac{k}{2} \log 2\pi - \frac{k}{2} \log \sigma^2 - \frac{1}{2} \log |M_k(\boldsymbol{\phi}_1)| - \frac{1}{2\sigma^2} (\mathbf{Y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1)^\top M_k^{-1}(\boldsymbol{\phi}_1) (\mathbf{Y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1) \\ &\quad - \frac{n-k}{2} \log 2\pi - \frac{n-k}{2} \log \sigma^2 - \frac{1}{2} \log |M_{n-k}(\boldsymbol{\phi}_2)| \\ &\quad - \frac{1}{2\sigma^2} (\mathbf{Y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2)^\top M_{n-k}^{-1}(\boldsymbol{\phi}_2) (\mathbf{Y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2) \\ &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \log |M_k(\boldsymbol{\phi}_1)| - \frac{1}{2\sigma^2} (\mathbf{Y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1)^\top M_k^{-1}(\boldsymbol{\phi}_1) (\mathbf{Y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1) \\ &\quad - \frac{1}{2} \log |M_{n-k}(\boldsymbol{\phi}_2)| - \frac{1}{2\sigma^2} (\mathbf{Y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2)^\top M_{n-k}^{-1}(\boldsymbol{\phi}_2) (\mathbf{Y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2). \end{aligned} \quad (4.14)$$

Reparametrizando (4.14) com (3.4), a menos de uma constante, temos

$$\ell(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{k}{2} \log \sigma^2 - \frac{1}{2\sigma^2} S(\boldsymbol{\pi}_1, \boldsymbol{\beta}_1) - \frac{1}{2} \log g_{p_1} - \frac{n-k}{2} \log \sigma^2 - \frac{1}{2\sigma^2} S(\boldsymbol{\pi}_2, \boldsymbol{\beta}_2) - \frac{1}{2} \log g_{p_2}, \quad (4.15)$$

em que $S(\boldsymbol{\pi}_d, \boldsymbol{\beta}_d) = \boldsymbol{\lambda}_d^\top D(\mathbf{Y}_d, \boldsymbol{\beta}_d)\boldsymbol{\lambda}_d$, é definida como em (3.7), $\boldsymbol{\lambda}_d^\top = (-1, \boldsymbol{\phi}_d^\top)$, $d = 1, 2$, $g_{p_1} = |M_k(\boldsymbol{\phi}_1)| = \prod_{j_1=1}^p (1 - \pi_{j_1}^2)^{-j_1}$, $g_{p_2} = |M_{n-k}(\boldsymbol{\phi}_2)| = \prod_{j_2=1}^p (1 - \pi_{j_2}^2)^{-j_2}$.

Uma importante suposição que estamos fazendo para a log-verossimilhança sob a hipótese de ponto de mudança é que a ordem do processo autoregressivos é a mesma *antes* e *depois* do ponto de mudança. Com isso, os índices p_1 e p_2 , assim como j_1 e j_2 , em (4.15) são apenas para diferenciar de qual partição do modelo eles referem.

A log-verossimilhança (4.15) tem máximo em

$$\begin{aligned}
\widehat{\beta}_1(\mathcal{B}(\boldsymbol{\pi}_1)) &= (\mathbf{X}_1^\top M_k^{-1}(\mathcal{B}(\boldsymbol{\pi}_1)) \mathbf{X}_1)^{-1} \mathbf{X}_1^\top M_k^{-1}(\mathcal{B}(\boldsymbol{\pi}_1)) \mathbf{Y}_1, \\
\widehat{\beta}_2(\mathcal{B}(\boldsymbol{\pi}_2)) &= (\mathbf{X}_2^\top M_{n-k}^{-1}(\mathcal{B}(\boldsymbol{\pi}_2)) \mathbf{X}_2)^{-1} \mathbf{X}_2^\top M_{n-k}^{-1}(\mathcal{B}(\boldsymbol{\pi}_2)) \mathbf{Y}_2, \\
\widehat{\sigma}^2(\mathcal{B}(\boldsymbol{\pi}_1), \mathcal{B}(\boldsymbol{\pi}_2)) &= \frac{1}{n} [(\mathbf{Y}_1 - \mathbf{X}_1 \widehat{\beta}_1(\mathcal{B}(\boldsymbol{\pi}_1)))^\top M_k^{-1}(\mathcal{B}(\boldsymbol{\pi}_1)) (\mathbf{Y}_1 - \mathbf{X}_1 \widehat{\beta}_1(\mathcal{B}(\boldsymbol{\pi}_1))) \\
&\quad + (\mathbf{Y}_2 - \mathbf{X}_2 \widehat{\beta}_2(\mathcal{B}(\boldsymbol{\pi}_2)))^\top M_{n-k}^{-1}(\mathcal{B}(\boldsymbol{\pi}_2)) (\mathbf{Y}_2 - \mathbf{X}_2 \widehat{\beta}_2(\mathcal{B}(\boldsymbol{\pi}_2)))] \\
&= \frac{1}{n} [S(\boldsymbol{\pi}_1, \widehat{\beta}_1(\mathcal{B}(\boldsymbol{\pi}_1))) + S(\boldsymbol{\pi}_2, \widehat{\beta}_2(\mathcal{B}(\boldsymbol{\pi}_2)))] \tag{4.16}
\end{aligned}$$

Substituindo os estimadores (4.16) em (4.15), também obtemos a log-verossimilhança concentrada sob H_1

$$\ell_c(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2 | \mathbf{Y}) = -\frac{n}{2} \left[\log S(\boldsymbol{\pi}_1, \widehat{\beta}_1(\mathcal{B}(\boldsymbol{\pi}_1))) + \log S(\boldsymbol{\pi}_2, \widehat{\beta}_2(\mathcal{B}(\boldsymbol{\pi}_2))) \right] - \frac{1}{2} \log g_{p_1} - \frac{1}{2} \log g_{p_2}$$

que ao ser maximizada numericamente em relação à $\boldsymbol{\pi}_d \in (-1, 1)^{p_d}$, obtemos as estimativas de $\boldsymbol{\pi}_1$ e $\boldsymbol{\pi}_2$. Assim como no caso sem ponto de mudança, os EMVs de β_d , e $\widehat{\sigma}^2$ são calculados como $\widehat{\beta}_d = \widehat{\beta}_d(\widehat{\boldsymbol{\pi}}_d)$, $d = 1, 2$, e $\widehat{\sigma}^2 = \widehat{\sigma}^2(\widehat{\boldsymbol{\pi}}_1, \widehat{\boldsymbol{\pi}}_2)$. Para as estimativas de ϕ_1 e ϕ_2 , basta tomarmos $\widehat{\phi}_1 = \mathcal{B}(\phi_1)$ e $\widehat{\phi}_2 = \mathcal{B}(\phi_2)$.

Sob H_1 , o SIC se torna

$$\begin{aligned}
SIC(k) &= n \{ [1 + \log 2\pi] + \log [(\mathbf{Y}_1 - \mathbf{X}_1 \widehat{\beta}_1)^\top M_k^{-1}(\widehat{\phi}_1) (\mathbf{Y}_1 - \mathbf{X}_1 \widehat{\beta}_1) \\
&\quad + (\mathbf{Y}_2 - \mathbf{X}_2 \widehat{\beta}_2)^\top M_{n-k}^{-1}(\widehat{\phi}_2) (\mathbf{Y}_2 - \mathbf{X}_2 \widehat{\beta}_2)] \} + \log |M_k^{-1}(\widehat{\phi}_1)| + \log |M_{n-k}^{-1}(\widehat{\phi}_2)| \\
&\quad + [2(s+p) + 1 - n] \log n. \tag{4.17}
\end{aligned}$$

Como $M_k(\widehat{\phi}_1)$ e $M_{n-k}(\widehat{\phi}_2)$ também têm problemas de singularidade, devemos reparametrizar (4.17)

$$\begin{aligned}
SIC(k) &= n \{ [1 + \log 2\pi] + \log [\boldsymbol{\lambda}_1^\top D(\mathbf{Y}_1, \widehat{\beta}_1) \boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2^\top D(\mathbf{Y}_2, \widehat{\beta}_2) \boldsymbol{\lambda}_2] \} + \log |M_k^{-1}(\widehat{\phi}_1)| \\
&\quad + \log |M_{n-k}^{-1}(\widehat{\phi}_2)| + [2(s+p) + 1 - n] \log n. \tag{4.18}
\end{aligned}$$

Finalmente, podemos proceder ao processo de decisão que, assim como na Seção 4.1, será uma escolha de modelos. Isto é, selecionamos o modelo sem ponto de mudança se $SIC(n) < SIC(k) \forall k = s+p, \dots, n-s-p$; ou selecionamos o modelo com ponto de mudança na posição \hat{k} se $SIC(\hat{k}) = \min_{s+p \leq k \leq n-s-p} SIC(k) < SIC(n)$. Escolhemos o intervalo de $k \in [s+p, n-s-p]$ para que tenhamos um número mínimo de observações em cada partição feita $(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{X}_1$ e $\mathbf{X}_2)$ de forma que possamos estimar os parâmetros para cada partição.

Retornando ao exemplo das bolsas de valores de Boston e Nova York, a Tabela 6 compara os resultados para os casos sem ponto de mudança, com ponto de mudança e com ponto de mudança e erro AR(1). Em um primeiro momento, notamos que os coeficientes

da regressão são bem distintos entre os dois últimos casos; além disso, ao considerarmos o erro como sendo autoregressivos de ordem 1, a variância é relativamente menor do que quando comparada aos outros casos. O ponto de mudança identificado foi o mesmo em ambos os modelos que o consideraram. A Figura 32 apresenta as retas estimadas em ambos os casos que consideraram o ponto de mudança.

Um questionamento que poderia surgir seria: qual o melhor modelo? Para respondê-lo, podemos fazer uso do critério de informação. Como o SIC é uma ferramenta que possibilita a escolha de modelos, escolheremos aquele que apresenta o menor valor de SIC dentre os valores calculados, um processo semelhante ao que foi feito na detecção do ponto de mudança. Assim, para o conjunto de dados das bolsas de valores de Boston e Nova York, escolhemos o modelo que leva em consideração a presença do ponto de mudança e o erro normal autoregressivos de ordem 1.

Tabela 6 – Estimativas dos coeficientes para os modelos sem ponto de mudança, com ponto de mudança e com ponto de mudança e erro AR(1).

<i>Sem Ponto de Mudança</i>	<i>Com Ponto de Mudança</i>	<i>Com Ponto de Mudança AR(1)</i>
Estimativas	Estimativas	Estimativas
$\widehat{\beta}_0$ -66,22	$\widehat{\beta}_0^1$ -110,31	$\widehat{\beta}_0^1$ -0,047
$\widehat{\beta}_1$ 0,01	$\widehat{\beta}_1^1$ 0,02	$\widehat{\beta}_1^1$ 0,010
$\widehat{\sigma}^2$ 1400,59	$\widehat{\beta}_0^2$ 11,07	$\widehat{\beta}_0^2$ 21,103
	$\widehat{\beta}_1^2$ 0,01	$\widehat{\beta}_1^2$ 0,006
	$\widehat{\sigma}^2$ 980,49	$\widehat{\phi}_1^1$ 0,880
	\widehat{k} 23	$\widehat{\phi}_1^2$ 0,442
		$\widehat{\sigma}^2$ 352,582
		\widehat{k} 23
<i>SIC</i> 357,94	<i>SIC</i> 351,07	<i>SIC</i> 331,206

Fonte: Elaborado pelo autor.

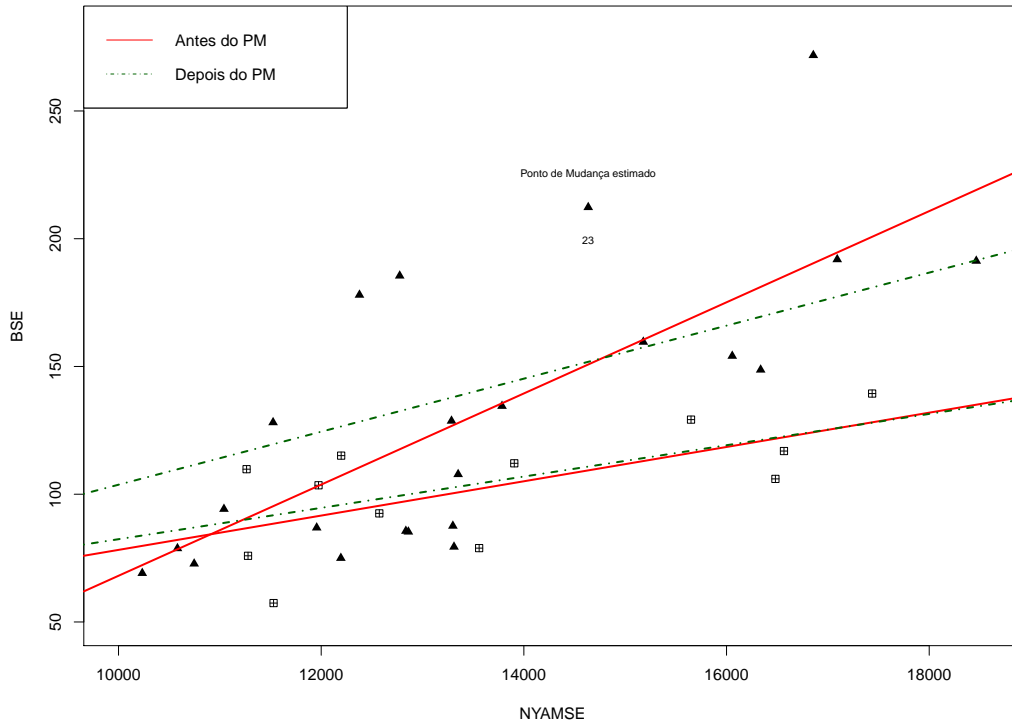
4.3 ESTUDO DE SIMULAÇÃO

O objetivo desta seção é apresentarmos a eficiência dos estimadores obtidos na Seção 4.2. Testamos nosso modelo em diferentes cenários mudando o tamanho n da amostra e a posição do ponto de mudança. Consideramos também cenários em que o ponto de mudança é suposto conhecido e também desconhecido.

4.3.1 Ponto de mudança conhecido

Aqui simulamos o modelo com diferença em todos os parâmetros da regressão e diferença em todos os coeficientes autoregressivos. Mantemos a ordem dos coeficientes

Figura 32 – Diagrama de dispersão da variável BSE versus NYAMSE. Triângulo representa os pontos *antes* do ponto de mudança e quadrado representa os pontos *depois* da mudança. As linhas tracejadas em verde representam as retas estimadas considerando o erro autoregressivos e as linhas contínuas em vermelho representam as retas estimadas considerando o erro normal.



Fonte: Elaborado pelo autor.

autoregressivos antes e depois do ponto de mudança.

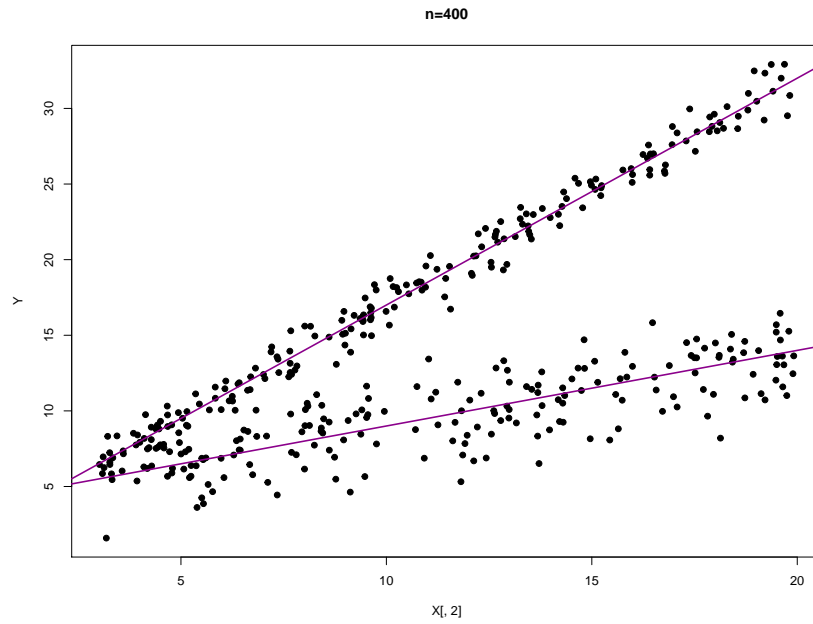
Simulamos o seguinte modelo

$$\begin{cases} y_i = 2.0 + 0.5x_i + \epsilon_i, & i = 1, \dots, k \\ \epsilon_i = 0.5\epsilon_{i-1} + 0.7\epsilon_{i-2} + \eta_i, & \eta_i \sim N(0, 0.4624), \quad i = 1, \dots, k \\ y_i = 4.0 + 1.5x_i + \epsilon_i, & i = k + 1, \dots, n \\ \epsilon_i = 0.8\epsilon_{i-1} + 0.9\epsilon_{i-2} + \eta_i, & \eta_i \sim N(0, 0.4624), \quad i = k + 1, \dots, n \end{cases} \quad (4.19)$$

Para a simulação do modelo (4.19), geramos uma amostra aleatória uniforme para a variável explicativa no intervalo $X_i \sim U(3, 20)$. A Figura 33 é um exemplo do modelo simulado (4.19) com a reta real.

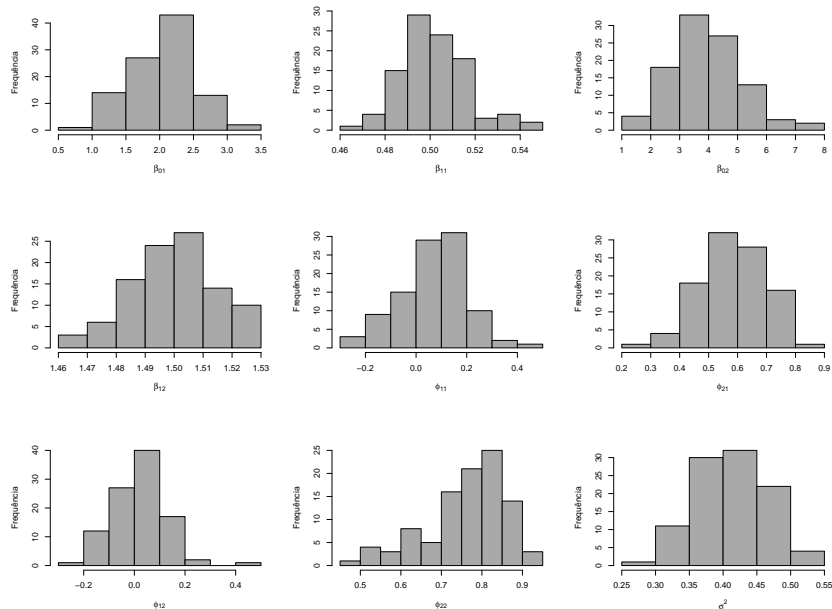
As Figuras 34 a 40 apresentam as frequências dos estimadores para as simulações feitas com diferentes tamanhos amostrais e 100 replicações. Estas simulações tem como objetivo mostrar a eficiência dos estimadores, evidenciando que, em média, o processo de estimação está correto.

Figura 33 – Diagrama de dispersão da variável dos dados simulados com a reta real evidenciada.



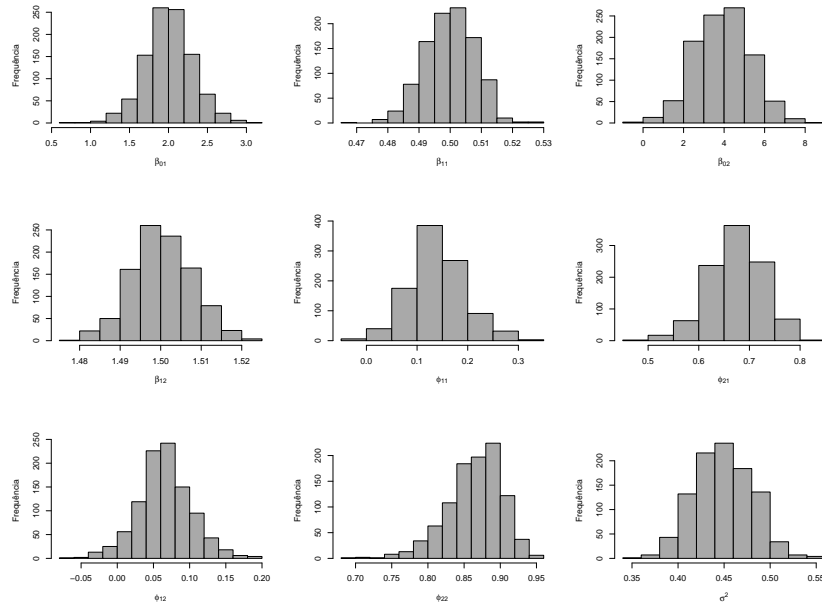
Fonte: Elaborado pelo autor.

Figura 34 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança na mediana dos dados, mudança em todos os parâmetros da regressão e nos coeficientes autoregressivos, $n = 100$.



Fonte: Elaborado pelo autor.

Figura 35 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança na mediana dos dados, mudança em todos os parâmetros da regressão e nos coeficientes autoregressivos, $n = 400$.



Fonte: Elaborado pelo autor.

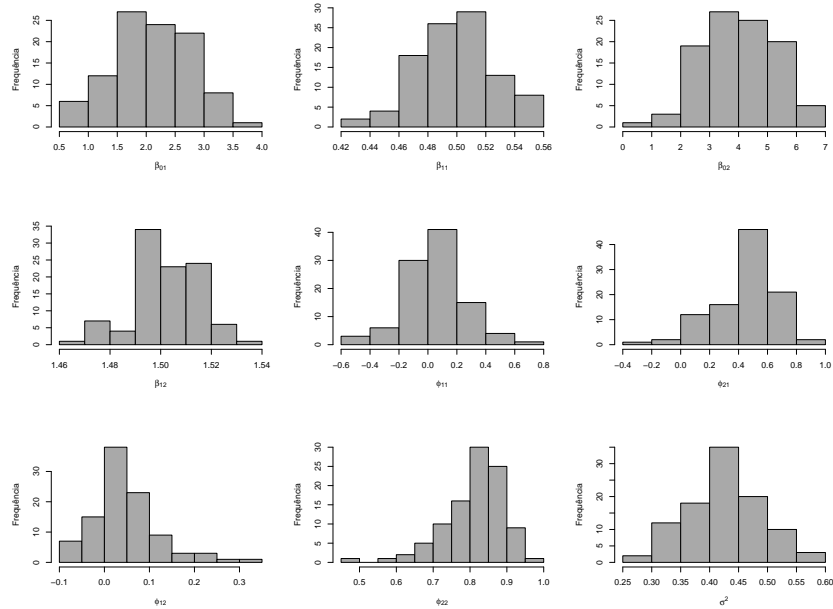
Tabela 7 – Comparação entre simulações de modelos de regressão linear simples com ponto de mudança e erro autoregressivo AR(2). Pontos de mudança na mediana dos dados e no primeiro quartil, com variação no tamanho da amostra. Em parêntesis, o desvio padrão amostral das simulações.

Estimativas	$n = 100$		$n = 400$		Real
	Mediana	1º Quartil	Mediana	1º Quartil	
$\hat{\beta}_0^1$	2,055(0,496)	2,110(0,681)	2,004(0,302)	2,050(0,463)	2
$\hat{\beta}_1^1$	0,502(0,015)	0,499(0,027)	0,500(0,008)	0,501(0,013)	0,5
$\hat{\beta}_0^2$	3,964(1,181)	4,009(1,260)	3,944(1,342)	4,016(1,340)	4
$\hat{\beta}_1^2$	1,500(0,015)	1,502(0,013)	1,500(0,007)	1,500(0,006)	1,5
$\hat{\phi}_1^1$	0,068(0,132)	0,054(0,210)	0,139(0,054)	0,118(0,083)	0,15
$\hat{\phi}_2^1$	0,584(0,112)	0,455(0,222)	0,673(0,053)	0,654(0,074)	0,7
$\hat{\phi}_1^2$	0,020(0,108)	0,046(0,069)	0,066(0,037)	0,070(0,027)	0,08
$\hat{\beta}_2^2$	0,760(0,100)	0,811(0,081)	0,865(0,038)	0,875(0,026)	0,9
$\hat{\sigma}^2$	0,411(0,053)	0,428(0,065)	0,449(0,031)	0,452(0,034)	0,4624

Fonte: Elaborado pelo autor.

A Tabela 7 resume os resultados dos histogramas das figuras 34, 35, 36 e 37. A partir dela, percebemos que a posição do ponto de mudança afeta um pouco a eficiência da estimação. Porém, ao aumentarmos o tamanho da amostra, como era esperado, o

Figura 36 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança no primeiro quartil dos dados, mudança em todos os parâmetros da regressão e nos coeficientes autoregressivos, $n = 100$.



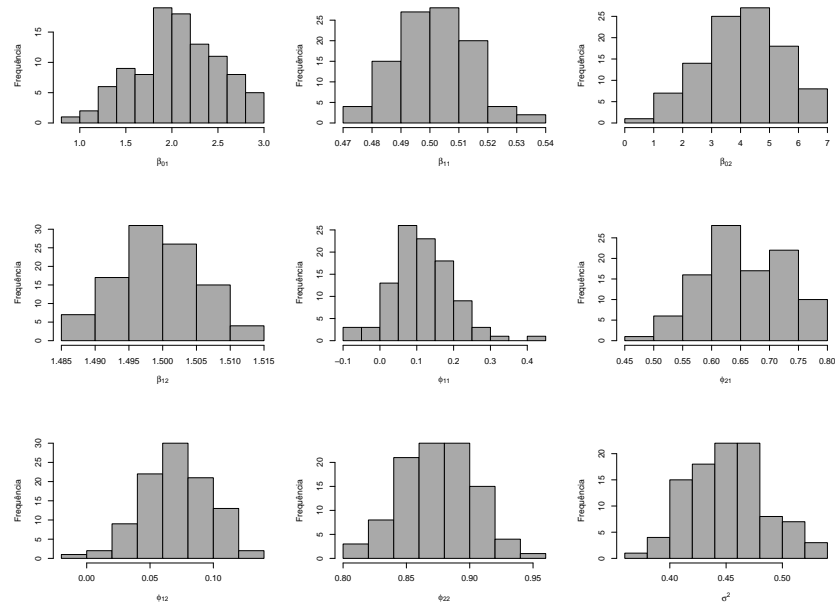
Fonte: Elaborado pelo autor.

estimador aparenta convergir para o valor real.

Já a Tabela 8 apresenta resultados para simulações com o mesmo intercepto, representados nas Figuras 38, 39 e 40. Este cenário parece afetar a estimação para o intercepto após o ponto de mudança (sobrescrito 2).

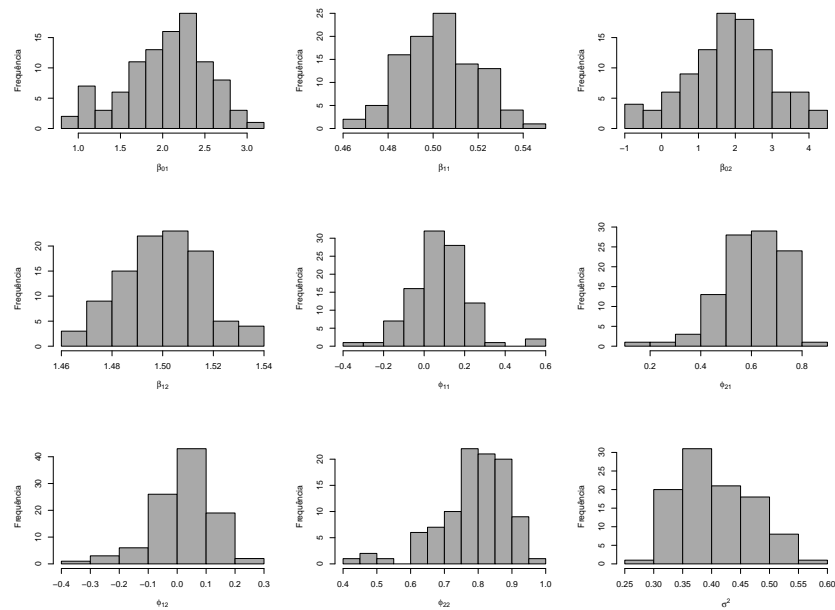
Finalizando o estudo de simulação com a posição do ponto de mudança conhecida, apresentamos uma última tabela com resultados para grandes amostras e grandes replicações. A Tabela 9 apresenta estes resultados. Note como os desvios padrão diminuem à medida que aumentamos o tamanho das amostras.

Figura 37 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança no primeiro quartil dos dados, mudança em todos os parâmetros da regressão e nos coeficientes autoregressivos, $n = 400$.



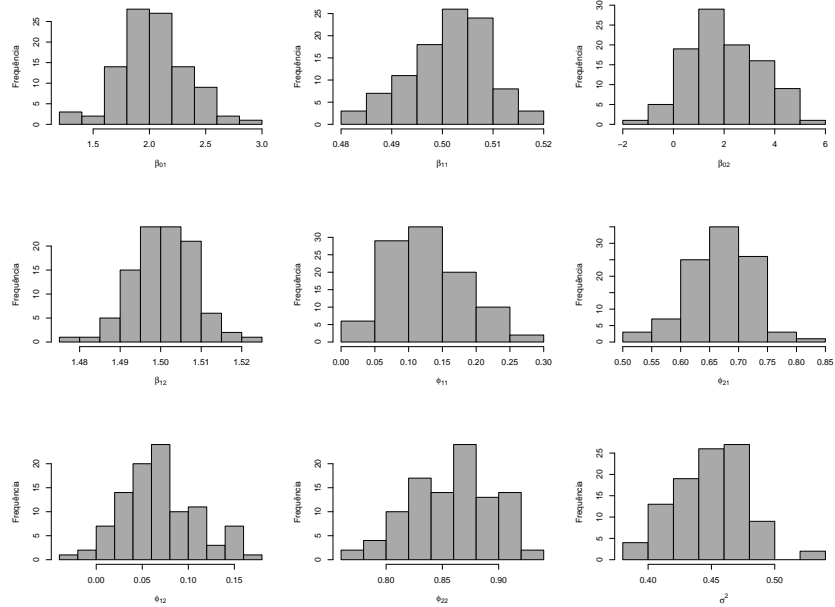
Fonte: Elaborado pelo autor.

Figura 38 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança na mediana dos dados, mesmo intercepto e mudança nos coeficientes autoregressivos, $n = 100$.



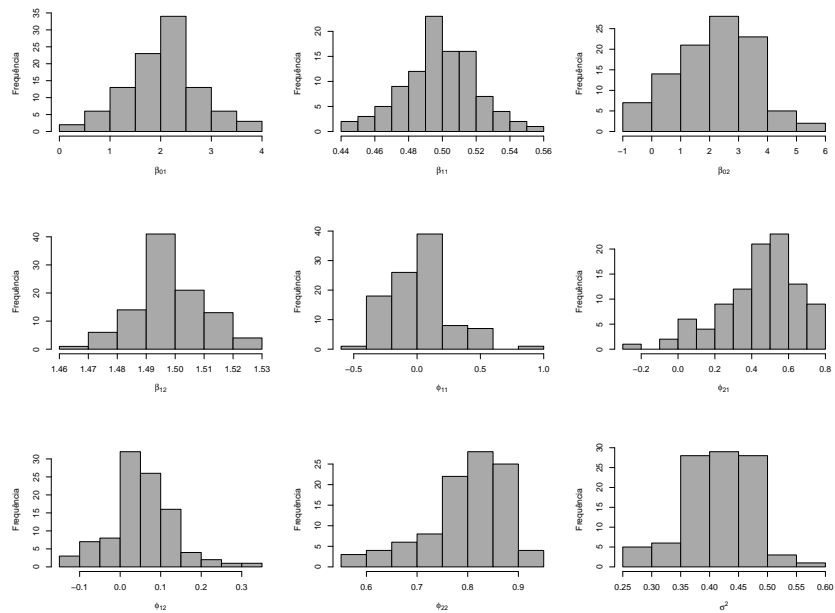
Fonte: Elaborado pelo autor.

Figura 39 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança na mediana dos dados, mesmo intercepto e mudança nos coeficientes autoregressivos, $n = 400$.



Fonte: Elaborado pelo autor.

Figura 40 – Histogramas com as frequências dos estimadores do modelo com ponto de mudança no primeiro quartil dos dados, mesmo intercepto e mudança nos coeficientes autoregressivos, $n = 100$.



Fonte: Elaborado pelo autor.

Tabela 8 – Comparação entre simulações de modelos de regressão linear simples com ponto de mudança e erro autoregressivo AR(2). Em parêntesis, o desvio padrão amostral das simulações. Para este caso, consideramos mesmo intercepto antes e depois do ponto de mudança.

Estimativas	Mediana		1° Quartil	Real
	$n = 100$	$n = 400$	$n = 100$	
$\hat{\beta}_0^1$	2,043(0,486)	2,033(0,292)	2,033(0,703)	2
$\hat{\beta}_1^1$	0,502(0,017)	0,501(0,008)	0,499(0,022)	0,5
$\hat{\beta}_0^2$	1,859(1,175)	1,999(1,377)	2,208(1,317)	2
$\hat{\beta}_1^2$	1,500(0,016)	1,501(0,008)	1,498(0,012)	1,5
$\hat{\phi}_1^1$	0,073(0,136)	0,129(0,056)	0,029(0,224)	0,15
$\hat{\phi}_2^1$	0,605(0,124)	0,669(0,053)	0,441(0,202)	0,7
$\hat{\phi}_1^2$	0,023(0,101)	0,068(0,040)	0,054(0,077)	0,08
$\hat{\beta}_2^2$	0,788(0,103)	0,859(0,037)	0,798(0,081)	0,9
$\hat{\sigma}^2$	0,407(0,061)	0,450(0,028)	0,416(0,060)	0,4624

Fonte: Elaborado pelo autor.

Tabela 9 – Resultados para simulação com 1000 replicações de cada um dos tamanhos amostrais abaixo. Em parêntesis, os desvios padrão amostrais.

	$n = 100$	$n = 400$	$n = 1000$	Real
$\hat{\beta}_0^1$	1,9919(0,5306)	2,0042(0,3016)	1,9816(0,2033)	2
$\hat{\beta}_1^1$	0,4998(0,0165)	0,4999(0,0080)	0,5002(0,0050)	0,5
$\hat{\beta}_0^2$	4,0411(1,1388)	3,9436(1,3416)	4,0015(1,1802)	4
$\hat{\beta}_1^2$	1,5001(0,0151)	1,5004(0,0074)	1,5000(0,0047)	1,5
$\hat{\phi}_1^1$	0,0981(0,1329)	0,1389(0,0537)	0,1440(0,0327)	0,15
$\hat{\phi}_2^1$	0,5906(0,1220)	0,6729(0,0531)	0,6902(0,0326)	0,7
$\hat{\phi}_1^2$	0,0292(0,1114)	0,0656(0,0369)	0,0760(0,0217)	0,08
$\hat{\beta}_2^2$	0,7603(0,1092)	0,8647(0,0376)	0,8837(0,0220)	0,9
$\hat{\sigma}^2$	0,4160(0,0632)	0,4493(0,0310)	0,4576(0,0212)	0,4624

Fonte: Elaborado pelo autor.

4.3.2 Ponto de mudança desconhecido

Para esta situação, apresentamos três diferentes cenários todos com tamanho de amostra $n = 400$ e número de replicações $l = 100$.

- Modelo com mesmo intercepto antes e depois do ponto de mudança com ponto de mudança no primeiro quartil, ilustrado pela Figura 41;
- Modelo com mesma inclinação antes e depois do ponto de mudança, isto é, duas retas paralelas com ponto de mudança na mediana ilustrado pela Figura 42,
- Modelo com retas concorrentes com ponto de mudança na mediana ilustrado pela Figura 43.

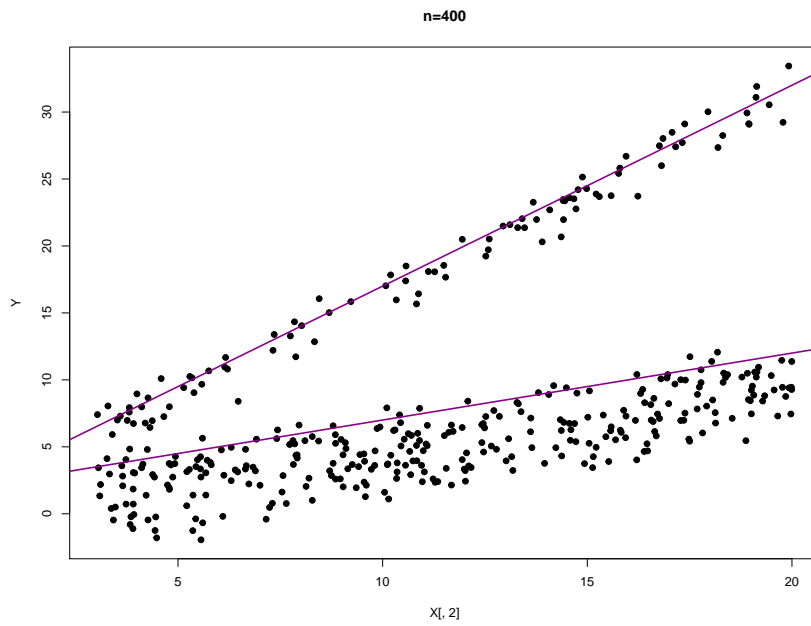
A Tabela 10 apresenta os resultados para os casos estudados. A Tabela 11 mostra os pontos de mudança identificados em cada caso, com a porcentagem de ocorrência de cada identificação. O procedimento desenvolvido por nós apresentou um resultado satisfatório, acertando a identificação em mais de 90% nos cenários simulados.

Tabela 10 – Resultados para modelos com mesmo intercepto, retas paralelas e retas concorrentes. Em parêntesis, os desvios-padrão amostrais.

	Mesmo Intercepto	Retas paralelas	Retas concorrentes
$\hat{\beta}_0^1$	2,054(0,435)	5,993(0,309)	20,045(0,307)
$\hat{\beta}_1^1$	0,499(0,012)	1,501(0,008)	-1,500(0,008)
$\hat{\beta}_0^2$	1,974(1,316)	2,058(1,360)	1,903(1,401)
$\hat{\beta}_1^2$	1,500(0,006)	1,500(0,007)	1,501(0,007)
$\hat{\phi}_1^1$	0,128(0,076)	0,137(0,053)	0,136(0,051)
$\hat{\phi}_2^1$	0,635(0,075)	0,668(0,053)	0,681(0,056)
$\hat{\phi}_1^2$	0,072(0,027)	0,069(0,037)	0,067(0,039)
$\hat{\beta}_2^2$	0,875(0,027)	0,865(0,032)	0,861(0,038)
$\hat{\sigma}^2$	0,449(0,025)	0,451(0,025)	0,450(0,032)

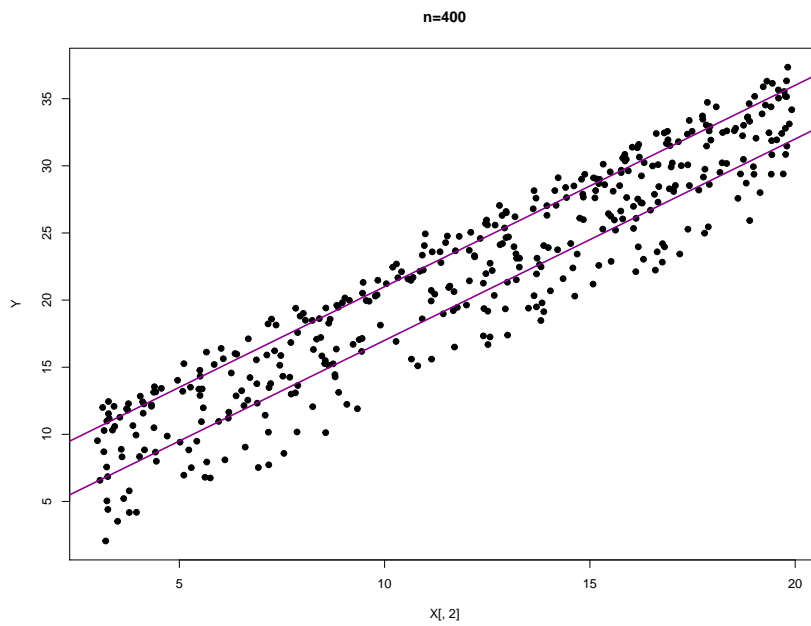
Fonte: Elaborado pelo autor.

Figura 41 – Diagrama de dispersão da variável dos dados simulados com a reta real evidenciada para o caso interceptos iguais e ponto de mudança no primeiro quartil.



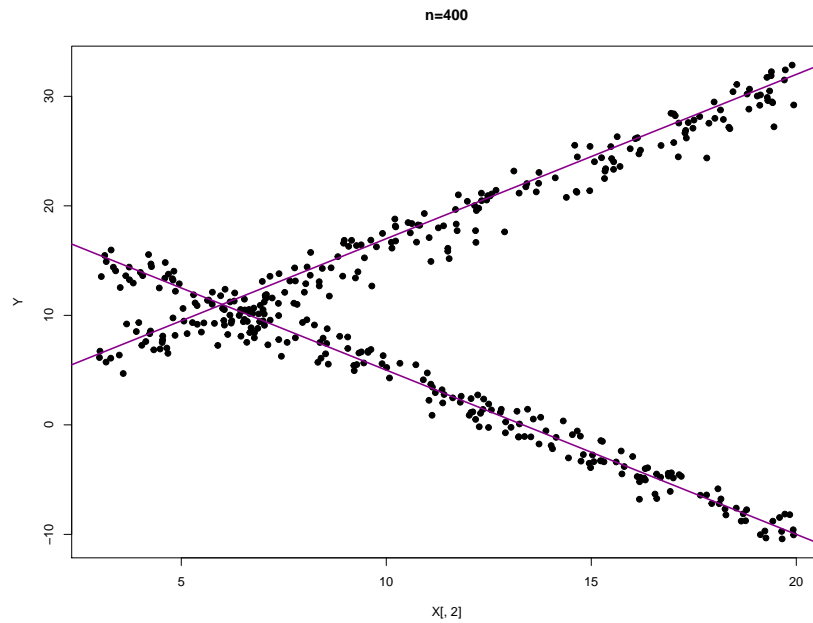
Fonte: Elaborado pelo autor.

Figura 42 – Diagrama de dispersão da variável dos dados simulados com a reta real evidenciada para o caso de retas paralelas e ponto de mudança na mediana.



Fonte: Elaborado pelo autor.

Figura 43 – Diagrama de dispersão da variável dos dados simulados com a reta real evidenciada para o caso de retas concorrentes e ponto de mudança na mediana.



Fonte: Elaborado pelo autor.

Tabela 11 – Posições do ponto de mudança identificadas e porcentagem de ocorrências para cada um dos três casos.

	Interceptos iguais ($k = 100$)	Retas paralelas ($k = 200$)	Retas concorrentes ($k = 200$)
\hat{k}	100	170	200
% ocorrências	100%	1%	99%

Fonte: Elaborado pelo autor.

5 Considerações Finais

Ao longo deste texto, procuramos apresentar uma abordagem para tratar de problemas que podem surgir durante uma análise estatística, em especial a identificação de pontos de mudança, e como lidar com autocorrelação em modelos de regressão linear simples e, é claro, quando há interseção entre estas duas situações. E, com isso, propor um novo modelo de regressão que lida com pontos de mudança e erros normais autoregressivos, com a suposição de que a ordem do processo autoregressivo se mantém após o ponto de mudança.

Acreditamos que a técnica proposta seja satisfatória, pelos resultados numéricos observados através de uma aplicação a dados reais e estudos de simulação. Em ambas as situações, nossa abordagem se mostrou mais capaz na estimação dos parâmetros, o que pôde ser observado na comparação dos critérios de informação que calculamos ao longo do Capítulo 4.

Sobre o Capítulo 3, para a detecção de ordem dos modelos AR, em [5, Subseção 6.2.4] há diferentes abordagens para identificação de modelos, dentre elas, critérios de informação. Em testes para a elaboração dos resultados apresentados neste texto, a utilização de critérios de seleção para a identificação da ordem do modelo se mostrou inviável. Especificamente para o banco de dados no Capítulo 3 (Companhia Blaisdell), foi identificada ordem 11 e quando utilizamos esta ordem na rotina no R que desenvolvemos ocorreram problemas de singularidade, devido às dimensões das matrizes que seriam calculadas.

Já no Capítulo 2, apresentamos uma revisão sobre modelos de regressão linear simples, com algumas extensões utilizando notação matricial. Essas extensões são necessárias quando lidamos com regressão linear múltipla, pois facilitam a obtenção dos estimadores de máxima verossimilhança e a implementação de algoritmos para estimação. Em relação aos modelos autoregressivos, apresentamos as principais definições sobre séries temporais e modelos autoregressivos.

Para trabalhos futuros, podemos considerar outras distribuições para os erros, como distribuições com caudas mais pesadas - t de Student, por exemplo. Ou utilizar métodos como o procedimento de Cochrane-Orcutt [16].

REFERÊNCIAS

- [1] AKAIKE, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**(6). Páginas 716-723.
- [2] AKASHI, F.; DETTE, H.; LIU, Y. (2018) Change-Point Detection in Autoregressive Models with no Moment Assumptions. *Journal of Time Series Analysis*. **39**(5). Páginas 763-786.
- [3] BARNFORFF-NIELSEN, O.; SCHOU, G. (1973) On the parametrization of autoregressive models by partial autocorrelations. *Journal of Multivariate Analysis*, **45**(4). Páginas 408-419.
- [4] BAI, J.(1997) Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, **79**(4). Páginas 551-563.
- [5] BOX, G.E.P.; JENKINS, G.M.; REINSEL, G.C. *Time series analysis: forecasting and control*. John Wiley & Sons, 2008.
- [6] CASELLA, G.; BERGER, R.L. *Inferência Estatística - tradução da 2ª edição norte-americana*. Cengage Learning, 2010.
- [7] CHEN, J. (1998) Testing for a change point in linear regression models. *Communications in Statistics - Theory and Methods*, **27:10**(1998). Páginas 2481-2493.
- [8] CHEN, J.; GUPTA, A.K. *Parametric statistical change point analysis - With applications to genetics, medicine and finance*. Birkäuser, 2012. Páginas 1-5, 139-154.
- [9] DURBIN, J.; WATSON, G.S. (1950) Testing for serial correlation in least squares regression: I. *Biometrika*, **37**(3/4). Páginas 409-428.
- [10] DURBIN, J.; WATSON, G.S. (1951) Testing for serial correlation in least squares regression: II. *Biometrika*, **38**(1/2). Páginas 159-177.
- [11] ECKLEY, I.; FEARNHEAD, P.; KILLICK, R. (2011). Analysis of changepoint models. Em BARBER, D.; CEMGIL, A.; CHIAPPA, S. (Eds.) *Bayesian Time Series Models*. Cambridge: Cambridge University Press, 2011. Páginas 205-224.
- [12] KUTNER, M.H.; NACHTSHEIN, C.J.; NETER, J.; LI, W.; (2004). *Applied linear statistical models*. McGraw-Hill/Irwin, 2004. Páginas 488-489.
- [13] HOLBERT, D. (1982) A Bayesian analysis of a switching linear model. *Journal of Econometrics*, **19**. Páginas 170-176.
- [14] JOHNSON, R.A.; WICHERN, D.W. *Applied multivariate statistical analysis*. Pearson, 2008.
- [15] MCLEOD, A.I.; ZHANG, Y. (2006) Partial autocorrelation parametrization for subset autoregression. *Journal of Time Series Analysis*, **27**(4). Páginas 599-612.
- [16] MONTGOMERY, D.C.; PECK, E.A.; VINING, G.G. *Introduction to Linear Regression Analysis*. John Wiley & Sons, 2012.

- [17] R CORE TEAM. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. 2018. <https://www.R-project.org/>.
- [18] SCHUMACHER, F.L.; LACHOS, V.; DEY, D. (2017) Censored regression models with autoregressive errors: A likelihood-based perspective. *Canadian Journal of Statistics*, **45**. Páginas 375-392.
- [19] SCHWARZ, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6**(2). Páginas 461-464.
- [20] WICHERN, D.W.; MILLER, R.B.; HSU, D. (2015) Changes of variance in first-order autoregressive time series models-with an application. *Journal of the royal statistical society. Series C (Applied Statistics)*, **25**(3). Páginas 248-256.
- [21] ZEILEIS, A.; HOTHORN, T. (2002) Diagnostic Checking in Regression Relationships. *R News*, **2**(3). Páginas 7-10.

APÊNDICE A – Rotinas em R para o modelo proposto

Os códigos em R a seguir foram elaborados para a obtenção de estimativas para o modelo proposto no Capítulo 4. Também estão presentes as rotinas para a estimação do modelo apresentado no Capítulo 3.

```
# Funções e funções auxiliares para as estimações:
## MRLAR
## MRLN-PM
## MRL-AR-PM
#
## Necessário carregar todas as funções.
#
# transforma pi em phi
estphit = function(pit) {
p = length(pit)
Phi = matrix(0,ncol=p,nrow=p)
if (p>1) {
diag(Phi) = pit
for (j in 2:p) {
for (k in 1:(j-1)) {
Phi[j,k] = Phi[j-1,k] - pit[j]*Phi[j-1,j-k]
}
}
return(Phi[p,])
}
else return(pit)
}
#
# gera matriz de covariância a partir dos phis
MatArp<-function(phi,n) {
p = length(phi)
if (n==1) Rn = 1
else Rn= toeplitz(ARMAacf(ar=phi, ma=0, lag.max = n-1))
rhos = ARMAacf(ar=phi, ma=0, lag.max = p)[(1:p)+1]
return(Rn/(1-sum(rhos*phi)))
}
#
# Dbeta entra na estimação
```

```

Dbeta = function(beta,y,x,q) {
n = dim(y)[1]
D = matrix(0,(q+1),(q+1))
for (ii in 1:(q+1)) {
for (jj in 1:(q+1)) {
D[ii,jj] = sum((y-x%%beta)[ii:(n+1-jj)]*(y-x%%beta)[jj:(n+1-ii)])
}
}
return(D)
}
#
#####estimar pi -- normal
lcc = function(pit,y,x) {
n = dim(y)[1]
phi = matrix(estphit(pit))
q = dim(phi)[1]
lambda = matrix(c(-1,phi))
betahat = solve(t(x)%%solve(MatArp(phi,n))%%x)%%
t(x)%%solve(MatArp(phi,n))%%y
spi = t(lambda)%%Dbeta(betahat,y,x,q)%%lambda
gp = 1
for (i in 1:q) gp = gp*((1-pit[i]^2)^(-i))
l = as.numeric(-n/2*log(spi) -1/2*log(gp))
return(-l)
}
#
### SIC - PM
SIC_H0 <- function(Y,X,betahat){
p <- dim(betahat)[2]
m <- dim(Y)[1]
SIC <- m*(log(t(Y-X%%betahat)%%(Y-X%%betahat))+(log(2*pi)+1))+(p+1-m)*log(m)
return(SIC)
}

SIC_H1 <- function(Y,Y1,Y2,X1,X2,betahat1,betahat2){
p <- dim(betahat1)[2]
m <- dim(Y)[1]
SIC <- m*(log(t(Y1-X1%%betahat1)%%(Y1-X1%%betahat1))+
t(Y2-X2%%betahat2)%%(Y2-X2%%betahat2))+

```

```

        (log(2*pi)+1))+(2*p+1-m)*log(m)
return(SIC)
}
#
### SIC-PM-AR
SIC_H0_caso3 <- function(Y,X,betahat,phihat){
p <- dim(betahat)[1]
q <- dim(phihat)[1]
m <- dim(Y)[1]
lambda <- matrix(c(-1,phihat))
SIC <- m*(log(t(lambda)%*%Dbeta(betahat,Y,X,q)%*%lambda)+(log(2*pi)+1))
      +(p+q+1-m)*log(m)+2*log(det(MatArp(phihat,m)))
return(SIC)
}
#
SIC_H1_caso3 <- function(Y,Y1,Y2,X1,X2,betahat1,betahat2,phihat1,phihat2,k){
p <- dim(betahat1)[1]
q <- dim(phihat1)[1]
m <- dim(Y)[1]
lambda1 <- matrix(c(-1,phihat1))
lambda2 <- matrix(c(-1,phihat2))
SIC <- m*((log(2*pi)+1)+log((t(lambda1)%*%Dbeta(betahat1,Y1,X1,q)%*%lambda1) +
      (t(lambda2)%*%Dbeta(betahat2,Y2,X2,q)%*%lambda2)))) +
      log(det(MatArp(phihat1,k))) + log(det(MatArp(phihat2,(m-k)))) +
      (2*(p+q)+1-m)*log(m)
return(SIC)
}
#
#
### Estimação
#
# MRL-AR
ajusteMRLAR <- function(Y,X,q){
m <- length(Y)
beta1 <- solve(t(X)%*%X)%*%t(X)%*%Y
pi1 <- as.numeric(pacf((Y - X)%*%beta1),lag.max = q,plot=F)$acf)
pi1 <- optim(pi1,lcc,y=Y,x=X,lower = rep(-.999,q), upper = rep(0.999,q),
      method='L-BFGS-B')$par
phi1 <- estphit(pi1)

```

```

lambda <- matrix(c(-1,phi1))
beta1 <- solve(t(X)%*%solve(MatArp(phi1,m))%*%X)%*%
          t(X)%*%solve(MatArp(phi1,m))%*%Y
spi <- t(lambda)%*%Dbeta(beta1,Y,X,q)%*%lambda
sigmae <- as.numeric(spi/m)
theta <- c(beta1,sigmae,phi1)
r <- length(theta)
return(theta)
}
#
# MRLN - PM
ajusteMRLN_PM <- function(Y,X){
p <- dim(X)[2]
m <- dim(Y)[1]
### sob hipótese nula
beta_H0 <- solve(t(X)%*%X)%*%t(X)%*%Y
sigma_H0 <- t(Y-X)%*%beta_H0)%*%(Y-X)%*%beta_H0)/m
theta_H0 <- c(beta_H0,sigma_H0)
###
theta_H1 <- matrix(NA,nrow = m,ncol = 2*p+3)
SICHO <- SIC_H0(Y,X,beta_H0)
SICH1 <- matrix(NA, nrow = m, ncol = 1)
### sob hipótese alternativa
for (k in p:(m-p)) {
Y1 <- Y[1:k,]
X1 <- X[1:k,]
Y2 <- Y[(k+1):m,]
X2 <- X[(k+1):m,]
beta1_H1 <- solve(t(X1)%*%X1)%*%t(X1)%*%Y1
beta2_H1 <- solve(t(X2)%*%X2)%*%t(X2)%*%Y2
sigma_H1 <- (t(Y1-X1)%*%beta1_H1)%*%(Y1-X1)%*%beta1_H1) +
t(Y2-X2)%*%beta2_H1)%*%(Y2-X2)%*%beta2_H1)/m
theta_H1[k,1:dim(beta1_H1)[1]] <- t(beta1_H1)
theta_H1[k,(dim(beta1_H1)[1]+1):(2*dim(beta2_H1)[1])] <- t(beta2_H1)
theta_H1[k,2*p+1] <- sigma_H1
theta_H1[k,2*p+2] <- k
SICH1[k,] <- SIC_H1(Y,Y1,Y2,X1,X2,beta1_H1,beta2_H1)
theta_H1[k,2*p+3] <- SICH1[k,]
}

```

```

k_est <- which.min(SICH1)
if(SICHO > SICH1[k_est,]) {
  return(theta_H1[k_est,])
} else return(theta_H0)
}
#
#
# MRLN - PM - AR(p)
## Mudança nos betas e nos coeficientes AR, com variância =  $\sigma^2 * Mn$ 
## Pelo ajusteMRLAR tem que passar: Y, X, q(ordem do AR).
#
ajusteMRLN_PM_AR_caso3 <- function(Y,X,q){
  p <- dim(X)[2]
  m <- dim(Y)[1]
  ### sob hipótese nula
  theta_H0 <- ajusteMRLAR(Y,X,q)
  beta_H0 <- matrix(theta_H0[1:p])
  phi_H0 <- matrix(theta_H0[(p+1):(p+q)])
  ###
  SICHO <- SIC_H0_caso3(Y,X,beta_H0,phi_H0)
  #theta_H1 <- matrix(NA,nrow = m,ncol = 2*(p+q)+2)
  theta_H1 <- matrix(NA,nrow = m,ncol = 2*(p+q)+3)
  SICH1 <- matrix(NA, nrow = m, ncol = 1)
  ### sob hipótese alternativa
  for (k in (p+q):(m-p-q)) {
    Y1 <- matrix(Y[1:k,])
    X1 <- X[1:k,]
    Y2 <- matrix(Y[(k+1):m,])
    X2 <- X[(k+1):m,]
    # chute inicial para os betas
    beta1_H10 <- solve(t(X1)%*%X1)%*%t(X1)%*%Y1
    beta2_H10 <- solve(t(X2)%*%X2)%*%t(X2)%*%Y2
    #
    # chute inicial para phis
    # phi da parte 1
    pi1 <- as.numeric(pacf(((Y1 - X1%*%beta1_H10)),lag.max = q,plot=F)$acf)
    pi1 <- optim(pi1,lcc,y=Y1,x=X1,lower = rep(-.999,q),
               upper = rep(0.999,q),method='L-BFGS-B')$par
    # phi da parte 2

```



```

pi2 <- as.numeric(pacf(((Y2 - X2**beta2_H10))),lag.max = q,plot=F)$acf)
pi2 <- optim(pi2,lcc,y=Y2,x=X2,lower = rep(-.999,q),
            upper = rep(0.999,q),method='L-BFGS-B')$par
phi1 <- matrix(estphit(pi1))
lambda1 <- matrix(c(-1,phi1))
phi2 <- matrix(estphit(pi2))
lambda2 <- matrix(c(-1,phi2))
# estimacão betas
beta1_H1 <- solve(t(X1)**solve(MatArp(phi1,k))**X1)**
              t(X1)**solve(MatArp(phi1,k))**Y1
beta2_H1 <- solve(t(X2)**solve(MatArp(phi2,(m-k)))**X2)**
              t(X2)**solve(MatArp(phi2,(m-k)))**Y2
# estimacao variancia
sigmae2 <- as.numeric(((t(lambda1)**Dbeta(beta1_H1,Y1,X1,q)**lambda1) +
(t(lambda2)**Dbeta(beta2_H1,Y2,X2,q)**lambda2))/m)
theta_H1[k,1:dim(beta1_H1)[1]] <- t(beta1_H1)
theta_H1[k,(dim(beta1_H1)[1]+1):(2*dim(beta2_H1)[1])] <- t(beta2_H1)
theta_H1[k,(2*dim(beta2_H1)[1]+1):(2*dim(beta2_H1)[1]+q)] <- t(phi1)
theta_H1[k,(2*dim(beta2_H1)[1]+q+1):(2*(dim(beta2_H1)[1]+q))] <- t(phi2)
theta_H1[k,2*(p+q)+1] <- sigmae2
theta_H1[k,2*(p+q)+2] <- k
SICH1[k,] <- SIC_H1_caso3(Y,Y1,Y2,X1,X2,beta1_H1,beta2_H1,phi1,phi2,k)
theta_H1[k,2*(p+q)+3] <- SICH1[k,]
}
k_est <- which.min(SICH1)
if(SICH0 > SICH1[k_est,]) {
#return(list(theta_H1[k_est,], SICH1))
# muito MUITO pesado
return(theta_H1[k_est,])
} else return(theta_H0)
}
#

```

APÊNDICE B – Rotinas para simulação

```

## replic: quantas simulações serão feitas

pit1 <- c(.5,.7)
phit1 <- estphit(pit1)

pit2 <- c(.8,.9)
phit2 <- estphit(pit2)

q <- length(phit1)
desv_pad <- 0.68
desv_pad^2

# diferença em todos os parâmetros da regressão
#beta1 <- matrix(c(2,1.5),2,1)
#beta2 <- matrix(c(4,0.5),2,1)

# mesmo intercepto
#beta1 <- matrix(c(2,1.5),2,1)
#beta2 <- matrix(c(2,0.5),2,1)

# Paralelas
#beta1 <- matrix(c(6,1.5),2,1)
#beta2 <- matrix(c(2,1.5),2,1)

# Concorrentes
#beta1 <- matrix(c(20,-1.5),2,1)
#beta2 <- matrix(c(2,1.5),2,1)
p <- as.numeric(dim(beta1)[1])

n <- 200
replic <- 100

thetaAR <- matrix(0,replic,2*(p+q)+3)

###change-points:
# k=n change-point is in the median
# k=n/2 (3n/2) change-point is in the 1st (3rd) quantile

```

```

inic <- Sys.time()
for (i in 1:replic){
eta1 <- rnorm((n),0,desv_pad)
erro1 <- c()
erro1[1] <- eta1[1]
erro1[2] <- eta1[2]
for(t in 3:(n)) erro1[t] = phit1[1]*erro1[t-1] + phit1[2]*erro1[t-2] + eta1[t]
eta2 <- rnorm((n),0,desv_pad)
erro2 <- c()
erro2[1] <- eta2[1]
erro2[2] <- eta2[2]
for(t in 3:(n)) erro2[t] = phit2[1]*erro2[t-1] + phit2[2]*erro2[t-2] + eta2[t]
X <- matrix(1,2*n,2)
X[,2] <- runif(2*n,3,20)
Y <- matrix(0,2*n,1)
Y[1:(n),] <- X[1:(n),]%*%beta1 + matrix(erro1)
Y[((n)+1):(2*n),] <- X[((n)+1):(2*n),]%*%beta2 + matrix(erro2)
thetaAR[i,] <- ajusteMRLN_PM_AR_caso3(Y,X,q)
print(i)
}
fin <- Sys.time()

```