



ANÁLISE DE MÚLTIPLOS PONTOS DE MUDANÇA EM MODELOS
NORMAL MULTIVARIADOS

Leonardo Brandão Freitas do Nascimento

Dissertação de Mestrado apresentada ao
Programa de Pós Graduação em Estatística da
Universidade Federal de Minas Gerais, como
parte dos requisitos necessários à obtenção do
título de Mestre em Estatística.

Orientadora: Prof^a Rosângela Helena Loschi,
D.Sc.

Belo Horizonte
Janeiro de 2017

ANÁLISE DE MÚLTIPLOS PONTOS DE MUDANÇA EM MODELOS
NORMAL MULTIVARIADOS

Leonardo Brandão Freitas do Nascimento

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE
PÓS GRADUAÇÃO EM ESTATÍSTICA DA UNIVERSIDADE FEDERAL DE
MINAS GERAIS, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
OBTENÇÃO DO GRAU DE MESTRE EM ESTATÍSTICA.

Examinada por:

Prof^a. Rosangela Helena Loschi, D.Sc.

Prof. Clécio da Silva Ferreira, D.Sc.

Cristiano de Carvalho Santos, D.Sc.

Prof. Flávio Bambilra Gonçalves, D.Sc.

BELO HORIZONTE, MG – BRASIL
JANEIRO DE 2017

Freitas do Nascimento, Leonardo Brandão

Análise de Múltiplos Pontos de Mudança em Modelos Normal Multivariados/Leonardo Brandão Freitas do Nascimento. – Belo Horizonte: UFMG/ICEx, 2017.

VIII, 105 p.: il.; 29, 7cm.

Orientadora: Prof^a Rosangela Helena Loschi, D.Sc.

Dissertação (mestrado) – UFMG/ICEx, Área de Concentração: Estatística, 2017.

Referências Bibliográficas: p. 102 – 105.

1. Modelo Partição Produto. 2. Múltiplas Mudanças.
3. Normal Multivariada. 4. Dados Econômicos. I.
D.Sc., Prof^a Rosangela Helena Loschi,. II. Universidade
Federal de Minas Gerais, UFMG, Área de Concentração:
Estatística. III. Título.

Agradecimentos

A Deus que através dos meus estudos me permite auxiliá-lo na criação.

Aos membros da banca, professores Clécio da Silva Ferreira, Flávio Bambirra Gonçalves e Rosângela Helena Loschi e ao pesquisador Cristiano de Carvalho Santos que se disponibilizaram de fazer parte da banca e pelas sugestões para a melhoria do trabalho. Em especial, agradeço ao Cristiano de Carvalho Santos pela amizade e pelos conselhos antes mesmo da defesa. Também agradeço de forma especial a professora Rosângela Helena Loschi que me orientou pacientemente e sempre dedicada a resolver os problemas oriundos do trabalho, além da amizade e compreensão do fato de eu ter que defender antes do período esperado.

Aos meus familiares, amigos e amigas que durante esse tempo de mestrado contribuíram para o desenvolvimento do trabalho ou simplesmente pelo fato da companhia de cada um.

Agradeço à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo financiamento desta pesquisa através do Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais.

Resumo da Dissertação apresentada ao Programa de Pós Graduação em Estatística da Univerisdade Federal de Minas Gerais, como parte dos requisitos necessários para a obtenção do grau de Mestre em Estatística. (M.Sc.)

ANÁLISE DE MÚLTIPLOS PONTOS DE MUDANÇA EM MODELOS NORMAL MULTIVARIADOS

Leonardo Brandão Freitas do Nascimento

Janeiro/2017

Orientadora: Prof^a Rosangela Helena Loschi, D.Sc.

Nesse trabalho, propõe-se uma extensão do Modelo Partição Produto para a identificação de múltiplos pontos de mudança, ao longo do tempo, no vetor de médias e na matriz de variância e covariância de uma sequência de dados com distribuição normal multivariada. Para isso, distribuições *a priori* conjugadas foram utilizadas para estimar o vetor de médias e a matriz de variância e covariância ao longo do tempo. Também propõe-se realizar uma comparação de cada parâmetro sequencialmente. Para este fim, constrói-se intervalos de mais alta densidade (intervalos HPD) *a posteriori* para a diferença de parâmetros em sucessivos instantes de tempo. Para avaliar o modelo proposto, foram considerados alguns cenários simulados e realizada uma aplicação em dados financeiros, mais especificamente uma análise do impacto da saída do Reino Unido da União Europeia.

Sumário

Lista de Figuras	vii
Lista de Tabelas	viii
Introdução	1
1 Modelo Partição Produto para dados sequencialmente observados: uma revisão	4
1.1 Modelo Partição Produto	5
1.2 Construção do Modelo Partição Produto	5
1.3 Métodos computacionais	13
2 Identificando múltiplas mudanças na média e na covariância de da- dos Normal Multivariados	16
2.1 Inferência Bayesiana no modelo normal multivariado	17
2.2 Construção do MPP	19
2.3 Falta de identificabilidade na identificação da mudança	21
3 Análise de sensibilidade do Modelo Partição Produto para Normal Multivariada	23
3.1 Cenário 1: mudança na média	24
3.2 Cenário 2: mudança na variância	39
3.3 Cenário 3: mudanças em todos os parâmetros	52
4 Avaliando o efeito do <i>Brexit</i> na economia	59
4.1 Banco de dados	60
4.2 Especificações <i>a priori</i>	62
4.3 Análise	63
5 Conclusão	76
A Distribuições	79

B	Demonstração da Proposição 3.1.1	82
C	Algumas propriedades matriciais	89
D	Gráficos	90
	Referências Bibliográficas	102

Lista de Figuras

3.1	Valores das séries para o cenário 1.	25
3.2	Probabilidade <i>a posteriori</i> do número de blocos, cenário 1	28
3.3	Probabilidade <i>a posteriori</i> de cada instante ser um ponto de mudança, cenário 2	30
3.4	Estimativas do parâmetro μ_k da série 1, cenário 1	32
3.5	Estimativas do parâmetro μ_k da série 2, cenário 1	34
3.6	Intervalo HPD de 95% para $\mu_k - \mu_{k-1}$, série 2 e cenário 1	35
3.7	Estimativas do parâmetro σ_k^2 da série 1, cenário 1	36
3.8	Estimativas do parâmetro σ_k^2 da série 2, cenário 1	37
3.9	Estimativas do parâmetro $\sigma_{(12)k}^2$, cenário 1	38
3.10	Valores das séries para o cenário 1.	39
3.11	Probabilidade <i>a posteriori</i> para número de blocos, cenário 2	42
3.12	Probabilidade <i>a posteriori</i> de cada instante ser um ponto de mudança, cenário 2	43
3.13	Estimativas do parâmetro μ_k da série 1, cenário 2	45
3.14	Estimativas do parâmetro μ_k da série 2, cenário 2	47
3.15	Estimativas do parâmetro σ_k^2 da série 1, cenário 2	48
3.16	Estimativas do parâmetro σ_k^2 da série 2, cenário 2	49
3.17	Intervalo HPD de 95% para $\sigma_k^2 - \sigma_{k-1}^2$, séries 2 e cenário 2	50
3.18	Estimativas do parâmetro $\sigma_{(12)k}^2$, cenário 2	51
3.19	Valores das séries para o cenário 3.	52
3.20	Probabilidade <i>a posteriori</i> do número de blocos, cenário 3	53
3.21	Distribuição <i>a posteriori</i> da probabilidade de mudança, cenário 3	54
3.22	Probabilidade <i>a posteriori</i> de cada instante ser um ponto de mudança, cenário 3	55
3.23	Estimativas do parâmetro μ_k da série 1 e 2, cenário 3	56
3.24	Intervalo HPD de 95% para $\mu_k - \mu_{k-1}$, séries 1 e 2, cenário 3	57
3.25	Estimativas dos parâmetros σ_k^2 e $\sigma_k(12)^2$, cenário 3	58
4.1	Retornos dos índices financeiros da Alemanha, EUA, França, Reino Unido e Suíça.	61

4.2	Retornos dos índices financeiros da Espanha, Grécia, Itália e Portugal.	62
4.3	Probabilidade <i>a posteriori</i> dos blocos, aplicação	64
4.4	Distribuição <i>a posteriori</i> da probabilidade de mudança, aplicação . . .	65
4.5	Probabilidade <i>a posteriori</i> de cada instante ser ponto de mudança, aplicação	65
4.6	Estimação para μ_k , estudo de caso	67
4.7	Estimação para μ_k , estudo de caso	68
4.8	Estimação para μ_k , estudo de caso	69
4.9	Estimação para σ_k^2 , estudo de caso	70
4.10	Estimação para σ_k^2 , estudo de caso	71
4.11	Estimação para σ_k^2 , estudo de caso	72
4.12	Estimação para covariância, estudo de caso	73
4.13	Estimação para covariância, estudo de caso	74
4.14	Estimação para covariância, estudo de caso	75
D.1	Cadeia de μ_{50} e μ_{51} , cenário 1 e série 2	90
D.2	Distribuição de $\mu_{51} - \mu_{50}$, cenário 1 e série 2	91
D.3	Cadeias da probabilidade de mudança, cenário 1	91
D.4	Cadeia de σ_{51}^2 e $\sigma_{51}^2 - \sigma_{50}^2$, cenário 1 e série 2	91
D.5	Distribuição <i>a posteriori</i> da probabilidade de mudança, cenário 1 . . .	92
D.6	Cadeia de μ_{51} e μ_{51} , cenário 2 e série 1	93
D.7	Cadeia de σ_{52}^2 e σ_{52}^2 , cenário 2 e série 2	93
D.8	Distribuição $\sigma_{53}^2 - \sigma_{52}^2$, cenário 2 e série 2	94
D.9	Distribuição <i>a posteriori</i> da probabilidade de mudança, cenário 2 . . .	95
D.10	Cadeias da média, variância e covariância, estudo de caso	96
D.11	Cadeias da probabilidade de mudança, estudo de caso	97
D.12	Estimativas para covariância, estudo de caso	97
D.13	Estimativas para covariância, estudo de caso	98
D.14	Estimativas para covariância, estudo de caso	99
D.15	Estimativas para covariância, estudo de caso	100
D.16	Estimativas para covariância, estudo de caso	101
D.17	Estimativas para covariância, estudo de caso	101

Lista de Tabelas

1.1	Relação entre a partição ρ e o número de blocos B	5
1.2	Relação entre ρ e U	13
3.1	Estatísticas <i>a priori</i> , cenário 1	26
3.2	Probabilidade <i>a posteriori</i> das partições, cenário 1	27
3.3	Estimativas <i>a posteriori</i> para a probabilidade de ponto de mudança, cenário 1	28
3.4	Estimativas <i>a priori</i> , cenário 2	40
3.5	Probabilidade <i>a posteriori</i> das partições, cenário 2	41
3.6	Estimativas para a probabilidade de ponto de mudança, cenário 2 . .	42
3.7	Média e variância <i>a priori</i> para $l = 1, 2$	52
3.8	Probabilidade <i>a posteriori</i> das partições, cenário 3	53
4.1	Média e variância dos parâmetros, para $l, t = 1, \dots, 9$	62
4.2	Probabilidade <i>a posteriori</i> das partições, aplicação	63
4.3	Estimativas <i>a posteriori</i> para a probabilidade de mudança, aplicação.	64

Introdução

A identificação de pontos de mudança é um problema estatístico relevante, pois possibilita conjecturar sobre os fenômenos que os causaram e auxilia na avaliação de ações a serem tomadas. Nesse sentido, um ponto de mudança pode ser entendido como uma quebra estrutural em uma série temporal ou, simplesmente, um valor atípico que ocorre em um banco de dados.

O problema de identificação de pontos de mudança, em dados observados sequencialmente, esta presente em diversas áreas como, por exemplo, em estudo hidrológicos (Perreault *et al.* [36]), climatológicos (Ruggieri [40]), econômicos (Arellano-Valle *et al.* [2]) e sociais (Loschi *et al.* [27]). Problemas de pontos de mudança também ocorrem em segmentação de textos como pode ser visto em Kehagias *et al.* [17].

Nesse sentido, se faz necessária a utilização de modelos estatísticos em conjunto com ferramentas computacionais capazes de reconhecer - los adequadamente e, conseqüentemente, ajudar nas tomadas de decisões. Assim, diversas abordagens e metodologias foram desenvolvidas para a captação dos pontos de mudança em conjunto de dados observados sequencialmente, entre as quais destacam-se a metodologia de máxima verossimilhança (Hinkley [13], Bhattacharya [5], Liu [21]) e procedimentos *Bayesianos* baseados em métodos paramétricos e não paramétricos (Martínez *et al.* [31], Allen *et al.* [1] e Hartigan [12]).

Do ponto de vista *Bayesiano* paramétrico e considerando apenas um ponto de mudança, destaca-se o trabalho de Smith [42] que considerou o problema de mudança na média de uma seqüência de variáveis aleatórias com distribuições normal e binomial. Lee & Heghinian [20] também estudaram uma mudança na média em uma seqüência de variáveis aleatórias independentes com distribuição normal, variância desconhecida e comum. Trabalhos aplicados na área econômica podem ser vistos em Booth & Smith [6], Diaz [9], Holbert [14] e Salazar [41]. Ainda sobre a perspectiva *Bayesiana* paramétrica, cita-se os trabalhos de Perreault *et al.* [37], que considerou um ponto de mudança no vetor de médias e Son & Kim [43], que considerou um ponto de mudança tanto no vetor de médias quanto na matriz de covariância, em uma seqüência de variáveis aleatórias independentes com distribuição normal multivariada.

Modelos que identificam apenas um ponto de mudança são muito restritivos,

pois em muitas situações a série pode apresentar várias mudanças. Hartigan [12] propôs o Modelo Partição Produto (MPP) em que considera a suposição da existência de múltiplos pontos de mudança. No MPP tanto o número de pontos de mudança quanto os instantes em que ocorreram são quantidades desconhecidas as quais devem ser estimadas, logo, tornando o modelo mais flexível. Barry & Hartigan [3] desenvolveram uma versão do MPP para a identificação de pontos de mudança em dados observados sequencialmente, no qual apenas blocos contíguos são considerados. Posteriormente, Barry & Hartigan [4] aplicaram o MPP para inferir sobre múltiplos pontos de mudança na média da distribuição normal com variância constante e desconhecida. Uma grande contribuição de Barry & Hartigan [4] foi o desenvolvimento de um algoritmo para gerar da distribuição *a posteriori* de uma partição com blocos contíguos, em que usam o amostrador de Gibbs. Aplicações e extensões do MPP podem ser vistos em Loschi *et al.* [28], Monteiro *et al.* [32], Loschi *et al.* [24], Quintana & Iglesias [38], Müller *et al.* [34] e entre muitos outros

A detecção de múltiplas mudanças em séries temporais multivariadas é de grande interesse e um problema ainda pouco explorado. Se estas séries são correlacionadas, a ocorrência de alguma mudança no comportamento de uma delas pode gerar mudança em alguma outra. Nesse cenário, o presente trabalho considera que uma sequência de vetores aleatórios, $\mathbf{Y}_1, \dots, \mathbf{Y}_n$, de dimensão $p \times 1$ são independentes e distribuídos segundo uma distribuição normal multivariada com vetores de médias $\boldsymbol{\mu}_k$, $p \times 1$, e matriz de covariância $\boldsymbol{\Sigma}_k$, $p \times p$, para $k = 1, \dots, n$. Assim sendo, temos como objetivos utilizar o MPP para identificar múltiplos pontos de mudanças no vetor de médias e na matriz de variância-covariância. Além de estimar $\boldsymbol{\mu}_k$ e $\boldsymbol{\Sigma}_k$ para todo k , também estima-se o número de pontos de mudança e as posições onde as mudanças ocorreram. Distribuições *a priori* conjugadas são utilizadas para $\boldsymbol{\mu}_k$ e $\boldsymbol{\Sigma}$.

O modelo apresentado nesse trabalho é uma extensão ao proposto em Cheon & Kim [8] e em Moura [33]. Moura [33], por exemplo, utilizou o modelo partição produto para a identificação de mudanças na matriz de variância e covariância para uma sequência de dados com distribuição normal multivariada centrado no vetor nulo. Cheon & Kim [8] aplicaram o algoritmo Monte Carlo com aproximação estocástica para a detecção de múltiplos pontos de mudanças no vetor de médias e na matriz de variância e covariância de dados com distribuição normal multivariada. No entanto, Cheon & Kim [8] consideraram uma distribuição *a priori* vaga para o vetor de médias, ou seja, $\pi(\boldsymbol{\mu}_k) = 1$ e estimam apenas a partição dos dados, o número de blocos e a probabilidade de cada instante ser um ponto de mudança, não fornecendo estimativas para os parâmetros no instante k e, conseqüentemente, não havendo a identificação em qual parâmetro e em qual variável houve a mudança. Além desses trabalhos, destacam-se também o de Lavielle & Teyssiere [19], por exemplo, que propõem um método adaptativo para a identificação de mudança

na estrutura da matriz de covariância. Mabaera *et al.* [29] aplicaram o método de máxima verossimilhança para a detecção de múltiplos pontos de mudança em séries multivariadas. Recentemente, James & Matteson [15] disponibilizaram o pacote *ecp* no *software* R Core Team [39], no qual utiliza da abordagem não paramétrica para a análise de múltiplos pontos de mudança em dados multivariados.

Em estudos de pontos de mudança usando o MPP, comumente ocorrem problemas de falta de identificabilidade. Por exemplo, no problema de identificação de múltiplas mudanças na média e variância de uma sequência de variáveis normalmente distribuídas (Loschi & Cruz [22]), a mudança real pode ocorrer apenas na média e não na variância. No entanto, estudos simulados mostram que as estimativas fornecidas pelo modelo indicarão também uma mudança na variância e em instantes próximos a mudança detectada na média. Problemas similares a este também ocorrerão no contexto multivariado. Neste caso, podemos, por exemplo, inferir que há mudança em todas as série quando apenas uma destas experimentou a mudança. Para analisar os resultados e minimizar este tipo de problema da falta de identificação, propõe-se realizar uma comparação de cada parâmetro sequencialmente. Para este fim, constrói-se intervalos de mais alta densidade (intervalos HPD) *a posteriori* para a diferença de parâmetros em sucessivos instantes de tempo. O intervalo HPD se relaciona com a evidência *a posteriori* fornecida pelo teste de significância bayesiano completo (FBST) introduzido por Bragança Pereira & Stern [7]. Assim, pode-se considera-lo para testar se a diferença de cada parâmetro em instantes sucessivos de tempo é significamente diferente de zero, concluindo-se diferença significativa quando o valor zero não pertence ao intervalo.

Para avaliar o desempenho do modelo, considerou-se alguns bancos de dados simulados. Além disso, também realizou-se uma aplicação em que consiste em avaliar o efeito do *Brexit* na economia de oito países. Esse termo é a abreviação das palavras em inglês *Britain* (Grã-Bretanha) e *exit* (saída) e designa a saída do Reino Unido da União Europeia. Para esse estudos foram considerados nove índices financeiros: DAX (*Deutscher Aktienindex*) da Alemanha, IBEX 35 (*Iberia Index*) da Espanha, CAC 40 (*Cotation Assistée en Continu*) da França, ATG (*Athens General*) da Grécia, FTSE MIB (*Financial Times Stock Exchange Milano Indice di Borsa*) da Itália, PSI-20 (*Portuguese Stock Index*) de Portugal, FTSE 100 Reino Unido (*Financial Times Stock Exchange*), SMI (Swiss Market Index) da Suíça e S&P 500 (Standard & Poor's 500) dos Estados Unidos. Tais índices são responsáveis por medir o desempenho econômico dos países em estudo

Este trabalho está organizado como segue. No Capítulo 1, é apresentado uma revisão do MPP desenvolvido em Barry & Hartigan [3] e Barry & Hartigan [4], as extensões proposta em Loschi & Cruz [25] e o método computacional para obter-se as distribuições *a posteriori* das partições. No Capítulo 2, é definido o modelo partição

produto para a distribuição Normal Multivariada e é apresentado um algoritmo para realizar os testes sequenciais. No Capítulo 3, é realizada uma análise de sensibilidade do modelo proposto através de três cenários sob diferentes óticas. No Capítulo 4, é analisado o efeito do *Brexit* na economia de oito países.

Capítulo 1

Modelo Partição Produto para dados sequencialmente observados: uma revisão

O Modelo Partição Produto foi definido, na sua forma mais geral, em Hartigan [12] e considera a suposição da existência de múltiplos pontos de mudança em um determinado banco de dados. Dessa forma, o MPP é mais flexível quando comparado aos modelos que consideram a presença de apenas um ponto de mudança. No MPP tanto o número de pontos de mudanças quanto os instantes em que estes ocorreram são quantidades a serem estimadas. Barry & Hartigan [3] particularizaram o MPP para a identificação de pontos de mudança em dados observados sequencialmente por assumirem que apenas blocos contíguos são considerados. Barry & Hartigan [3] forneceram expressões analíticas para as distribuições *a posteriori* dos parâmetros que indexam a função de distribuição dos dados e para suas esperanças. Posteriormente, Barry & Hartigan [4] aplicaram o MPP para inferir sobre múltiplos pontos de mudança na média de uma sequência de dados normalmente distribuídos e assumindo variância comum e desconhecida em cada instante do tempo. Além disso, propuseram um algoritmo baseado no amostrador de Gibbs para gerar da distribuição *a posteriori* da partição aleatória quando apenas blocos contíguos são possíveis. Mais tarde, Loschi *et al.* [23] implementou o amostrador de Gibbs para gerar da distribuição *a posteriori* do número de pontos de mudanças e para os instantes nos quais estes ocorreram. Loschi *et al.* [23] também considerou a coesão definida em Yao [44], no qual depende de uma probabilidade p de haver mudança em algum instante, assumindo uma distribuição *a priori* degenerada para a p . Loschi *et al.* [26] fez uma extensão de Loschi *et al.* [23] assumindo distribuições *a priori* não degeneradas para a p . Posteriormente, Ferreira *et al.* [10] apresenta uma versão do MPP no qual considera os blocos correlacionados.

Neste capítulo será feito uma revisão do MPP. Na Seção 1.1 são apresentada as definições e conceitos básicos inerentes ao MPP. Na Seção 1.2 é definido o MPP no caso paramétrico e são apresentadas as distribuições *a priori* e *a posteriori* da partição aleatória, do número de blocos e a probabilidade *a posteriori* de cada instante ser um ponto de mudança. Por fim, na Seção 1.3, é descrito o método computacional que é utilizado para a obtenção das estimativas de interesse do MPP.

1.1 Modelo Partição Produto

Sejam y_1, \dots, y_n uma série temporal observada e I o conjunto formado pelos índices $\{1, \dots, n\}$ que indexam tais observações. Denote por ρ a partição aleatória do conjunto $I \cup \{0\}$ e B o número de blocos da partição ρ . Denote por $[ij]$ o subconjunto de I formado pelos indices $\{i + 1, \dots, j\}$, para $i < j$ e $i, j \in I \cup \{0\}$, e por $y_{[ij]}$ o bloco formado pelas observações y_{i+1}, \dots, y_j .

Assumindo que somente blocos contíguos de observações são possíveis, cada valor da partição aleatória ρ é da forma $\rho = \{i_0, i_1, \dots, i_b\}$, em que $0 = i_0 < i_1 < \dots < i_b = n$, para $b \in I$, a qual divide conjunto de observações em $B = b$ blocos vizinhos da seguinte forma:

$$[y_1, \dots, y_{i_1}] [y_{i_1+1}, \dots, y_{i_2}] \dots [y_{i_{b-1}+1}, \dots, y_{i_b}].$$

A ocorrência da partição $\rho = \{i_0, i_1, \dots, i_b\}$ implica na ocorrência de $b - 1$ mudanças no comportamento da série observada nos instantes $i_1 + 1, \dots, i_{b-1} + 1$. Para melhor compreensão, ver Tabela (1.1).

Exemplo 1.1.1 *Seja y_1, y_2, y_3 uma série temporal observada. Então, tem-se que $I \cup 0 = \{0, 1, 2, 3\}$ e os dados podem ser divididos de 2^{n-1} formas diferentes como mostra a Tabela 1.1.*

Tabela 1.1: Relação entre a partição ρ e o número de blocos B .

Bloco de observações	Valores de ρ	Valores de B
$[y_1, y_2, y_3]$	$\{i_0 = 0, i_1 = 3\}$	1
$[y_1], [y_2, y_3]$	$\{i_0 = 0, i_1 = 1, i_2 = 3\}$	2
$[y_1, y_2], [y_3]$	$\{i_0 = 0, i_1 = 2, i_2 = 3\}$	2
$[y_1], [y_2], [y_3]$	$\{i_0 = 0, i_1 = 1, i_2 = 2, i_3 = 3\}$	3

Definidos esses elementos básicos, na próxima seção será apresentado o MPP em sua versão paramétrica.

1.2 Construção do Modelo Partição Produto

Seja $\theta_1, \dots, \theta_n$ uma sequência de parâmetros sobre o espaço paramétrico Θ e Y_1, \dots, Y_n uma sequência de variáveis aleatórias. Considere que, dados $\theta_1, \dots, \theta_n$, as variáveis Y_1, \dots, Y_n são independentes e com comportamento dado por $f_1(y_1|\theta_1), \dots, f_n(y_n|\theta_n)$, respectivamente. No MPP assume-se que:

- 1 Dado $\rho = \{i_0, i_1, \dots, i_b\}$, a sequência de parâmetros $\theta_1, \dots, \theta_n$ é particionada em b blocos contíguos, no qual $\theta_{[ij]}$ denota o parâmetro comum que indexa a distribuição das variáveis aleatórias do bloco $Y_{[ij]}$, ou seja

$$\theta_k = \theta_{[i_{r-1}i_r]} \quad \forall \quad i_{r-1} < k \leq i_r, \quad k = 1, \dots, n \quad \text{e} \quad r = 1, \dots, b.$$

Desta forma, o vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ pode ser reescrito da seguinte forma:

$$\boldsymbol{\theta} = \sum_{r=1}^b (\theta_{[i_{r-1}i_r]} \mathbf{1}\{i_{r-1} < 1 \leq i_r\}, \dots, \theta_{[i_{r-1}i_r]} \mathbf{1}\{i_{r-1} < n \leq i_r\}),$$

onde $\mathbf{1}\{A\}$ denota a função indicadora do evento A

- 2 Dado $\rho = \{i_0, i_1, \dots, i_b\}$, tem-se que os parâmetros comuns $\theta_{[i_0i_1]}, \dots, \theta_{[i_{b-1}i_b]}$ são independentes e $\theta_{[ij]}$ possui distribuição *a priori* $\pi_{[ij]}(\theta)$ com $\theta \in \Theta_{[ij]}$, onde $\Theta_{[ij]}$ é o espaço paramétrico de $\theta_{[ij]}$.

O MPP estabelece a distribuição conjunta de ρ , das observações e dos parâmetros como mostra a definição abaixo.

Definição 1.2.1 *A quantidade aleatória $((Y_1, \theta_1), \dots, (Y_n, \theta_n); \rho)$ possui o seu comportamento descrito pelo MPP paramétrico, denotado por $((Y_1, \theta_1), \dots, (Y_n, \theta_n); \rho) \sim MPP$, se:*

- (1) *A distribuição a priori para ρ é uma distribuição produto, ou seja:*

$$P(\rho = \{i_0, i_1, \dots, i_b\}) = \frac{\prod_{r=1}^b c_{[i_{r-1}i_r]}}{\sum_C \prod_{r=1}^b c_{[i_{r-1}i_r]}},$$

onde $c_{[ij]}$ é a coesão do bloco $[ij]$ (ver detalhes a seguir) e C é o conjunto de todas as possíveis partições de I em b blocos contíguos $\forall b \in I$.

- (2) *Condicional em $\rho = \{i_0, i_1, \dots, i_b\}$, tem-se que as observações em diferentes blocos são independentes e as observações em cada bloco $y_{[ij]}$ possuem distribuição*

indexada pelo parâmetro $\theta_{[ij]}$, ou seja:

$$f(y_1, \dots, y_n; \theta_1, \dots, \theta_n | \rho = \{i_0, i_1, \dots, i_b\}) = \prod_{r=1}^b \prod_{k=i_{r-1}+1}^{i_r} f(y_k | \theta_{[ij]}) \pi_{[i_{r-1}i_r]}(\theta_{[ij]}).$$

(3) Condicional em ρ , a distribuição a priori para $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ é

$$\begin{aligned} \pi(\boldsymbol{\theta}) &= \pi(\theta_{[i_0i_1]}) \cdots \pi(\theta_{[i_{b-1}i_b]}) \\ &= \prod_{r=1}^b \pi(\theta_{[i_{r-1}i_r]}). \end{aligned}$$

Uma partição tende a agregar em um mesmo bloco observações com comportamentos semelhantes. Diante disso, para descrever a incerteza inicial sobre ρ , Hartigan [12] define $c_{[ij]}$ a coesão associada ao bloco de observações $[ij]$. Segundo Hartigan [12], $c_{[ij]}$ é um valor numérico e não negativo que representa o grau de similaridade existente entre observações do bloco $[ij]$, podendo ser interpretada, quando os dados são sequencialmente observados, como uma probabilidade de transição da cadeia de Markov $Z_a, a \in \{0, \dots, n\}$, onde Z_a assume valores em $\{i_0, i_1, \dots, i_b\}$, isto é, corresponde ao instante em que ocorre a a -ésima mudança na estrutura da distribuição. Desse modo, tem-se que a coesão do bloco $c_{[ij]}$ é a probabilidade do instante da a -ésima mudança ser o j dado que a $(a-1)$ -ésima mudança ocorreu no instante i ([3]).

Como consequência da Definição 1.2.1, tem-se que a distribuição conjunta de y_1, \dots, y_n dado ρ , é dada por

$$f(y_1, \dots, y_n | \rho = \{i_0, i_1, \dots, i_b\}) = \prod_{r=1}^b f_{[i_{r-1}i_r]}(y_{[i_{r-1}i_r]}),$$

onde

$$f_{[i_{r-1}i_r]}(y_{[i_{r-1}i_r]}) = \int \prod_{k=i_{r-1}+1}^{i_r} f(y_k | \theta) \pi_{[i_{r-1}i_r]}(\theta) d\theta$$

é denominada fator de dados e corresponde a distribuição preditiva *a priori* associada ao bloco $[ij]$.

A partir da Definição 1.2.1, pode ser mostrado que a distribuição *a priori* do número de blocos na partição ρ é dado por:

$$P(B = b) = \sum_{C'} \frac{\prod_{r=1}^b c_{i_{r-1}i_r}}{\sum_C \prod_{r=1}^b c_{i_{r-1}i_r}},$$

onde C' é um subconjunto de todas as partições em I com exatamente b blocos.

Barry & Hartigan [3] também mostraram que as distribuições *a posteriori* para ρ e B são dadas, respectivamente, por:

$$P(\rho|y_1, \dots, y_n) \propto \prod_{r=1}^b c_{i_{r-1}i_r}^*$$

$$P(B = b|y_1, \dots, y_n) \propto \sum_{C'} \prod_{r=1}^b c_{i_{r-1}i_r}^* \quad \forall b \in I,$$

onde $c_{[ij]}^* = c_{[ij]}f_{[ij]}(y_{[ij]})$, $\forall i, j \in \{0, \dots, n\}$ e $i < j$, denota a coesão *a posteriori* do bloco $[ij]$.

Em geral, a distribuição *a posteriori* de ρ fornece uma boa informação sobre a probabilidade de cada instante ser um ponto de mudança. Loschi & Cruz [25] propôs um método para calcular a probabilidade *a posteriori* de cada instante ser um ponto de mudança.

Seja C_k um subgrupo de C que contém todas as partições que incluem o instante k como ponto de mudança, ou seja, cada partição em C_k é da forma $\{i_0, \dots, i_{l-1}, i_l = k-1, i_{l+1}, \dots, i_b\}$ para algum $l \in I$. Seja o evento A_k que denota o k -ésimo instante é um ponto de mudança, para $k = 2, \dots, n$. Portanto, a probabilidade de um instante k ser um ponto de mudança é

$$P(A_k|y_1, \dots, y_n) = \sum_{C_k} P(\{i_0, \dots, i_{l-1}, i_l = k-1, i_{l+1}, \dots, i_b\}|y_1, \dots, y_n)$$

$$\propto \sum_{C_k} \left(\prod_{r=1}^{l-1} c_{[i_{r-1}i_r]}^* \right) c_{[i_{l-1}(k-1)]}^* c_{[(k-1)i_{l+1}]}^* \left(\prod_{r=l+1}^b c_{[i_{r-1}i_r]}^* \right).$$

Similarmente, a probabilidade *a posteriori* para dois instantes específicos, k' e k^* , serem pontos de mudanças pode ser obtida a partir da probabilidade de $A_{k'} \cap A_{k^*}$, e assim por diante.

Considerando o item (2) da Definição 1.2.1, observa-se que o comportamento *a posteriori* do parâmetro $\theta_{[ij]}$ depende do bloco $y_{[ij]}$ a que está associado. Portanto, como mostrado em Barry & Hartigan [3], a distribuição *a posteriori* para cada θ_k , $k = 1, \dots, n$, é dada por

$$f(\theta_k|y_1, \dots, y_n) = \sum_{i=0}^{k-1} \sum_{j=k}^n r_{[ij]}^* f(\theta_k|y_{[ij]}), \quad (1.1)$$

onde $r_{[ij]}^*$ denota a relevância *a posteriori* do bloco $[ij]$, a qual é definida como sendo, $r_{[ij]}^* = P([ij] \in \rho|Y_1, \dots, Y_n)$. Barry & Hartigan [3] também mostraram que assumindo a perda quadrática, a estimativa produto de θ_k é obtida através da média

a posteriori de cada θ_k , ou seja,

$$\hat{\theta}_k = E(\theta_k | y_1, \dots, y_n) = \sum_{i=0}^{k-1} \sum_{j=k}^n r_{[ij]}^* E(\theta_k | y_{[ij]}).$$

A escolha da coesão $c_{[ij]}$ expressa a percepção inicial do pesquisador em relação ao grau de similaridade entre as observações pertencentes ao bloco $[ij]$ ou sobre o quão provável tais observações pertencerem ao mesmo cluster. Portanto, a escolha da coesão é fundamental no processo de inferência. Algumas coesões utilizadas frequentemente são:

- $c_{[ij]} = 1$: assumindo esta coesão para todos os clusters, tem-se que a distribuição *a priori* para ρ será uma uniforme no conjunto de todas as partições em blocos contíguos de I , ou seja, $P(\rho = \{i_0, \dots, i_b\}) = 2^{-(n-1)}$;
- $c_{[ij]} = \frac{1}{j-i}$: assumindo esta coesão, tem-se que a concepção inicial do pesquisador revela que as observações são poucos similares, pois este tipo de coesão atribui alta probabilidade a blocos com poucas observações, induzindo partições com muitos blocos.
- $c_{[ij]} = j-i$: assumindo esta coesão, tem-se que a concepção inicial do pesquisador revela que as observações são similares, pois este tipo de coesão atribui pesos maiores a blocos com muita observações, formando, com grande probabilidade, partições com poucos blocos.

No presente trabalho, será adotada a coesão proposta por Yao [44]. Seja p a probabilidade de que uma nova mudança ocorra em algum instante da sequência de observações. Então, a coesão *a priori* para o bloco $[ij]$ é dada por:

$$c_{[ij]} = \begin{cases} p(1-p)^{j-i-1}, & \text{se } j < n, \\ (1-p)^{i-j-1}, & \text{se } j = n, \end{cases} \quad (1.2)$$

onde $i, j \in I \cup 0$ e $i < j$. Tal coesão corresponde à probabilidade que uma nova mudança ocorra após $j-i$ instantes, dado que ocorreu uma mudança no instante i . Condicional em p , as distribuições *a priori* de ρ e B são, respectivamente, dadas por:

$$p(\rho = i_0, \dots, i_b | p) = p^{b-1} (1-p)^{n-b}, \quad b \in I,$$

$$P(B = b|p) = \binom{n-1}{b-1} p^{b-1} (1-p)^{n-b}, \quad b \in I.$$

Atribuindo uma distribuição *a priori* $\pi(p)$, obtem-se as distribuições *a priori* de ρ e B como segue. A distribuição *a priori* de ρ é dada por

$$\begin{aligned} P(\rho = \{i_0, \dots, i_b\}) &= \int_0^1 (\rho = \{i_0, \dots, i_b\} | p) \pi(p) dp \\ &= \int_0^1 p^{b-1} (1-p)^{n-b} \pi(p) dp, \end{aligned}$$

$$\begin{aligned} P(B = b) &= \int_0^1 P(B = b | p) \pi(p) dp \\ &= \binom{n-1}{b-1} \int_0^1 p^{b-1} (1-p)^{n-b} \pi(p) dp. \end{aligned}$$

Como consequência, tem-se que as distribuições *a posteriori* de ρ e B são dadas, respectivamente, por:

$$\begin{aligned} P(\rho = \{i_0, \dots, i_b\} | y_1, \dots, y_n) &= \frac{P(\rho; y_1, \dots, y_n)}{P(y_1, \dots, y_n)} \\ &\propto \int_0^1 P(y_1, \dots, y_n; \rho, p) dp \\ &\propto \int_0^1 P(y_1, \dots, y_n | \rho, p) P(\rho | p) \pi(p) dp \\ &\propto \int_0^1 \prod_{r=1}^b f_{i_{r-1}i_r}(y_{i_{r-1}i_r}) p^{b-1} (1-p)^{n-b} \pi(p) dp \\ &\propto \prod_{r=1}^b f_{i_{r-1}i_r}(y_{i_{r-1}i_r}) \int_0^1 p^{b-1} (1-p)^{n-b} \pi(p) dp \quad (1.3) \end{aligned}$$

e

$$\begin{aligned}
P(B = b \mid y_1, \dots, y_n) &= \sum_{C'} P(\rho = \{i_0, \dots, i_b\} \mid y_1, \dots, y_n) \\
&= \sum_{C'} \frac{P(\rho = \{i_0, \dots, i_b\}, y_1, \dots, y_n)}{P(y_1, \dots, y_n)} \\
&\propto \sum_{C'} \int_0^1 P(y_1, \dots, y_n \mid \rho, p) P(\rho \mid p) \pi(p) dp \\
&\propto \binom{n-1}{b-1} \prod_{r=1}^b f_{i_{r-1}i_r}(y_{i_{r-1}i_r}) \int_0^1 p^{b-1} (1-p)^{n-b} \pi(p) dp.
\end{aligned}$$

Além disso, tem-se que a probabilidade *a posteriori* de um particular instante k ser um ponto de mudança e a distribuição *a posteriori* de p são dadas, respectivamente, dada por:

$$\begin{aligned}
P(A_k \mid y_1, \dots, y_n) &= \sum_{C_k} P(\rho = \{i_0, \dots, i_{l-1}, i_l = k-1, i_{l+1}, \dots, i_b\} \mid y_1, \dots, y_n) \\
&= \sum_{C_k} \frac{P(\rho = \{i_0, \dots, i_{l-1}, i_l = k-1, i_{l+1}, \dots, i_b\}; y_1, \dots, y_n)}{P(y_1, \dots, y_n)} \\
&\propto \sum_{C_k} \int_0^1 P(y_1, \dots, y_n \mid \rho, p) P(\rho \mid p) \pi(p) dp \\
&\propto \sum_{C_k} \prod_{r=1}^b f(y_{i_{r-1}i_r}) \int_0^1 p^{b-1} (1-p)^{n-b} \pi(p) dp, \tag{1.4}
\end{aligned}$$

e

$$\begin{aligned}
P(p \mid y_1, \dots, y_n) &= \frac{P(y_1, \dots, y_n; p)}{P(y_1, \dots, y_n)} \\
&\propto \sum_C P(y_1, \dots, y_n; \rho, p) \\
&\propto \sum_C P(y_1, \dots, y_n \mid \rho, p) P(\rho \mid p) \pi(p) \\
&\propto \sum_C \prod_{r=1}^b f(y_{i_{r-1}i_r}) p^{b-1} (1-p)^{n-b} \pi(p), \tag{1.5}
\end{aligned}$$

onde C_k é o subconjunto de C que contém todas as partições que incluem o instante k como ponto de mudança.

Para completar as especificações do modelo pode-se utilizar uma distribuição *a priori* para p . Neste trabalho, $p \sim \text{Beta}(\alpha, \beta)$, com $\alpha > 0$ e $\beta > 0$. Consequentemente as, distribuições *a priori* de ρ e B serão, respectivamente:

$$\begin{aligned}
P(\rho = \{i_0, \dots, i_b\}) &= \int_0^1 P(\rho = \{i_0, \dots, i_b\} | p) \pi(p) dp \\
&= \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + b - 1) \Gamma(n + \beta - b)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(n + \alpha + \beta - 1)}
\end{aligned}$$

$$\begin{aligned}
P(B = b) &= \int_0^1 P(B = b | p) \pi(p) dp \\
&= \binom{n-1}{b-1} \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + b - 1) \Gamma(n + \beta - b)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(n + \alpha + \beta - 1)}.
\end{aligned}$$

Note que $B = W + 1$, onde W é uma variável aleatória com distribuição Beta-Binomial de parâmetros $n - 1$, α e β , para $\alpha > 0$ e $\beta > 0$, denotada por $W \sim Bb(n - 1, \alpha, \beta)$. A função de probabilidade de W é dada por:

$$f(W | n-1, \alpha, \beta) = \binom{n-1}{w} \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + w) \Gamma(n + \beta - w - 1)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(n + \alpha + \beta - 1)}, \quad w = 0, 1, \dots, n-1.$$

A esperança e a variância de W é dada, respectivamente, por

$$E(W) = (n - 1) \frac{\alpha}{\alpha + \beta}$$

e

$$Var(W) = (n - 1) \frac{\alpha\beta(\alpha + \beta + n - 1)}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Consequentemente, tem-se que a esperança e a variância da variável aleatória B são dadas, respectivamente, por

$$E(B) = (n - 1) \frac{\alpha}{\alpha + \beta} + 1$$

e

$$Var(B) = (n - 1) \frac{\alpha\beta(\alpha + \beta + n - 1)}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

A partir da Equação (1.3), tem-se que a distribuição a posteriori de ρ é:

$$P(\rho = \{i_0, \dots, i_b\} | y_1, \dots, y_n) \propto \prod_{r=1}^b f_{i_{r-1}i_r}(y_{i_{r-1}i_r}) \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + b - 1) \Gamma(n + \beta - b)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(n + \alpha + \beta - 1)}. \quad (1.6)$$

Enquanto a posteriori para B é dada por:

$$P(B = b \mid y_1, \dots, y_n) \propto \binom{n-1}{b-1} \prod_{r=1}^b f_{i_{r-1}i_r}(y_{i_{r-1}i_r}) \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + b - 1) \Gamma(n + \beta - b)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(n + \alpha + \beta - 1)}$$

Além disso, das equações (1.4) e (1.5) tem-se que as distribuições *a posteriori* de A_k e p são dadas, respectivamente, por:

$$P(A_k \mid y_1, \dots, y_n) \propto \sum_{C_k} \prod_{r=1}^b f(y_{i_{r-1}i_r}) \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + \beta - 1) \Gamma(n + \beta - b)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(n + \alpha + \beta - 1)}$$

$$P(p \mid y_1, \dots, y_n) \propto \sum_C \prod_{r=1}^b f(y_{i_{r-1}i_r}) p^{\alpha+b-2} (1-p)^{n+\beta-b-1}. \quad (1.7)$$

1.3 Métodos computacionais

Observe que não é trivial gerar amostras de ρ a partir da distribuição condicional completa, pois a dimensão de ρ não é fixa e o número de partições possíveis cresce a medida que n aumenta. Para contornar esta dificuldade Barry & Hartigan [4] propôs uma transformação de variáveis em que ρ é representada como um vetor cujas componentes são variáveis binárias. Ao fazer isto, pode-se gerar da distribuição *a posteriori* de ρ via amostrador de Gibbs. Tal procedimento é mostrado a seguir.

Considere uma quantidade aleatória U_r que reflete se houve ou não mudança no instante $r + 1$, ou seja,

$$U_r = \begin{cases} 1, & \text{se } \theta_r = \theta_{r+1}, \\ 0, & \text{se } \theta_r \neq \theta_{r+1}, \end{cases}$$

para $r = 1, \dots, n - 1$. Assumindo esta transformação, cada valor de ρ pode ser identificado por um valor do vetor $\mathbf{U} = (U_1, \dots, U_{n-1})$, isto é,

$$\rho = \{i_0, \dots, i_b\} \Leftrightarrow (U_1 = 1, \dots, U_{i_1-1} = 1, U_{i_1} = 0, U_{i_1+1} = 1, U_{i_2-1} = 0, U_{i_2} = 0, \dots, U_{i_{b-1}-1} = 1, U_{i_{b-1}} = 0, U_{i_{b-1}+1} = 1, \dots, U_{i_b} = 1).$$

Para melhor compreensão, ver Tabela 1.2.

Uma das vantagens de se considerar a transformação \mathbf{U} é que este terá dimensão fixa em cada parte do algoritmo. Então, cada partição $(U_1^s, \dots, U_{n-1}^s)$, $s \geq 1$, é gerada a partir das distribuições condicionais completas *a posteriori*, isto é, para r -

Tabela 1.2: Relação entre ρ e U

Bloco de observações	Valores de ρ	Valores de U
$[y_1, y_2, y_3]$	$\{i_0 = 0, i_1 = 3\}$	$\{U_1 = 1, U_2 = 1\}$
$[y_1], [y_2, y_3]$	$\{i_0 = 0, i_1 = 1, i_2 = 3\}$	$\{U_1 = 0, U_2 = 1\}$
$[y_1, y_2], [y_3]$	$\{i_0 = 0, i_1 = 2, i_2 = 3\}$	$\{U_1 = 1, U_2 = 0\}$
$[y_1], [y_2], [y_3]$	$\{i_0 = 0, i_1 = 1, i_2 = 2, i_3 = 3\}$	$\{U_1 = 0, U_2 = 0\}$

ésimo elemento no passo s do algoritmo, U_r^s de \mathbf{U} , é gerada a partir da distribuição condicional

$$U_r | U_1^s, \dots, U_{r-1}^s, U_{r+1}^{s-1}, \dots, U_{n-1}^{s-1}, p, y_1, \dots, y_n,$$

para $r = 1, \dots, n-1$. Como U_r é uma variável binária, para gerar amostras de \mathbf{U} , sugere-se considerar a seguinte razão:

$$R_r = \frac{P(U_r = 1 | V_r^s, p, y_1, \dots, y_n)}{P(U_r = 0 | V_r^s, p, y_1, \dots, y_n)},$$

para $r = 1, \dots, n-1$ e $V_r^s = \{U_1^s = u_1, \dots, U_{r-1}^s = u_{r-1}, U_{r+1}^{s-1} = u_{r+1}, \dots, U_{n-1}^{s-1} = u_{n-1}\}$. Assumindo a coesão apresentada na Equação (1.2) e fazendo $p \sim \text{Beta}(\alpha, \beta)$, com $\alpha > 0$ e $\beta > 0$, então o valor de R_r é dado por:

$$R_r = \frac{f_{[xl]}(y_{[xl]})\Gamma(n + \beta - b + 1)\Gamma(b + \alpha - 2)}{f_{[xr]}(y_{[xr]})f_{[rl]}(y_{[rl]})\Gamma(b + \alpha - 1)\Gamma(n + \beta - b)},$$

para $b = 1, \dots, n$,

$$x = \begin{cases} \max\{r^*; 0 < r^* < r, U_{r^*}^s = 0\}, & \text{se } U_{r^*}^s = 0, r^* \in \{1, \dots, r-1\}, \\ 0, & \text{caso contrário,} \end{cases}$$

e

$$l = \begin{cases} \min\{r^*; r < r^* < n, U_{r^*}^{s-1} = 0\}, & \text{se } U_{r^*}^{s-1} = 0, r^* \in \{r+1, \dots, n-1\}, \\ n, & \text{caso contrário.} \end{cases}$$

para mais detalhes, ver Loschi & Cruz [22].

Usando o amostrador de Gibbs, gera-se amostra de \mathbf{U} considerando o seguinte critério para a simulação da variável aleatória U_r^s :

$$U_r^s = \begin{cases} 1, & \text{se } R_r \geq \frac{1-z}{z}, \\ 0, & \text{caso contrário,} \end{cases}$$

onde $r = 1, \dots, n-1$ e $z \sim U(0, 1)$.

Após gerar um número grande de amostras do objeto aleatório (U_1, \dots, U_{n-1}) ,

a estimativa Monte Carlo para a probabilidade *a posteriori* de cada partição $\rho = \{i_0, \dots, i_b\}$ é dada pela proporção de vetores $(U_1 = 1, \dots, U_{i_1-1} = 1, U_{i_1} = 0, U_{i_1+1} = 1, U_{i_2-1} = 0, U_{i_2} = 0, \dots, U_{i_{b-1}-1} = 1, U_{i_{b-1}} = 0, U_{i_{b-1}+1} = 1, \dots, U_{i_b} = 1)$ que apareceram no total de amostras úteis. Além disso, a amostra da distribuição *a posteriori* para número de blocos em ρ pode ser obtida considerando-se a seguinte relação

$$B = 1 + \sum_{r=1}^{n-1} (1 - U_r).$$

Consequentemente, para cada vetor gerado $(U_1^s, \dots, U_{n-1}^s)$ obtém-se um valor $B^s = 1 + \sum_r^{n-1} (1 - U_r^s)$ e a estimativa Monte Carlo para probabilidade *a posteriori* do número de blocos é dada por

$$P(B = b | y_1, \dots, y_n) = \frac{\sum_{s=1}^T \mathbf{1}\{B^s = b\}}{T}, \forall b \in I,$$

onde T é o número total de amostras *a posteriori* e $\mathbf{1}\{B^s = b\}$ é a função indicadora.

Além disso, Loschi & Cruz [25] mostra que a estimativa Monte Carlo para a probabilidade *a posteriori* do instante k ser um ponto de mudança é dada por

$$P(A_k | y_1, \dots, y_n) = \frac{N}{T},$$

onde $k = 2, \dots, n$, N é o número de vetores observados com $U_{k-1} = 0$ e T é o tamanho da amostra da distribuição *a posteriori*.

Assumindo que $p \sim \text{Beta}(\alpha, \beta)$, com $\alpha > 0$ e $\beta > 0$, como mostrado em Loschi & Cruz [25], tem-se que cada amostra da distribuição *a posteriori* de p é da seguinte distribuição condicional completa *a posteriori*:

$$p^s | \theta, \rho, y_1, \dots, y_n \sim \text{beta}(b^s + \alpha - 1; n + \beta - b^s),$$

onde b^s é o número de blocos observados no vetor $(U_1^s, \dots, U_{n-1}^s)$.

No próximo capítulo, aplica-se o MPP para a identificação de pontos de mudança no vetor de médias e na matriz de variância e covariância de dados multivariados e normalmente distribuídos. Os resultados estendem o trabalho de Moura [33], pois assume que o vetor de médias também pode sofrer modificações ao longo do tempo.

Capítulo 2

Identificando múltiplas mudanças na média e na covariância de dados Normal Multivariados

O estudo de séries temporais multivariadas permite avaliar a influência do comportamento que uma determinada série exerce sobre as demais. No estudo de dados financeiros, por exemplo, tal influência é denominada contágio e pode ser mensurada através da correlação entre as séries. Nesse sentido, a identificação de múltiplos pontos de mudanças no contexto multivariado é mais uma forma de estudar as relações existente entre as séries.

Diante disso, alguns métodos foram utilizados para a identificação de múltiplos pontos de mudanças, no qual destacam-se o de Moura [33] e Cheon & Kim [8]. Moura [33], por exemplo, utilizou o modelo partição produto para a identificação de mudanças na matriz de variância e covariância para uma sequência de dados com distribuição normal p -variada. Cheon & Kim [8] aplicaram o algoritmo Monte Carlo com aproximação estocástica para a detecção de múltiplos pontos de mudanças no vetor de médias e na matriz de variância e covariância de dados com distribuição normal p -variada. No entanto, Cheon & Kim [8] consideraram uma distribuição *a priori* vaga para o vetor de médias, ou seja, $\pi(\boldsymbol{\mu}_k) = 1$ e estimam apenas a partição dos dados, o número de blocos e a probabilidade de cada instante ser um ponto de mudança, não fornecendo estimativas para os parâmetros no instante k e, conseqüentemente, não havendo a identificação em qual parâmetro e em qual variável houve a mudança.

Neste capítulo, tem-se como meta estudar o comportamento de séries temporais multivariadas e avaliar se há mudanças ao longo do tempo tanto no vetor de médias quanto na matriz de variância-covariância. Assim, tem-se como avaliar, por exemplo, se as correlações entre as séries se modificam ao longo do tempo. Para isso,

será considerado o Modelo Partição Produto para a detecção de múltiplos pontos de mudanças no vetor de médias e na matriz de variância e covariância de uma sequência de vetores aleatórios com distribuição normal multivariada, no qual, diferente de Cheon & Kim [8], distribuições *a priori* conjugadas serão utilizadas para o vetor de médias e para matriz de variância covariância. Estimativas do vetor de médias e da matriz de variância e covariância para cada instante do tempo também serão fornecidas. Além disso, será apresentado um procedimento para comparações sequencias dos parâmetros de forma a auxiliar na identificação de mudanças significativas nos parâmetros de diferentes séries. Na Seção 2.1 são apresentadas alguns resultados úteis para a construção do modelo partição produto para dados multivariados normalmente distribuídos (MPPNM). Na Seção 2.2 é apresentada a construção do MPPNM. Na Seção 2.3 é fornecido o algoritmo para gerar da distribuição *a posteriori* conjunta de $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ e para realizar as comparações sequenciais dos parâmetros.

2.1 Inferência Bayesiana no modelo normal multivariado

Neste primeiro momento, serão apresentados alguns resultados úteis para a construção do MPPNM apresentado na Seção 2.2. As funções densidade de probabilidade das distribuições consideradas neste trabalho, normal p -variada, t -matriz variada e Wishart-inversa, são apresentadas no Apêndice A. As propriedades matriciais utilizadas no trabalho estão no Apêndice C.

Considere uma amostra aleatória $\mathbf{y}_1, \dots, \mathbf{y}_n$ de vetores $p \times 1$ provenientes de uma distribuição normal p -variada com vetor de médias $\boldsymbol{\mu}$ de dimensão $p \times 1$ e matriz de variância-covariância $\boldsymbol{\Sigma}$ de dimensão $p \times p$, a qual é denotada por $\mathbf{y}_1, \dots, \mathbf{y}_n \mid \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Considerando Equação (A.1) do Apêndice A, obtém-se que a verossimilhança, associada a este experimento, é dada por:

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y}_1, \dots, \mathbf{y}_n) = (2\pi)^{(-np/2)} |\boldsymbol{\Sigma}|^{-n/2} \times \exp \left\{ -\frac{1}{2} [(n-1)tr[\boldsymbol{\Sigma}^{-1}\mathbf{S}] + n(\bar{\mathbf{y}} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu})] \right\},$$

onde

$$\bar{\mathbf{y}} = \sum_{k=1}^n \frac{1}{n} \mathbf{y}_k$$

é o vetor de médias amostrais de dimensão $p \times 1$ e

$$\mathbf{S} = \frac{1}{n-1} \sum_{k=1}^n (\mathbf{y}_k - \bar{\mathbf{y}})(\mathbf{y}_k - \bar{\mathbf{y}})^t$$

é a matriz de variância covariância amostral de dimensão $p \times p$. O desenvolvimento função de verossimilhança é apresentado no Apêndice B e, para mais detalhes, ver Johnson *et al.* [16].

Uma forma alternativa de escrever, em termos matriciais, a função de verossimilhança é:

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{Y}) = (2\pi)^{(-np/2)} |\boldsymbol{\Sigma}|^{-n/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{1}_n \boldsymbol{\mu}^t)^t \mathbb{I}_n (\mathbf{Y} - \mathbf{1}_n \boldsymbol{\mu}^t) \right] \right\}, \quad (2.1)$$

onde $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^t$, \mathbb{I}_n é a matriz identidade de ordem $n \times n$ e $\mathbf{1}_n$ denota um vetor de uns de ordem $n \times 1$. A verossimilhança apresentada na Equação (2.1) foi utilizada para no cálculo da distribuição preditiva *a priori*. A Proposição 2.1.1 apresenta alguns resultados úteis quando distribuições *priori* conjugadas são utilizadas para descrever a incerteza sobre $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. A demonstração dessa proposição pode ser encontrada no Apêndice B.

Denote por $WI(\mathbf{D}, d)$ a distribuição Wishart-inversa com parâmetros $d \in \mathfrak{R}^+$, tal que $d \geq p$ e \mathbf{D} é uma matriz positiva definida de dimensão $p \times p$. Denote por $t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ a distribuição t-Student multivariada com locação $\boldsymbol{\mu} \in \mathfrak{R}^p$, matriz de dispersão $\boldsymbol{\Sigma} \in \mathfrak{R}^{p \times p}$ e graus de liberdade ν . Denote por $\mathcal{T}_{n \times p}(\nu, \mathbf{M}, \boldsymbol{\Omega}, \boldsymbol{\Sigma})$ a distribuição t-Student matriz variada, onde $\boldsymbol{\Omega}$ é uma matriz positiva definida de dimensão $n \times n$, $\boldsymbol{\Sigma}$ é uma matriz positiva definida de dimensão $p \times p$, \mathbf{M} uma matriz de locação de dimensão $n \times p$ e ν é o grau de liberdade. As respectivas funções densidades de probabilidade e algumas propriedades encontram-se no Apêndice A.

Proposição 2.1.1 *Seja $\mathbf{y}_1, \dots, \mathbf{y}_n$ uma amostra aleatória de vetores, $p \times 1$, de uma distribuição normal p -variada, ou seja, $\mathbf{y}_k | \boldsymbol{\mu}, \boldsymbol{\Sigma} \stackrel{iid}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, onde $\boldsymbol{\mu}$ é um vetor de médias de dimensão $p \times 1$, $\boldsymbol{\Sigma}$ é a matriz de variância-covariância de dimensão $p \times p$ e $k = 1, \dots, n$. Se, a priori, $\boldsymbol{\mu} | \boldsymbol{\Sigma} \sim N_p\left(\mathbf{m}, \frac{1}{v} \boldsymbol{\Sigma}\right)$ e $\boldsymbol{\Sigma} \sim WI(\mathbf{D}, d)$, em que \mathbf{m} é um vetor de médias de dimensão $p \times 1$ e v e d são números reais positivos, então tem-se que:*

a) *A distribuição a posteriori conjunta de $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ é tal que*

$$\begin{aligned} \boldsymbol{\mu} | \boldsymbol{\Sigma}, \mathbf{y}_1, \dots, \mathbf{y}_n &\sim N_p(\mathbf{M}, (v+n)^{-1} \mathbf{D}^*); \\ \boldsymbol{\Sigma} | \mathbf{y}_1, \dots, \mathbf{y}_n &\sim WI(\mathbf{D}^*, d+n); \end{aligned}$$

b) a distribuição a posteriori de $\boldsymbol{\mu}$ é

$$\boldsymbol{\mu} | \mathbf{y}_1, \dots, \mathbf{y}_n \sim t_{d+n+1-p}(\mathbf{M}, (d+n+1-p)^{-1}(v+n)^{-1}\mathbf{D}^*);$$

c) a distribuição preditiva a priori de \mathbf{Y} é

$$\mathbf{Y} \sim \mathcal{T}_{n \times p} \left(d+1-p, \mathbf{1}_n \mathbf{m}^t, \mathbb{I}_n + \frac{\mathbf{1}_n \mathbf{1}_n^t}{v}, \mathbf{D} \right),$$

onde $\mathbf{M} = (n+v)^{-1}(n\bar{\mathbf{y}} + v\mathbf{m})$ e $\mathbf{D}^* = \mathbf{D} + (n-1)\mathbf{S} + \frac{nv}{n+v} [(\bar{\mathbf{y}} - \mathbf{m})(\bar{\mathbf{y}} - \mathbf{m})^t]$.

Os resultados apresentados na Proposição 2.1.1 são conhecidos e serão úteis na obtenção das distribuições *a posteriori* envolvidas no MPP desenvolvido na próxima seção.

2.2 Construção do MPP

O objetivo nesta seção é construir um modelo que permita identificar múltiplas mudanças, ao longo do tempo, no vetor de médias e na matriz de covariância quando os dados são normalmente distribuídos. Para isso, será utilizado o modelo descrito no Capítulo 1.

Seja $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ uma sequência de vetores aleatórios de dimensão $p \times 1$, onde $\mathbf{Y}_k = (Y_{1k}, \dots, Y_{pk})^t$, $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n$ é uma sequência de vetores das médias de dimensão $p \times 1$ e $\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_n$ é uma sequência de matrizes de variância covariância, de ordem $p \times p$, simétrica e positiva definida. O modelo será construído assumindo que, para $k = 1, \dots, n$,

$$\begin{aligned} \mathbf{Y}_k | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k &\stackrel{iid}{\sim} N_p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \\ \boldsymbol{\mu}_k | \boldsymbol{\Sigma}_k &\stackrel{iid}{\sim} N_p(\mathbf{m}, \frac{1}{v}\boldsymbol{\Sigma}_k) \\ \boldsymbol{\Sigma}_k &\stackrel{iid}{\sim} WI(\mathbf{D}, d), \end{aligned}$$

onde \mathbf{m} é um vetor de médias, $p \times 1$, \mathbf{D} é uma matriz positiva de dimensão $p \times p$ e v e d são reais positivos.

Considerando a notação similar a apresentada no Capítulo 1, denote por

$$\mathbf{Y}_{[ij]} = (\mathbf{y}_{i+1}, \mathbf{y}_{i+2}, \dots, \mathbf{y}_j)^t = \begin{pmatrix} y_{(i+1)1} & \cdots & y_{(i+1)p} \\ y_{(i+2)1} & \cdots & y_{(i+2)p} \\ \vdots & \ddots & \vdots \\ y_{j1} & \cdots & y_{jp} \end{pmatrix}$$

a matriz bloco de observações de dimensão $(j - i) \times p$. Além disso, denote por $\boldsymbol{\mu}_{[ij]}$ e $\boldsymbol{\Sigma}_{[ij]}$ o vetor de médias e a matriz de covariância que indexam a distribuição das observações no bloco $\mathbf{Y}_{[ij]}$. Adicionalmente, assuma que

- (i) As observações $\mathbf{y}_{i+1}, \mathbf{y}_{i+2}, \dots, \mathbf{y}_j$ são independentes e identicamente distribuídas, condicionalmente em $\boldsymbol{\mu}_{[ij]}$ e $\boldsymbol{\Sigma}_{[ij]}$, com uma distribuição normal p -variada, isto é:

$$\mathbf{Y}_k | \boldsymbol{\mu}_{[ij]}, \boldsymbol{\Sigma}_{[ij]} \stackrel{iid}{\sim} N_p(\boldsymbol{\mu}_{[ij]}, \boldsymbol{\Sigma}_{[ij]}) \quad \text{para } k = i + 1, \dots, j;$$

- (ii) Dada uma partição $\rho = \{i_0, \dots, i_n\}$, existem os parâmetros comuns $(\boldsymbol{\mu}_{[i_0 i_1]}, \boldsymbol{\Sigma}_{[i_0 i_1]}), \dots, (\boldsymbol{\mu}_{[i_{b-1} i_b]}, \boldsymbol{\Sigma}_{[i_{b-1} i_b]})$ que são independentes e identicamente distribuídos tais que

$$\begin{aligned} \boldsymbol{\mu}_{[ij]} | \boldsymbol{\Sigma}_{[ij]} &\sim N_p\left(\mathbf{m}, \frac{1}{v} \boldsymbol{\Sigma}_{[ij]}\right) \\ \boldsymbol{\Sigma}_{[ij]} &\sim WI(\mathbf{D}, d). \end{aligned}$$

- (iii) Dada uma partição $\rho = \{i_0, \dots, i_n\}$ e os parâmetros comuns $(\boldsymbol{\mu}_{[i_0 i_1]}, \boldsymbol{\Sigma}_{[i_0 i_1]}), \dots, (\boldsymbol{\mu}_{[i_{b-1} i_b]}, \boldsymbol{\Sigma}_{[i_{b-1} i_b]})$, as observações em diferentes blocos $\mathbf{Y}_{i_0, i_1}, \dots, \mathbf{Y}_{i_{b-1}, i_b}$ são independentes.

Denote por $n_r = i_r - i_{r-1}$ e $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n)$, onde $\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ para $k = 1, \dots, n$. Considerando as suposições (i) e (iii), obtém-se a verossimilhança como

$$\begin{aligned} L(\boldsymbol{\theta}, \rho | \mathbf{y}_1, \dots, \mathbf{y}_n) &= \prod_{r=1}^b f(\mathbf{y}_{[i_{r-1} i_r]} | \boldsymbol{\mu}_{[i_{r-1} i_r]}, \boldsymbol{\Sigma}_{[i_{r-1} i_r]}) \\ &= \prod_{r=1}^b \prod_{k=i_{r-1}+1}^{i_r} f(\mathbf{y}_k | \boldsymbol{\mu}_{[i_{r-1} i_r]}, \boldsymbol{\Sigma}_{[i_{r-1} i_r]}) \\ &= \prod_{r=1}^b \left((2\pi)^{-\frac{n_r p}{2}} |\boldsymbol{\Sigma}_{[i_{r-1} i_r]}|^{-\frac{n_r}{2}} \exp \left\{ -\frac{1}{2} \left\{ (n_r - 1) \text{tr} \left[\boldsymbol{\Sigma}_{[i_{r-1} i_r]}^{-1} \mathbf{S}_{[i_{r-1} i_r]} \right] \right\} \right\} \right. \\ &\quad \left. \times \exp \left\{ n_r (\bar{\mathbf{y}}_{[i_{r-1} i_r]} - \boldsymbol{\mu}_{[i_{r-1} i_r]})^t \boldsymbol{\Sigma}_{[i_{r-1} i_r]}^{-1} (\bar{\mathbf{y}}_{[i_{r-1} i_r]} - \boldsymbol{\mu}_{[i_{r-1} i_r]}) \right\} \right) \end{aligned} \quad (2.2)$$

Como consequência das suposições anteriores, segue que a distribuição *a posteriori* de $\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ é dada pela expressão em 1.1, em que

$$f(\boldsymbol{\theta}_k | \mathbf{y}_{i+1}, \dots, \mathbf{y}_j) = f(\boldsymbol{\mu}_k | \boldsymbol{\Sigma}_k, \mathbf{y}_{i+1}, \dots, \mathbf{y}_j) f(\boldsymbol{\Sigma}_k | \mathbf{y}_{i+1}, \dots, \mathbf{y}_j).$$

Dos resultados apresentados na Proposição 2.1.1, segue que

$$\begin{aligned}\bar{\boldsymbol{\mu}}_k \mid \boldsymbol{\Sigma}_k, \rho, \mathbf{y}_{i+1}, \dots, \mathbf{y}_j &\sim N_p(\mathbf{M}_{[ij]}, (j-i+v)^{-1}\boldsymbol{\Sigma}_{[ij]}), \\ \boldsymbol{\Sigma}_k \mid \rho, \mathbf{y}_{i+1}, \dots, \mathbf{y}_j &\sim WI(\mathbf{D}_{[ij]}^*, d+j-i),\end{aligned}$$

onde $\mathbf{M}_{[ij]} = (j-i+v)^{-1}((j-i)\bar{\mathbf{y}}_{[ij]} + v\mathbf{m})$, $\mathbf{D}_{[ij]}^* = \mathbf{D} + (j-i-1)\mathbf{S}_{[ij]} + \frac{(j-i)v}{j-i+v} [(\bar{\mathbf{y}}_{[ij]} - \mathbf{m})(\bar{\mathbf{y}}_{[ij]} - \mathbf{m})^t]$, $\mathbf{S}_{[ij]} = \frac{1}{j-i-1} \sum_{k=i+1}^j (\mathbf{y}_k - \bar{\mathbf{y}}_{[ij]})(\mathbf{y}_k - \bar{\mathbf{y}}_{[ij]})^t$ e $\bar{\mathbf{y}}_{[ij]} = \sum_{k=i+1}^j \frac{1}{j-i} \mathbf{y}_k$.

As distribuições *a posteriori* para ρ e p são obtidas, respectivamente, das Equações (1.6) e (1.7), onde o fator dados é dado pela seguinte distribuição preditiva *a priori* por blocos

$$\mathbf{Y}_{[ij]} \sim \mathcal{T}_{(j-i) \times p} \left(d+1-p, \mathbf{1}_{(j-i)} \mathbf{m}^t, \mathbb{I}_{(j-i)} + \frac{\mathbf{1}_{(j-i)} \mathbf{1}_{(j-i)}^t}{v}, \mathbf{D} \right),$$

obtida no item (c) da Proposição 2.1.1.

2.3 Falta de identificabilidade na identificação da mudança

Na utilização do MPP na detecção de pontos de mudança podem ocorrer problemas de falta de identificabilidade, ou seja, a mudança pode ocorrer apenas no vetor de médias e, no entanto, o modelo identifica mudanças também na matriz de variância e covariância, por exemplo.

Para auxiliar na identificação dos parâmetros que, de fato, experimentaram mudança, minimizando o problema de falta de identificabilidade, aqui propõe-se considerar o estudo do comportamento da distribuição *a posteriori* de $\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}$ e para todo $k = 2, \dots, n$ testar as hipóteses

$$\begin{aligned}H_0 : \boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1} &= 0 \\ H_a : \boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1} &\neq 0.\end{aligned}$$

Caso H_0 seja aceita, tem-se evidências de que os parâmetros sucessivos no tempo não são substancialmente diferentes. A evidência a favor de H_0 pode ser obtida usando o teste de significância Bayesiano Completo (Bragança Pereira & Stern [7]). No entanto, decisões baseadas em tal teste é equivalente a decidir por H_0 se o valor zero pertencer a região de maior densidade *a posteriori* (regiões HPD). Considerando a estrutura do MPP, é mais simples obter-se as regiões HPD, uma vez que tem-se

um algoritmo eficiente para gerar das distribuições de cada $\boldsymbol{\theta}_k$. O seguinte algoritmo é proposto para amostrar da distribuição produto dos parâmetros $\boldsymbol{\theta}_k$ e da diferença *a posteriori* de $\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}$.

Para $s = 1, \dots, T$,

1. Gere uma partição ρ^s de $\pi(\rho | \boldsymbol{\mu}_k^{s-1}, (\boldsymbol{\Sigma}_k)^{s-1}, p, \mathbf{y}_1, \dots, \mathbf{y}_n)$;

2. Para cada $k = 1, \dots, n$

Encontre o bloco $[ij]$ na partição ρ^s tal que $k \in [ij]$;

Gere uma amostra $(\boldsymbol{\mu}_k^s, (\boldsymbol{\Sigma}_k)^s)$ a partir da distribuição *a posteriori* por bloco de $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, isto é:

$$\begin{aligned} \boldsymbol{\mu}_k^s | \rho^s, \boldsymbol{\Sigma}_k^s, \mathbf{y}_{i+1}, \dots, \mathbf{y}_j &\sim N_p(\mathbf{M}_{[ij]}, (j-i+v)^{-1} \boldsymbol{\Sigma}_k^s); \\ \boldsymbol{\Sigma}_k^s | \rho^s, \mathbf{y}_{i+1}, \dots, \mathbf{y}_j &\sim WI(\mathbf{D}_{[ij]}^*, d+j-i). \end{aligned}$$

3. Para $k = 2, \dots, n$, calcule $\boldsymbol{\theta}_k^{(s)} - \boldsymbol{\theta}_{k-1}^{(s)}$;

4. Determine a região HPD para $\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}$.

Loschi *et al.* [28] propôs uma aproximação da estimação produto a partir das seguintes estimações Monte Carlo:

$$\hat{\boldsymbol{\mu}}_k = \frac{\sum_{s=1}^T \boldsymbol{\mu}_k^s}{T} \quad \text{e} \quad \hat{\boldsymbol{\Sigma}}_k = \frac{\sum_{s=1}^T (\boldsymbol{\Sigma}_k)^s}{T}.$$

Além das estimativas produto proposta por Barry & Hartigan [3] e Barry & Hartigan [4], neste trabalho, também foram utilizadas como medidas resumos a mediana e intervalos de credibilidades *a posteriori* de cada parâmetro.

No próximo capítulo será realizada uma análise de sensibilidade para MPPNMV através de cenários que possibilita avaliar o desempenho do modelo em algumas situações.

Capítulo 3

Análise de sensibilidade do Modelo Partição Produto para Normal Multivariada

Neste capítulo, será feita uma análise de sensibilidade no MPPNM proposto no Capítulo 2. O objetivo é observar o efeito de diferentes especificações *a priori* para $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ nas inferências *a posteriori* das partições, do número de blocos, da probabilidade de cada instante ser um ponto de mudança e de haver uma mudança em um instante qualquer do tempo. Além disso, também estamos interessados em como essas diferentes especificações influenciam nas estimações de $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. Para isso, três cenários foram propostos.

Nos cenários 1 e 2 são geradas uma sequência de 100 amostras aleatórias de ordem 2×1 , no qual o vetor \mathbf{y}_k , é gerado a partir de uma distribuição normal $N_2(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, onde $\boldsymbol{\mu}_k$ é um vetor de médias de ordem 2×1 e $\boldsymbol{\Sigma}_k$ é a matriz de variância covariância de ordem 2×2 . Para o cenário 1, descrito na seção 3.1, foi considerada uma mudança na média de uma das séries. Para o cenário 2, descrito na 3.2, foi considerada uma mudança na variância de uma das séries.

Foi assumido que, dada a partição, os parâmetros comuns possuem as seguintes distribuições *a priori*

$$\boldsymbol{\mu}_{[ij]} | \boldsymbol{\Sigma}_{[ij]} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{v} \boldsymbol{\Sigma}_{[ij]} \right),$$
$$\boldsymbol{\Sigma}_{[ij]} \sim WI(\mathbf{D}, d),$$

em que diferentes valores de v , d e D serão considerados em cada cenário. Resumos das distribuições *a priori* consideradas podem ser encontrados nas Tabelas 3.1 e 3.4.

O cenário 3 é baseado na geração apresentada em Cheon & Kim [8]. Foi gerada uma série de tamanho 1000 de ordem 2×1 , no qual o vetor \mathbf{y}_k é gerado de uma

distribuição normal $N_2(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. Foram consideradas oito pontos de mudança em todos os parâmetros. Detalhes sobre os valores dos parâmetros considerados neste cenário estão na seção 3.3. Para analisar os dados foi assumido que, dada a partição, os parâmetros comuns possuem as seguintes distribuições *a priori*:

$$\begin{aligned} \boldsymbol{\mu}_{[ij]} | \boldsymbol{\Sigma}_{[ij]} &\sim N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{0.01} \boldsymbol{\Sigma}_{[ij]} \right], \\ \boldsymbol{\Sigma}_{[ij]} &\sim WI \left(\begin{pmatrix} 0.1 & 0.01 \\ 0.01 & 0.1 \end{pmatrix}, 4 \right). \end{aligned}$$

Assumindo essas especificações, as estimativas *a priori* para $\boldsymbol{\mu}_{[ij]}$ e $\boldsymbol{\Sigma}_{[ij]}$ são apresentadas na Tabela 3.7.

Como em Moura [33], assume-se que $p \sim Beta(2, 198)$ nos três cenários. Com isto, tem-se que a média e o desvio padrão *a priori* de p são, respectivamente, 0.01 e 0.0070181 e, conseqüentemente, a média e o desvio padrão *a priori* de b são, respectivamente, 1.99 e 1.20746, o que é compatível com o número de mudanças nas séries consideradas nos cenários 1 e 2.

Foram geradas 10000 amostras do vetor \mathbf{U} partindo de um vetor inicial ($U_1 = 0, U_2 = 0, \dots, U_{n-1} = 0$), onde $n = 100$ para os cenários 1 e 2 e $n = 1000$ para o cenário 3. Para a formação da amostra *a posteriori*, foram descartadas as 4000 primeiras gerações e, com *lag* 10, foram selecionadas 600 observações das 6000 restantes. O algoritmo foi codificado em R Core Team [39] e processado no computador Intel Core i7, 3.40GHz e 8 GB RAM. Os tempos estimados para os cenários 1 e 2 foram de aproximadamente 60 minutos. Para o cenário 3, foi de aproximadamente três dias. No cenário 3, houve o problema de *overflow*, e para contorná-lo foi utilizada a função *mpfr* do pacote *Rmpfr*, Maechler [30]. As cadeias de alguns parâmetros podem ser vistos no Apêndice D.

3.1 Cenário 1: mudança na média

Neste cenário, tem-se como objetivo avaliar o desempenho do MPPNM sobre diferentes óticas quando há uma mudança na média de apenas uma das séries. Para isso, uma série bivariada de tamanho 100 foi simulada considerando um ponto de mudança no instante 51 da seguinte forma:

$$\mathbf{Y}_k \sim \begin{cases} N_2 \left[\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{pmatrix} \right] & \text{para } k = 1, \dots, 50, \\ N_2 \left[\begin{pmatrix} 2 \\ 2.5 \end{pmatrix}, \begin{pmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{pmatrix} \right] & \text{para } k = 51, \dots, 100. \end{cases}$$

Esta série é apresentada na Figura 3.1. Para avaliar o efeito da escolha do parâmetro \mathbf{D} que indexa a distribuição *a priori* de $\Sigma_{[ij]}$, três situações são consideradas. As Situações 1, 2 e 3 são diferenciadas pelas escolhas de diferentes \mathbf{D} , assumindo, respectivamente, os valores $\mathbf{D} = \begin{pmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{pmatrix}$, $\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ e $\mathbf{D} = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$. Diferentes valores de d ($d = 4, d = 6$ e $d = 20$) e v ($v = 0.01, v = 1$ e $v = 10$) também são assumidos.

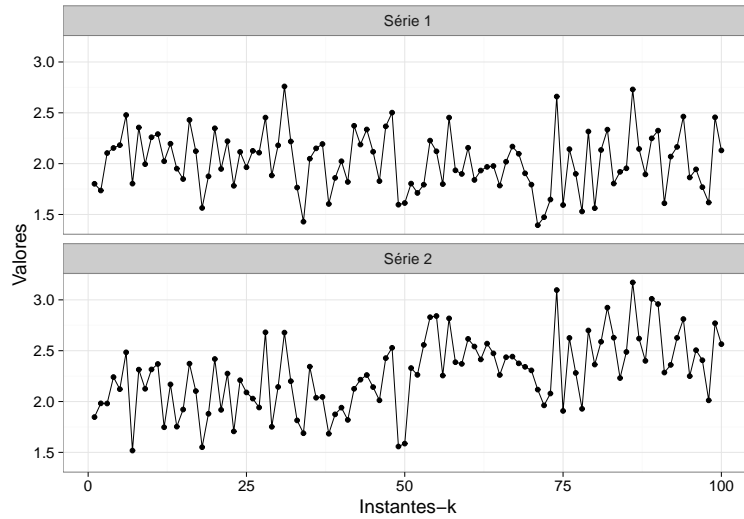


Figura 3.1: Valores das séries para o cenário 1.

A Tabela 3.1, mostra todas as distribuições *a priori* consideradas e alguns de seus resumos sobre o efeito das escolhas de \mathbf{D} , v e d nas estimativas *a priori* do vetor de médias e da matriz de covariância. Menores valores escolhidos para v indicam maior variabilidade e portanto a incerteza *a priori* sobre o vetor de médias aumenta. Em relação a escolha de d , percebe-se que o aumento do valor de d induz uma distribuição *a priori* muito informativa para a matriz de covariância.

Tabela 3.1: Estatísticas *a priori*, $l = 1, 2$.

Situação 1							
d	v	$E(\mu_l)$	$Var(\mu_l)$	$E(\sigma_l^2)$	$Var(\sigma_l^2)$	$E(\sigma_{12}^2)$	$Var(\sigma_{12}^2)$
4	0.01	0	10.0000	0.1000	-	0.09	-
4	1	0	0.1000	0.1000	-	0.09	-
4	10	0	0.0100	0.1000	-	0.09	-
6	0.01	0	3.3333	0.0333	0.0022	0.03	0.0053
6	1	0	0.0333	0.0333	0.0022	0.03	0.0053
6	10	0	0.0033	0.0333	0.0022	0.03	0.0053
20	0.01	0	0.5882	0.0059	0.0000	0.005294	0.0000
20	1	0	0.0059	0.0059	0.0000	0.005294	0.0000
20	10	0	0.0006	0.0059	0.0000	0.005294	0.0000
Situação 2							
d	v	$E(\mu_l)$	$Var(\mu_l)$	$E(\sigma_l^2)$	$Var(\sigma_l^2)$	$E(\sigma_{12}^2)$	$Var(\sigma_{12}^2)$
4	0.01	0	100.0000	1.0000	-	0	-
4	1	0	1.0000	1.0000	-	0	-
4	10	0	0.1000	1.0000	-	0	-
6	0.01	0	33.3333	0.3333	0.2222	0	0.0833
6	1	0	0.3333	0.3333	0.2222	0	0.0833
6	10	0	0.0333	0.3333	0.2222	0	0.0833
20	0.01	0	5.8824	0.0588	0.0005	0	0.0002
20	1	0	0.0588	0.0588	0.0005	0	0.0002
20	10	0	0.0059	0.0588	0.0005	0	0.0002
Situação 3							
d	v	$E(\mu_l)$	$Var(\mu_l)$	$E(\sigma_l^2)$	$Var(\sigma_l^2)$	$E(\sigma_{12}^2)$	$Var(\sigma_{12}^2)$
4	0.01	0	1000.0000	10.0000	-	0	-
4	1	0	10.0000	10.0000	-	0	-
4	10	0	1.0000	10.0000	-	0	-
6	0.01	0	333.3333	3.3333	22.2222	0	8.3333
6	1	0	3.3333	3.3333	22.2222	0	8.3333
6	10	0	0.3333	3.3333	22.2222	0	8.3333
20	0.01	0	58.8235	0.5882	0.0461	0	0.0218
20	1	0	0.5882	0.5882	0.0461	0	0.0218
20	10	0	0.0588	0.5882	0.0461	0	0.0218

Inferência sobre a partição, número de blocos e probabilidade de um instante ser mudança

A Tabela 3.2 mostra as probabilidades *a posteriori* das partições para as Situações 1, 2 e 3 considerando as diferentes escolhas de d e v . Na Situação 1, a escolha de $d = 20$ e $v = 10$ faz com que o modelo identifique a partição $\rho = \{0, 100\}$, que indica não ocorrência de mudança na série, com probabilidade 1. Escolhendo $d = 20$ e $v = 1$, o modelo identifica mudança no instante 49, o que se aproxima da partição real. Percebe-se também que para escolha de $v = 0.01$, a partição correta é identificada com probabilidade acima de 0.93683 para todos os três valores de d . Para outras combinações de v e d , o modelo identificou a partição correta com probabilidade 1. Percebe-se também que para Situação 1 o aumento do valor de d e a diminuição do valor de v leva o modelo à escolher a partição correta com alta probabilidade. Na Situação 2, percebe-se que houve um decréscimo da probabilidade *a posteriori* da partição correta se $d = 20$ e $v = 0.01$. Na escolha de $v = 1$, houve um crescimento da probabilidade *a posteriori* da partição correta quando d aumentou de 4 para 6. Para $d = 20$ obteve-se, *a posteriori*, probabilidade 1 para uma partição próxima da correta. A escolha de $v = 10$ fez com que o modelo identificasse a partição errada com probabilidade 1 para todos valores de d . Em relação a Situação 3, para todas as escolhas de d e v , o modelo identificou, erroneamente com probabilidade 1, a

partição que indica não ocorrência de mudança na série.

Tabela 3.2: Probabilidade *a posteriori* das partições mais prováveis.

Situação 1				
d	v	ρ	Frequência de ρ	$P(\rho \mathbf{y}_1, \dots, \mathbf{y}_n)$
4	0.01	{0, 50, 100}	599	0.9983
4	1	{0, 50, 100}	600	1
4	10	{0, 50, 100}	600	1
6	0.01	{0, 50, 100}	592	0.9867
6	1	{0, 50, 100}	600	1
6	10	{0, 50, 100}	600	1
20	0.01	{0, 50, 100}	581	0.9683
20	1	{0, 49, 100}	600	1
20	10	{0, 100}	600	1
Situação 2				
d	v	ρ	Frequência de ρ	$P(\rho \mathbf{y}_1, \dots, \mathbf{y}_n)$
4	0.01	{0, 50, 100}	600	1
4	1	{0, 50, 100}	537	0.895
4	10	{0, 100}	600	1
6	0.01	{0, 50, 100}	600	1
6	1	{0, 50, 100}	600	1
6	10	{0, 100}	599	0.9983
20	0.01	{0, 50, 100}	500	0.8333
20	1	{0, 49, 100}	600	1
20	10	{0, 100}	600	1
Situação 3				
d	v	ρ	Frequência de ρ	$P(\rho \mathbf{y}_1, \dots, \mathbf{y}_n)$
4	0.01	{0, 100}	600	1
4	1	{0, 100}	600	1
4	10	{0, 100}	600	1
6	0.01	{0, 100}	600	1
6	1	{0, 100}	600	1
6	10	{0, 100}	600	1
20	0.01	{0, 100}	600	1
20	1	{0, 100}	600	1
20	10	{0, 100}	600	1

Outra importante observação, é o fato de que as escolhas de \mathbf{D} influenciaram na identificação das partições. Na Situação 1, foi a qual obteve os melhores resultados. No entanto, vale ressaltar que mesmo com essa escolha o modelo identificou a partição errada quando houve a combinação $d = 20$ e $v = 10$. De maneira geral, nota-se se o valor \mathbf{D} for distante da matriz de covariância utilizada na geração das observações, o modelo tende a identificar partições do tipo $\rho = \{0, 100\}$ com alta probabilidade.

A Figura 3.2 mostra as probabilidades *a posteriori* para o número de blocos, considerando as distribuições *a priori* da Tabela 3.1. Na Situação 1, percebe-se que para $d \leq 20$ e $v \leq 1$ o modelo conseguiu identificar dois blocos com probabilidade alta. No entanto, considerando $d = 20$ e $v = 10$ o modelo identificou apenas 1 bloco com probabilidade 1. Na Situação 2, percebe-se que se $v = 10$, o modelo identifica apenas um bloco para qualquer que seja a escolha de d . Para as demais combinações, foi identificado dois blocos com alta probabilidade. Na Situação 3, todas as combinações de d e v levaram a identificação *a posteriori* de apenas um bloco.

A Tabela 3.3 apresenta as estimativas *a posteriori* para p . Em geral a média *a posteriori* de p é inferior ao que assumido *a priori* ($E(p) = 0.01$) e nos cenários onde isto não é observado, a média *a posteriori* é bem próxima de 0.01. Note que nas partições que indicam presença de dois blocos, as estimativas *a posteriori*, para a probabilidade de haver mudança em um instante qualquer, estão em torno de 0.01

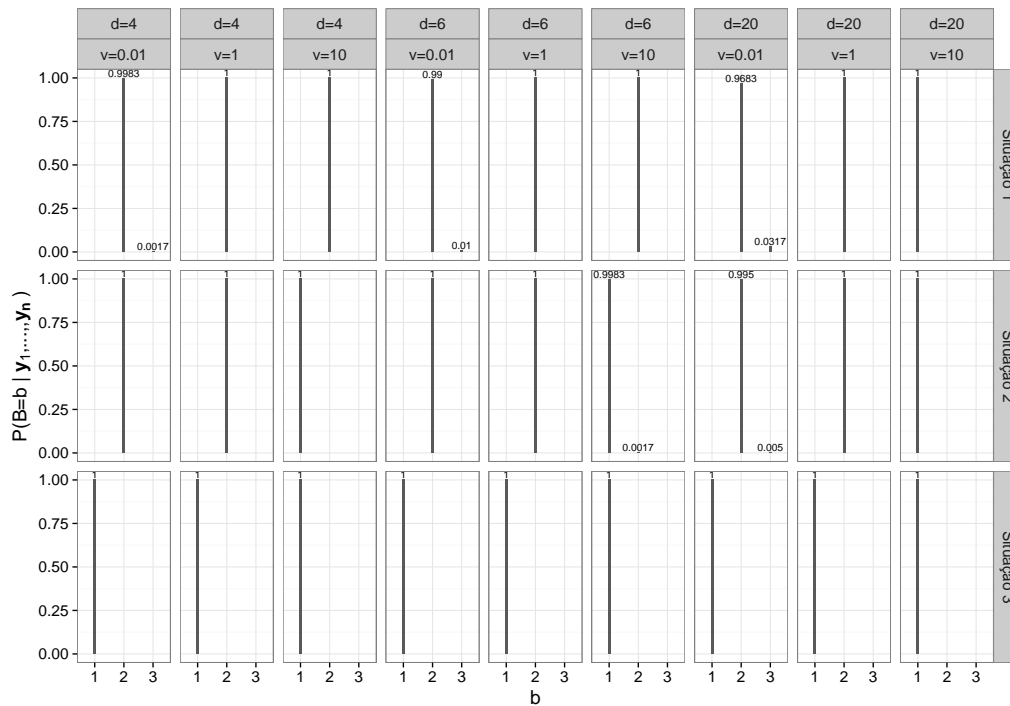


Figura 3.2: Probabilidade *a posteriori* para o número de blocos.

(média) ou 0.0092 (mediana), enquanto para aquelas que identificaram apenas um bloco as estimativas estão próximas de 0.0066 (média) ou 0.0054 (mediana). Na

Tabela 3.3: Estimativas *a posteriori* para a probabilidade p de ponto de mudança.

Situação 1				
d	v	Média	Mediana	Desvio padrão
4	0.01	0.0105	0.0092	0.0061
4	1	0.0104	0.0092	0.0060
4	10	0.0104	0.0092	0.0060
6	0.01	0.0102	0.0090	0.0063
6	1	0.0099	0.0089	0.0058
6	10	0.0099	0.0089	0.0058
20	0.01	0.0101	0.0089	0.0058
20	1	0.0098	0.0087	0.0058
20	10	0.0066	0.0058	0.0047
Situação 2				
d	v	Média	Mediana	Desvio padrão
4	0.01	0.0103	0.0092	0.0059
4	1	0.0103	0.0092	0.0057
4	10	0.0064	0.0054	0.0046
6	0.01	0.0099	0.0089	0.0058
6	1	0.0099	0.0089	0.0058
6	10	0.0066	0.0054	0.0046
20	0.01	0.0100	0.0089	0.0057
20	1	0.0099	0.0088	0.0059
20	10	0.0066	0.0058	0.0047
Situação 3				
d	v	Média	Mediana	Desvio padrão
4	0.01	0.0064	0.0054	0.0046
4	1	0.0064	0.0054	0.0046
4	10	0.0064	0.0054	0.0046
6	0.01	0.0066	0.0054	0.0046
6	1	0.0066	0.0054	0.0046
6	10	0.0066	0.0054	0.0046
20	0.01	0.0066	0.0058	0.0047
20	1	0.0066	0.0058	0.0047
20	10	0.0066	0.0058	0.0047

Figura D.5 (Apêndice D) são apresentadas as distribuições *a posteriori* da probabilidade de mudança p para as três situações.

A Figura 3.3 apresenta a probabilidade *a posteriori* de cada instante ser um ponto de mudança para as Situações 1 e 2. Na Situação 3, a probabilidade *a posteriori* de cada instante ser um ponto de mudança é zero em todos os instantes. Este comportamento também é observado na Situação 1 quando $v = 10$ e $d = 20$ e na Situação 2 quando $v = 10$. Na Situação 1, o instante 51 assume alta probabilidade ser mudança para todas escolhas de d e v , excetuando-se quando $v = 1$ e $d = 20$, onde identifica-se o instante 50 com alta probabilidade. Na Situação 2, a escolha de $v = 1$ e $d = 20$ também identificou o instante 50 com alta probabilidade de mudança. Para as demais combinações, o modelo identificou o instante 51 com alta probabilidade *a posteriori* de ser ponto de mudança.

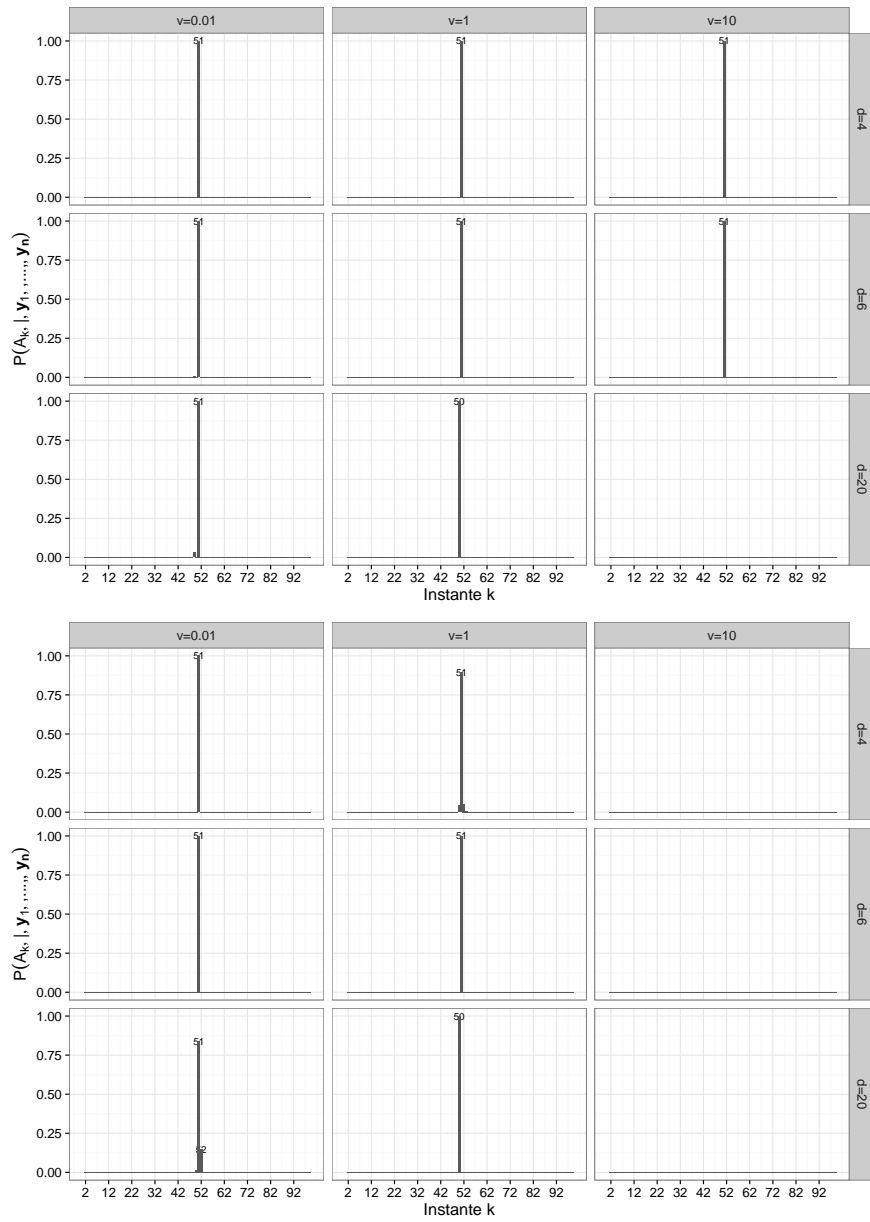


Figura 3.3: Probabilidade *a posteriori* de cada instante ser um ponto de mudança para as Situações 1 (topo) e 2 (base).

Análise de μ e Σ

A Figura 3.4 mostra as estimativas (média, mediana e HPD *a posteriori*) de μ_k para série 1 nas Situações 1, 2 e 3. Percebe-se que na Situação 1, para escolha de $v = 10$ houve uma subestimação, inclusive para as escolhas de d no qual o modelo identificou a partição correta com probabilidade 1. Para $v = 1$, note que para os instantes de 1 a 50 houve uma superestimação enquanto para 51 a 100 houve uma subestimação. Para $v = 0.01$, a estimativa apresentou o mesmo comportamento de $v = 1$. Para $v = 1$, apesar de haver uma superestimação, as estimativas estão mais próximas dos valores reais para os instantes de 1 a 50, quando comparados com as obtidas para os instantes 51 a 100. No entanto, para $v = 0.01$, apesar da subestimação, as estimativas estão mais próximas dos valores reais para os instantes de 51 a 100. O mesmo comportamento é identificado para a Situação 2. Percebe-se que, na Situação 3, as estimativas ficaram bem próximas dos valores reais para as escolhas de $v = 0.01$ e $v = 1$, porém neste mesmo caso o modelo identificou a partição errada dos dados.

Em relação a diferença entre as médias em tempos consecutivos da série 1, não houveram diferenças significativas em nenhuma das três situações, por isso os gráficos foram omitidos.

A Figura 3.5 mostra as estimativas de μ_k para série 2. Nota-se que para a situação 1, houve também uma subestimação das médias em cada instante quando $v = 10$, inclusive nos valores de d no qual o modelo identificou a partição correta com o probabilidade 1. Isso ocorre devido ao fato de que a escolha de $v = 10$ torna distribuição *a priori* de μ muita informativa. Observa-se que para as demais escolhas de v , o modelo conseguiu fornecer valores bem próximo dos reais. A Situação 2 apresentou um comportamento parecido com a Situação 1. Em relação a Situação 3, também observa-se o efeito da escolha de v na estimativa dos parâmetros.

Na Figura 3.6 podem serem observados os intervalos HPD com 95% de probabilidade para a diferença entre as médias em instantes consecutivos para série 2. Para a maioria dos casos, detectou-se que a média no instante 51 difere significativamente da média no instante 50, indicando haver uma mudança neste parâmetro ao longo do tempo. Para os casos das Situações 1 e 2, se $d = 20$ e $v = 1$ a média do instante 49 diferiu significativamente do instante 48. Observa-se que para $d = 4$, $v = 1$ e Situação 2, a média do instante 51 não diferenciou-se significavelmente da média do instante 50, apesar de o modelo ter identifica a partição correta dos dados com probabilidade de 0.895. O mesmo acontece quando assumido $v = 0.01$ e $d = 20$. Como na situação 3 o modelo identificou apenas partições do tipo $\rho = \{0, 100\}$ com probabilidade 1 para todas as escolhas de d e v , logo não existe ponto de mudança para essa situação, por isso o gráfico foi omitido. Na Figura D.2 do Apêndice D é

apresentada a distribuição da diferença entre $\mu_{51} - \mu_{50}$.

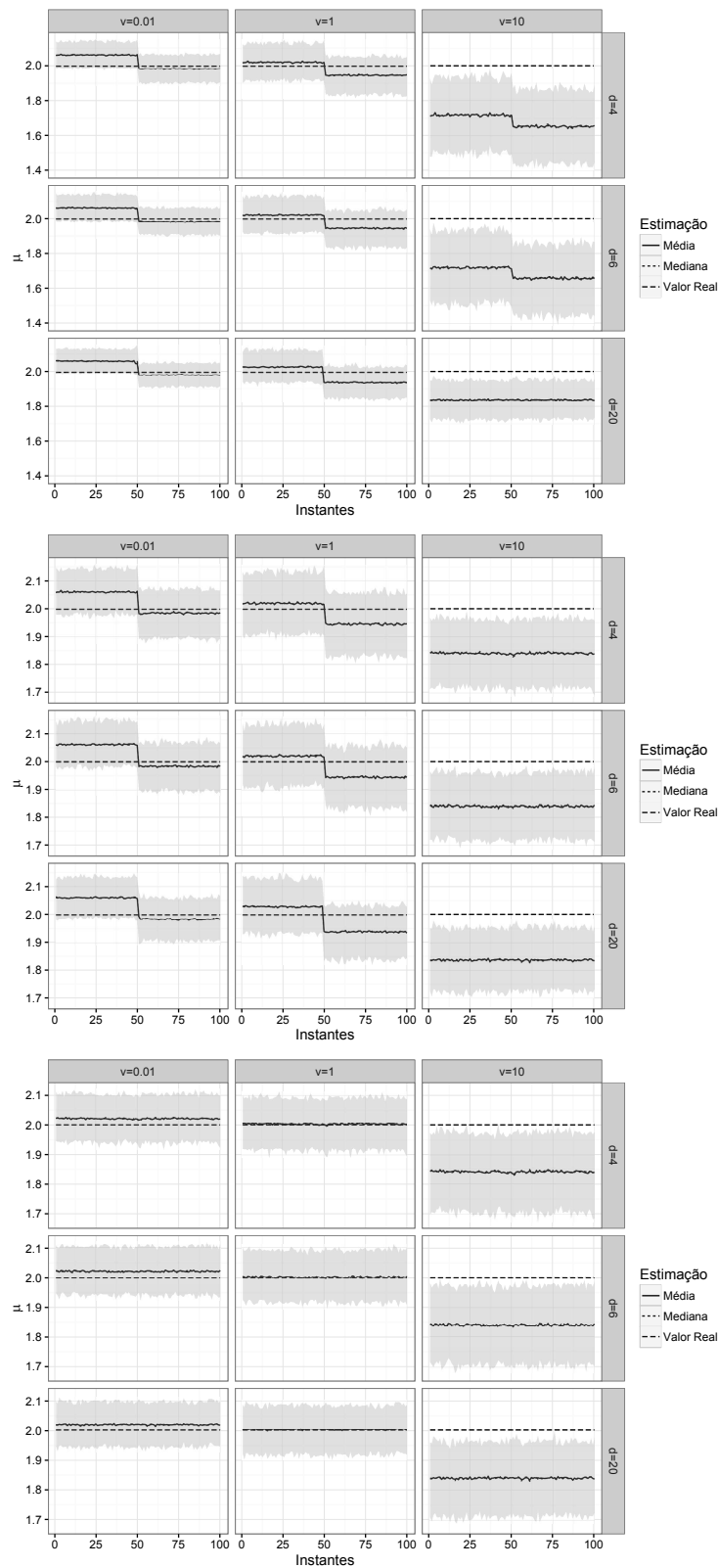


Figura 3.4: Estimativas *a posteriori* do parâmetro μ_k da série 1 para as Situações 1 (topo), 2 (meio) e 3 (base). A parte sombreada corresponde ao intervalo HPD com 95%.

A Figura 3.7 mostra a estimativa de σ_k^2 para série 1. Note que para $v = 10$ houve uma superestimação da variância nas três situações, inclusive nos casos em que a média *a priori* da variância da série 1 está subestimando ou era igual ao valor real. Percebe-se também que a medida que o valor de v diminui, as estimativas se aproximam dos valores reais. Tal efeito ocorre devido ao fato de que o valor de v está relacionado ao vetor de médias amostrais presente no parâmetro \mathbf{D}^* da distribuição *a posteriori* de Σ_k . Dessa forma, o aumento de v faz com que o vetor de médias amostrais tenham maior influência na estimativa da variância. Observa-se que a medida que o valor de d aumenta, as estimativas ficam mais próximas dos valores reais.

A partir do intervalo HPD da diferença entre os instantes consecutivos, percebe-se que não existem diferenças significativas entre σ_k^2 e σ_{k-1}^2 , evidenciando que a variância não se modifica ao longo do tempo (gráfico omitido).

A Figura 3.8 apresenta as estimativas de σ_k^2 da série 2. Percebe-se, novamente, o efeito proveniente das médias amostrais ponderada pela a escolha de v . Em relação $\sigma_k^2 - \sigma_{k-1}^2$, não houveram diferenças significativas entre os instantes consecutivos (gráfico omitido), inclusive na Situação 1 onde $v = 10$ e $d = 4$ ou $v = 10$ e $d = 6$.

As estimativas para a covariância podem ser observadas na Figura 3.9. Também pode-se notar o efeito da escolha do v na estimativa. Para $\sigma_{(12)k}^2 - \sigma_{(12)k-1}^2$ também não houve mudança significativa, indicando que a correlação não muda ao longo do tempo (gráfico omitido).

De forma geral, percebe-se que os modelos com as especificações $\mathbf{D} = \begin{pmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{pmatrix}$, $v = 0.01$, $d = 4$ e $\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $v = 0.01$ e $d = 4$ foram os que apresentaram melhores resultados conjuntamente para os parâmetros.

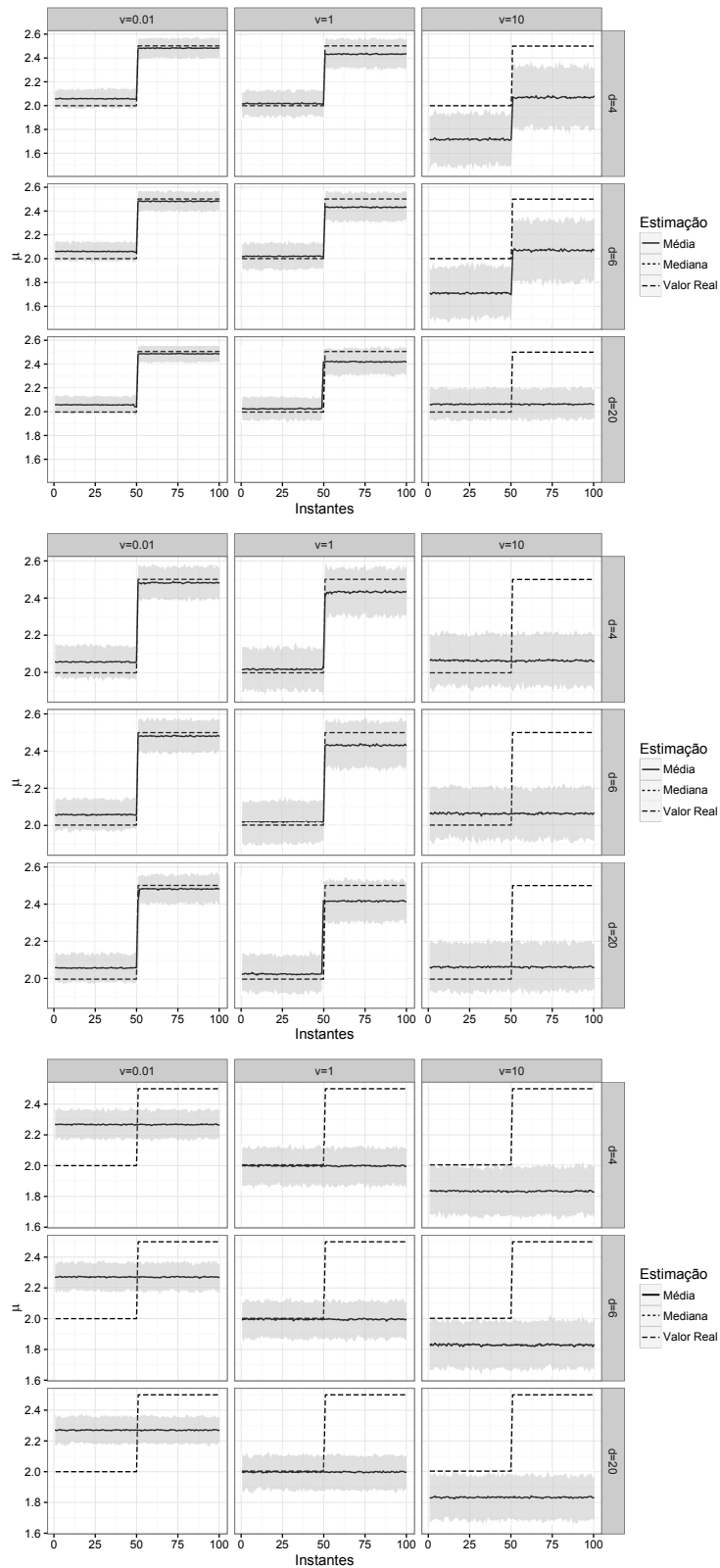


Figura 3.5: Estimativas *a posteriori* do parâmetro μ_k da série 2 para as Situações 1 (topo), 2 (meio) e 3 (base). A parte sombreada corresponde ao intervalo HPD com 95%.

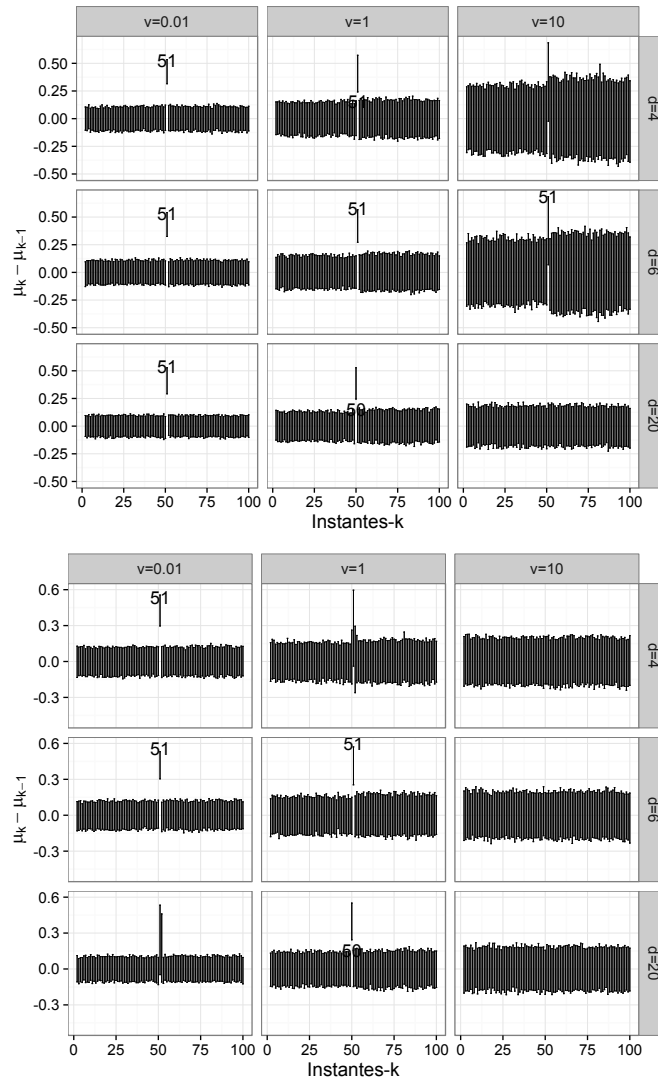


Figura 3.6: Intervalo HPD de 95% para $\mu_k - \mu_{k-1}$ da série 2, Situações 1 (topo) 2 (base).

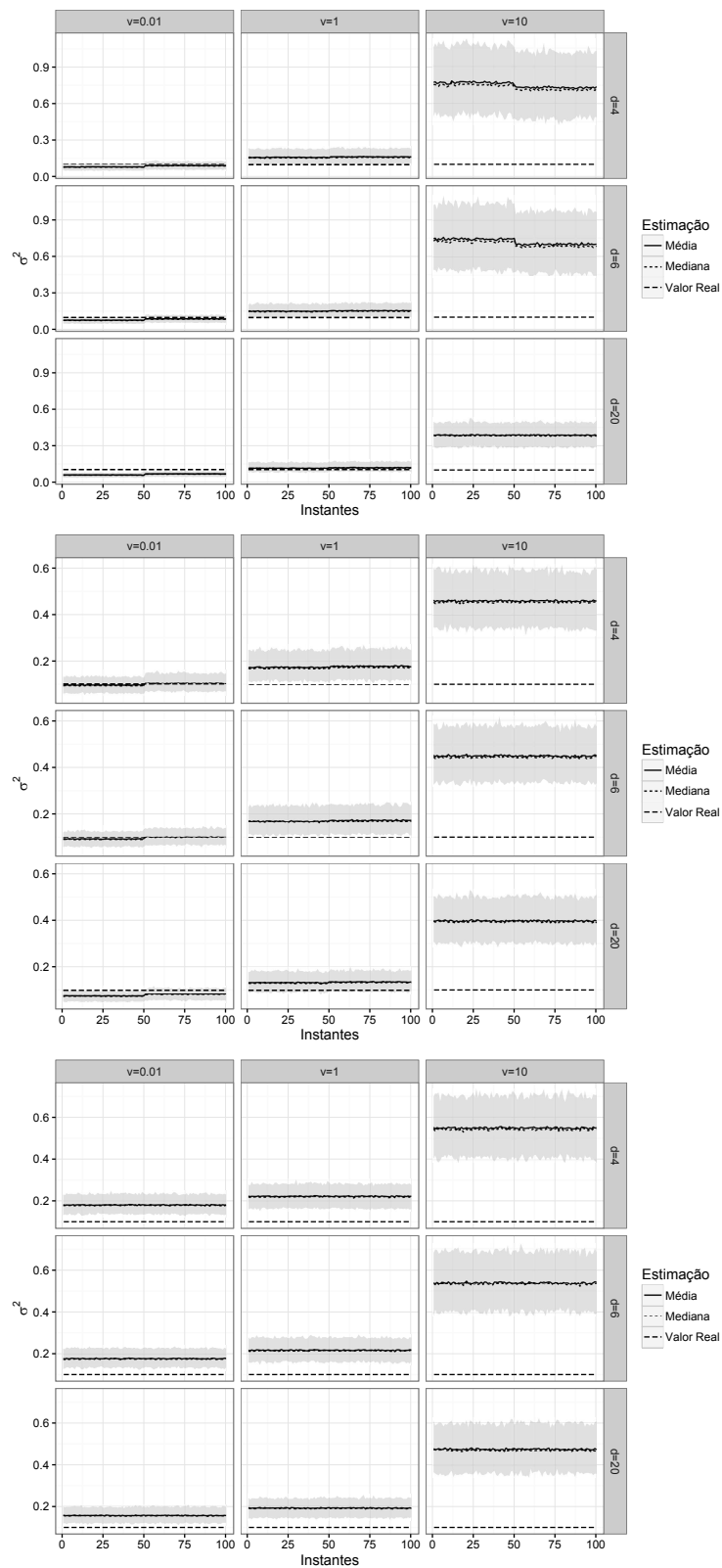


Figura 3.7: Estimativas do parâmetro σ_k^2 da série 1 para as Situações 1 (topo), 2 (meio) e 3 (base). A parte sombreada corresponde ao intervalo HPD de 95%.

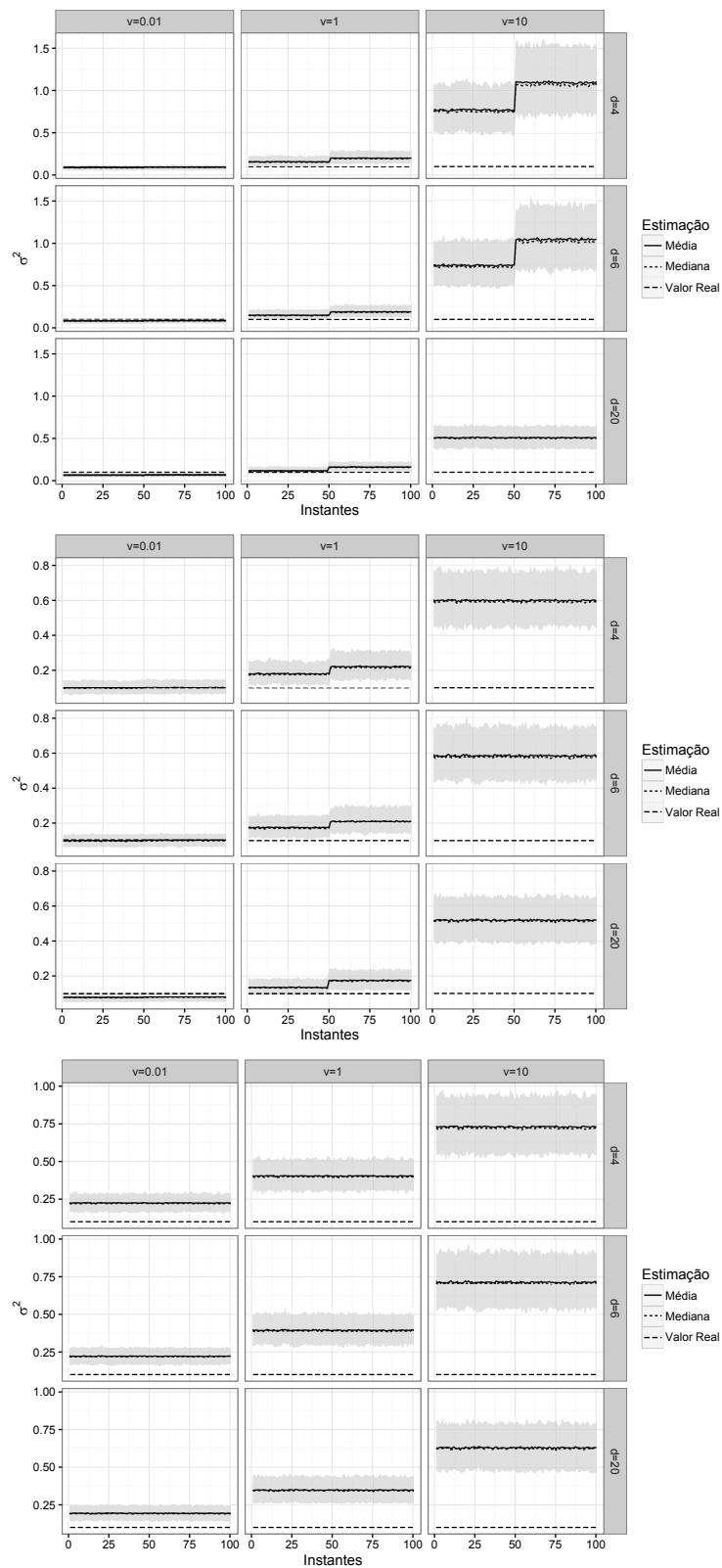


Figura 3.8: Estimativas do parâmetro σ_k^2 da série 2 para as Situações 1 (topo), 2 (meio) e 3 (base). A parte sombreada corresponde ao intervalo HPD de 95%.

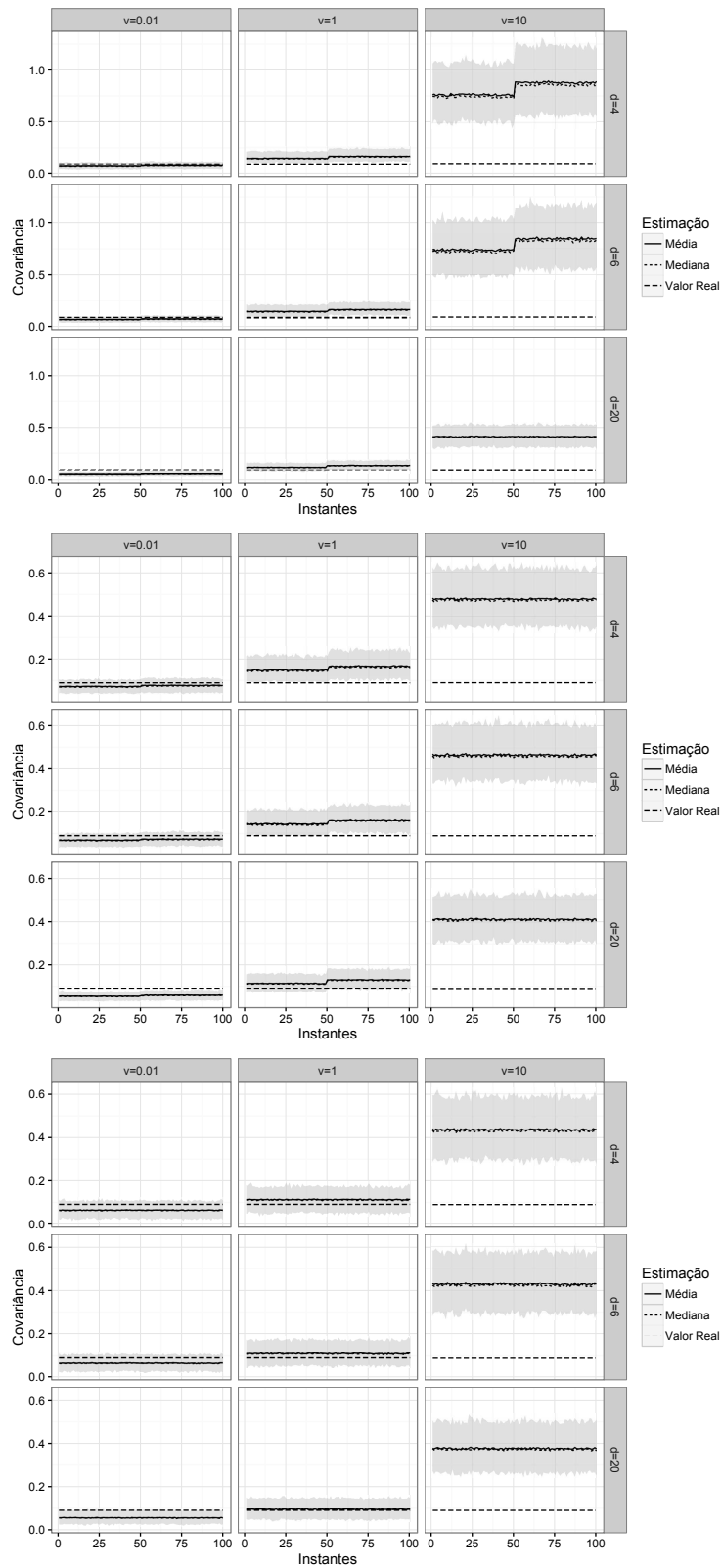


Figura 3.9: Estimativas do parâmetro $\sigma_{12(k)}^2$ para as Situações 1 (topo), 2 (meio) e 3 (base). A parte sombreada corresponde ao intervalo HPD de 95%.

3.2 Cenário 2: mudança na variância

Neste cenário, tem-se como objetivo avaliar o desempenho do MMPNM sobre diferentes óticas quando há uma pequena mudança na variância de apenas uma das séries e quando há uma alta correlação entre ambas. Para isso, uma série bivariada de tamanho 100 foi simulada considerando um ponto de mudança no instante 51 da seguinte forma:

$$\mathbf{Y}_k \sim \begin{cases} N_2 \left[\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{pmatrix} \right] & \text{para } k = 1, \dots, 50, \\ N_2 \left[\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 0.1 & 0.09 \\ 0.09 & 0.5 \end{pmatrix} \right] & \text{para } k = 51, \dots, 100. \end{cases}$$

Os dados simulados podem ser vistos na Figura 3.10.

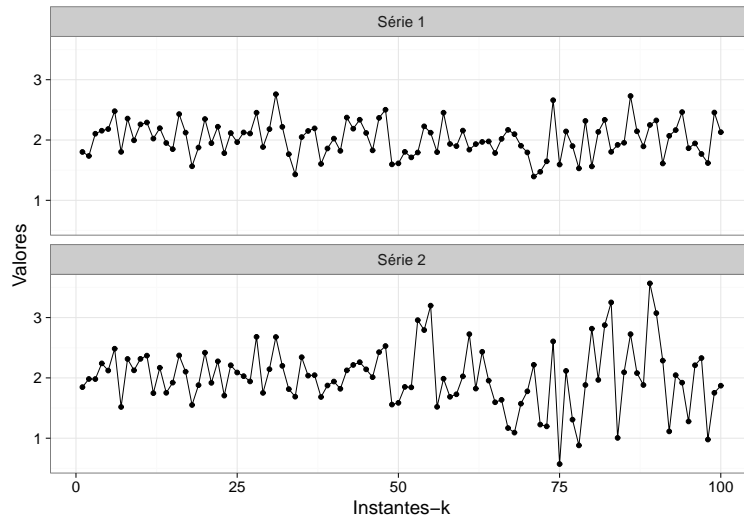


Figura 3.10: Valores das séries para o cenário 1.

Para avaliar o efeito da escolha do parâmetro \mathbf{D} que indexam a distribuição *a priori* de $\Sigma_{[ij]}$, três situações são consideradas. Para a Situação 1, define-se $\mathbf{D} = \begin{pmatrix} 0.1 & 0.09 \\ 0.09 & 0.5 \end{pmatrix}$. As escolhas de \mathbf{D} para Situações 2 e 3 são as mesmas apresentadas na Seção 3.1

O efeito da escolha de \mathbf{D} , v e d nas estimativas *a priori* parâmetros $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, quando a situação 1 é considerada, pode ser vista na Tabela 3.4.

Tabela 3.4: Estimativas *a priori*, $l = 1, 2$.

d	v	$E(\mu_l)$	$Var(\mu_1)$	$Var(\mu_2)$	$E(\sigma_1^2)$	$E(\sigma_{12}^2)$	$E(\sigma_2^2)$	$Var(\sigma_1^2)$	$Var(\sigma_2^2)$	$Var(\sigma_{12}^2)$
4	0.01	0	10.0000	50.0000	0.1000	0.0900	0.5000	-	-	-
4	1	0	0.1000	0.5000	0.1000	0.0900	0.5000	-	-	-
4	10	0	0.0100	0.0500	0.1000	0.0900	0.5000	-	-	-
6	0.01	0	3.3333	16.6667	0.0333	0.0300	0.1667	2.2222e-03	5.5556e-02	5.2917e-03
6	1	0	0.0333	0.1667	0.0333	0.0300	0.1667	2.2222e-03	5.5556e-02	5.2917e-03
6	10	0	0.0033	0.0167	0.0333	0.0300	0.1667	2.2222e-03	5.5556e-02	5.2917e-03
20	0.01	0	0.5882	2.9412	0.0059	0.0053	0.0294	4.6136e-06	1.1534e-04	1.2866e-05
20	1	0	0.0059	0.0294	0.0059	0.0053	0.0294	4.6136e-06	1.1534e-04	1.2866e-05
20	10	0	0.0006	0.0029	0.0059	0.0053	0.0294	4.6136e-06	1.1534e-04	1.2866e-05

Inferência sobre a partição, número de blocos e probabilidade de um instante ser mudança

Da Tabela 3.5, pode ser observada as probabilidades *a posteriori* das partições para as Situações 1, 2 e 3. Observa-se que tanto na Situação 1 quanto na Situação 2, a partição $\rho = \{0, 52, 100\}$ foi a mais identificada pelo modelo. Para a Situação 1, há uma diminuição da probabilidade a medida que d cresce e o valor de v é fixado em 0.01. Nota-se que para o valor de v fixo em 1 e o crescimento do valor de d , o modelo passou a identificar a partição $\rho = \{0, 100\}$. O mesmo comportamento ocorre para $v = 10$ em relação a Situação 2. Note que para Situação 3, o modelo identificou a partição $\rho = \{0, 100\}$ com probabilidade 1 para todas as combinações de v e d . Nota-se também que a medida que o valor de \mathbf{D} fica mais distante da matriz de covariância que gerou os dados, o modelo tende a identificar com probabilidade 1 a partição do tipo $\rho = \{0, 100\}$.

Na Figura 3.11 é apresentada a probabilidade *a posteriori* do número de blocos. Nota-se que para a Situação 1, o modelo identificou com alta probabilidade dois blocos, quantidade correta, para todas as combinações de v e d , exceto $d = 20$ e $v = 1$ ou $d = 20$ e $v = 10$. Em relação a Situação 2, nota-se que para o valor de $v = 0.01$, o modelo tende a identificar dois blocos com alta probabilidade. Na Situação 3, o modelo identificou um bloco com probabilidade 1 para todas as combinações de v e d .

A Tabela 3.6 apresenta as estimativas *a posteriori* da probabilidade de mudança p . Note que a probabilidade de mudança para as partições que obtiveram dois blocos estão em torno de 0.01 (média) ou 0.0092 (mediana), enquanto para aquelas que identificaram apenas um bloco as estimativas estão próximas de 0.0066 (média) ou 0.0054 (mediana). Na Figura D.9 (Apêndice D) são apresentadas as distribuições *a posteriori* da probabilidade de mudança p para as três situações.

Na Figura 3.12 é apresentada a probabilidade *a posteriori* de cada instante ser um ponto de mudança para as situações 1 e 2, pois para a Situação 3 a probabilidade

Tabela 3.5: Probabilidade a posteriori das partições mais prováveis.

Situação 1				
d	v	ρ	Frequência de ρ	$P(\rho \mathbf{y}_1, \dots, \mathbf{y}_n)$
4	0.01	{0, 52, 100}	464	0.7733
4	1	{0, 52, 100}	600	1
4	10	{0, 52, 100}	429	0.715
6	0.01	{0, 51, 100}	302	0.5033
6	1	{0, 52, 100}	600	1
6	10	{0, 100}	260	0.4333
20	0.01	{0, 52, 100}	322	0.5367
20	1	{0, 100}	590	0.9833
20	10	{0, 100}	600	1
Situação 2				
d	v	ρ	Frequência de ρ	$P(\rho \mathbf{y}_1, \dots, \mathbf{y}_n)$
4	0.01	{0, 52, 100}	599	0.9983
4	1	{0, 52, 100}	330	0.55
4	10	{0, 100}	600	1
6	0.01	{0, 52, 100}	599	0.9983
6	1	{0, 52, 100}	388	0.6467
6	10	{0, 100}	600	1
20	0.01	{0, 51, 100}	218	0.3633
20	1	{0, 100}	390	0.65
20	10	{0, 100}	600	1
Situação 3				
d	v	ρ	Frequência de ρ	$P(\rho \mathbf{y}_1, \dots, \mathbf{y}_n)$
4	0.01	{0, 100}	600	1
4	1	{0, 100}	600	1
4	10	{0, 100}	600	1
6	0.01	{0, 100}	600	1
6	1	{0, 100}	600	1
6	10	{0, 100}	600	1
20	0.01	{0, 100}	600	1
20	1	{0, 100}	600	1
20	10	{0, 100}	600	1

de cada instante ser um ponto de mudança é zero. Note que na Situação 1, o instante 53 assume alta probabilidade para algumas escolhas de d e v , em outras combinações o instante 52 também apresenta alta probabilidade. Para as demais combinações, todos os instantes assumem probabilidades muito pequenas ou zero. A Situação 2, apresenta um comportamento parecido com a Situação 1, assumindo alta probabilidade para o instante 53 e para algumas combinações de v e d há um confundimento com outros instantes.

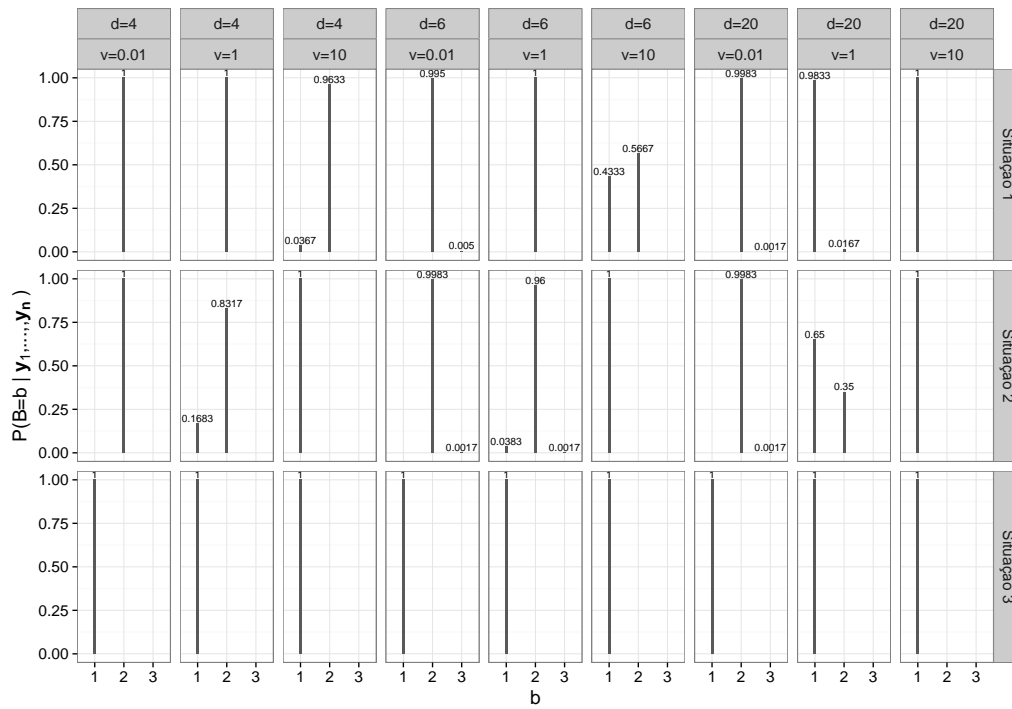


Figura 3.11: Probabilidade *a posteriori* para o número de blocos.

Tabela 3.6: Estimativas para a probabilidade p de ponto de mudança

Situação 1				
d	v	Média	Mediana	Desvio padrão
4	0.01	0.0101	0.0089	0.0058
4	1	0.0103	0.0091	0.0060
4	10	0.0106	0.0093	0.0062
6	0.01	0.0102	0.0089	0.0061
6	1	0.0099	0.0088	0.0059
6	10	0.0086	0.0078	0.0054
20	0.01	0.0102	0.0089	0.0061
20	1	0.0067	0.0058	0.0046
20	10	0.0066	0.0058	0.0047
Situação 2				
d	v	Média	Mediana	Desvio padrão
4	0.01	0.0103	0.0090	0.0060
4	1	0.0092	0.0082	0.0057
4	10	0.0064	0.0054	0.0046
6	0.01	0.0098	0.0088	0.0057
6	1	0.0098	0.0089	0.0058
6	10	0.0066	0.0054	0.0046
20	0.01	0.0099	0.0088	0.0059
20	1	0.0080	0.0066	0.0055
20	10	0.0066	0.0058	0.0047
Situação 3				
d	v	Média	Mediana	Desvio padrão
4	0.01	0.0064	0.0054	0.0046
4	1	0.0064	0.0054	0.0046
4	10	0.0064	0.0054	0.0046
6	0.01	0.0066	0.0054	0.0046
6	1	0.0066	0.0054	0.0046
6	10	0.0066	0.0054	0.0046
20	0.01	0.0066	0.0058	0.0047
20	1	0.0066	0.0058	0.0047
20	10	0.0066	0.0058	0.0047

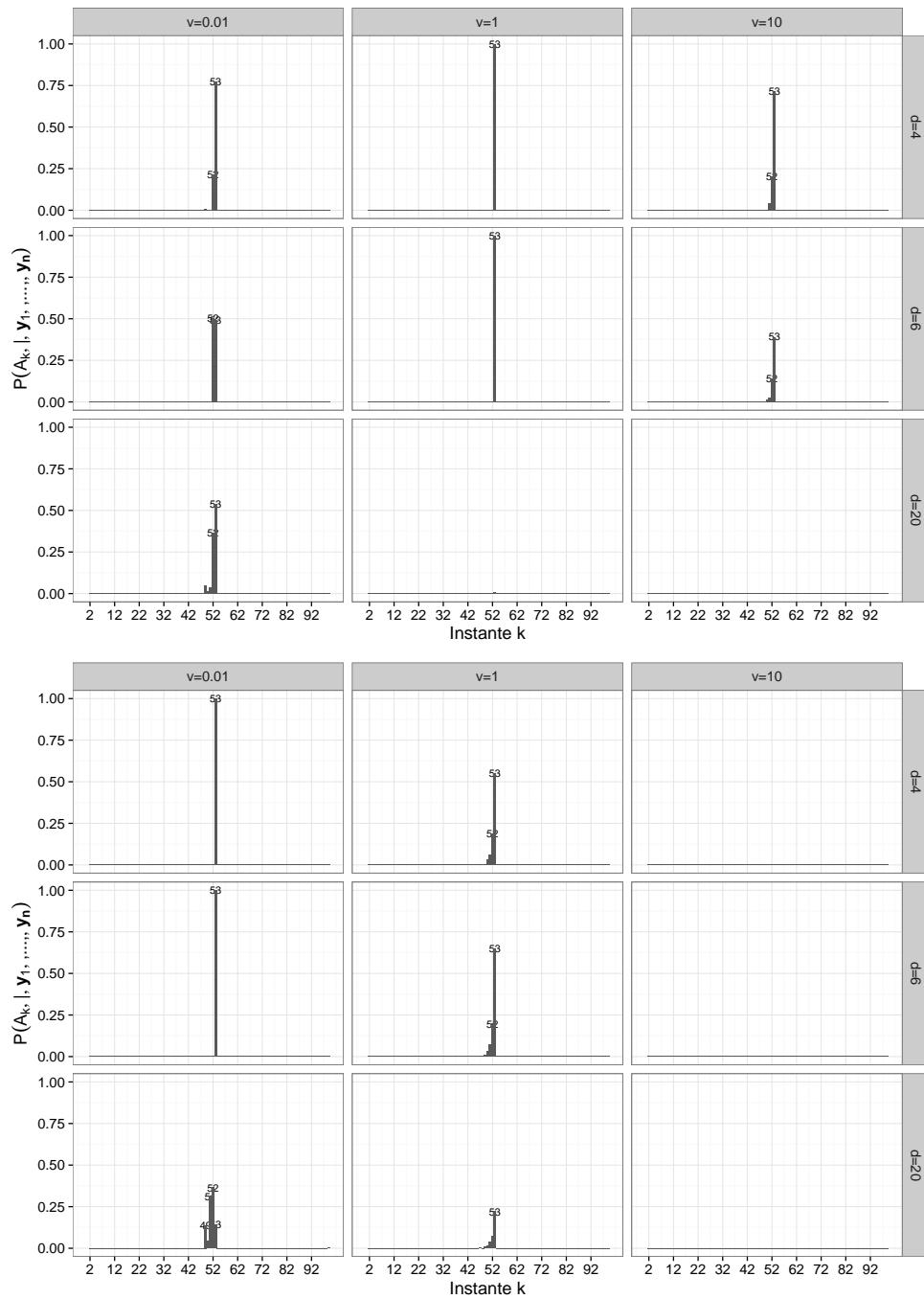


Figura 3.12: Probabilidade *a posteriori* de cada instante ser um ponto de mudança para as Situações 1 (topo) e 2 (base). Os números acima da barra representam os instantes.

Análise de μ_k e Σ_k

A Figura 3.13 mostra as estimativas (média, mediana e HPD de 95%) de μ_k da série 1. Nota-se que para Situações 1, 2 e 3, a escolha de $v = 10$ fez com que houvesse uma subestimação do modelo, isso ocorre devido ao fato que tal escolha de v torna a distribuição *a priori* muito informativa. Na Situação 1, percebe-se que se $v = 1$ e $d = 4$ o modelo conseguiu estimar bem próximo dos valores reais, lembrando que para tal escolha a $P(\rho = \{0, 52, 100\} \mid \mathbf{y}_1, \dots, \mathbf{y}_n) = 1$. Quando $v = 1$ e $d = 20$ tem-se estimativas muito perto do valor real, no entanto para tal combinação a $P(\rho = \{0, 100\} \mid \mathbf{y}_1, \dots, \mathbf{y}_n) = 1$. Na Situação 2, percebe-se que se $v = 1$ as estimativas ficaram muito próximas do valor real. No entanto, quando $v = 1$ e $d = 20$ o modelo identifica a partição $\rho = \{0, 100\}$ com probabilidade 1. Na escolha de $v = 0.01$, nota-se também que as estimativas estão bem próximas do valor real. Para essa escolha, o modelo identificou a partição $\rho = \{0, 52, 100\}$ para todos os valores de d . Em relação a Situação 3, as estimativas ficaram próximas do valor real, porém para essa Situação a partição identificada pelo modelo foi de $\rho = \{0, 100\}$. Na Figura D.6, Apêndice D, pode ser observado o gráfico de μ_k para duas combinações de v e d .

Na Figura 3.14 são apresentadas as estimativas de μ_k para a série 2 nas Situações 1, 2, e 3. Novamente, a Situação 3 apresenta estimativas próximas dos valores reais para $v \neq 10$, no entanto o modelo não identifica a partição correta. Para as Situações 1 e 2 as estimativas estão mais próximas dos valores reais. Também é possível notar o efeito da escolha de v nos valores estimados. Não houveram diferenças significativas entre $\mu_k - \mu_{k-1}$, por isso os gráficos foram omitidos.

Na Figura 3.15 são apresentadas as estimativas de σ_k^2 da série 1. Nota-se mais uma vez o efeito do vetor de médias em \mathbf{D}^* ponderada pelo valor de v . Percebe-se que as estimativas foram ficando mais próximas dos valores reais a medida que v decresce nas três situações. Percebe-se também que mesmo nas situações onde o modelo identificou a partição $\rho = \{0, 52, 100\}$, a estimação ficou próxima do valor real.

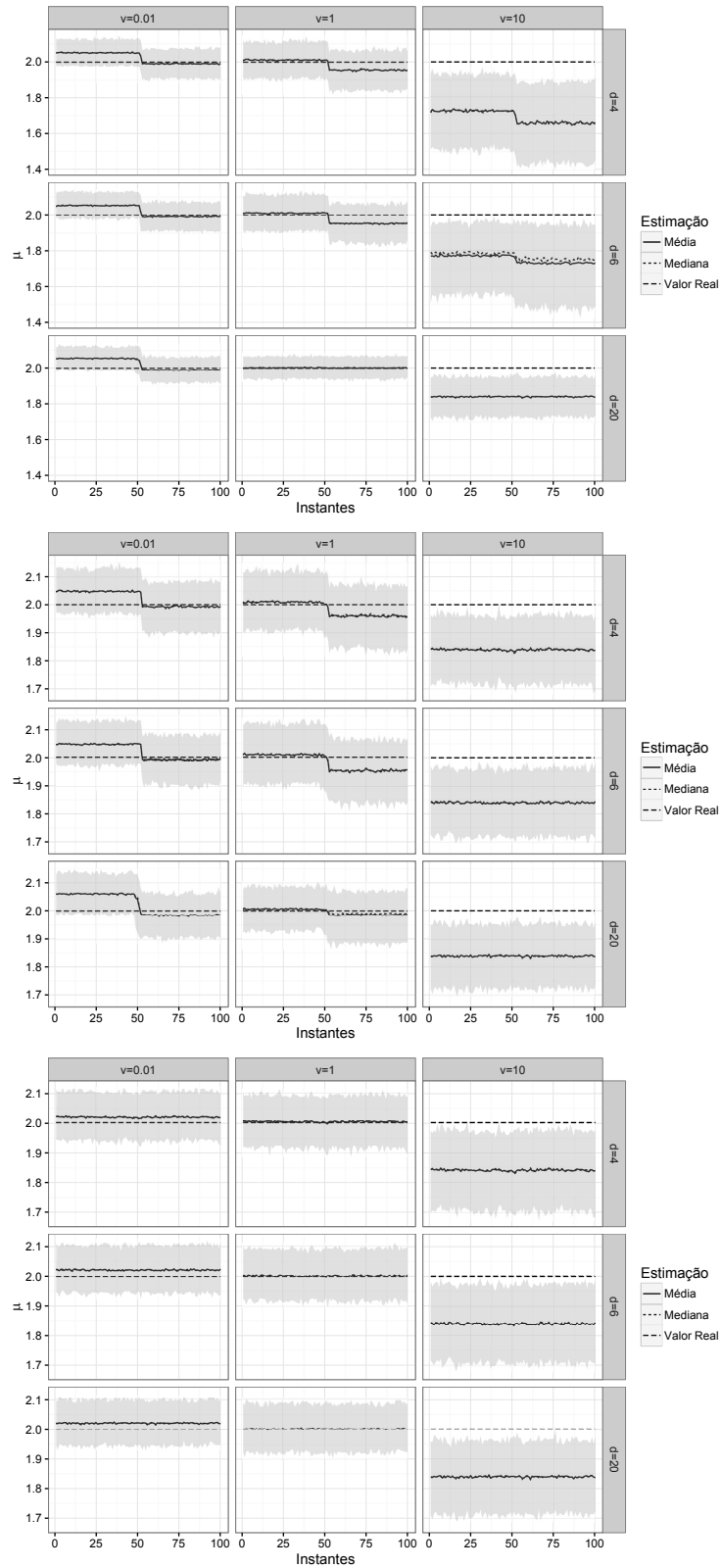


Figura 3.13: Estimativas *a posteriori* do parâmetro μ_k da série 1 para as Situações 1 (topo), 2 (meio) e 3 (base). A parte sombreada corresponde ao intervalo HPD com 95%.

A Figura 3.16 mostra as estimativas de σ_k^2 para a série 2. Na Situação 1, nota-se que se $d = 4$ o modelo estimou bem os valores exceto com a combinação $v = 10$. Para $d = 6$ houve um comportamento parecido com $d = 4$. Em relação a $d = 20$, apenas a combinação com $v = 0.01$ gerou uma estimação que identificasse a quebra. Nota-se que se $d = 20$ e $v = 1$ ou $d = 20$ e $v = 10$, o modelo identificou a partição $\rho = \{0, 100\}$ com probabilidade 1. Em relação a Situação 2, percebe-se que se $v = 0.01$ há uma subestimação a medida que d cresce. Tal acontecimento, está relacionado a fato de que a escolha de $v = 0.01$ e $d = 20$ faz com que o modelo identifique a partição $\rho = \{0, 51, 100\}$ com uma probabilidade de 0.3633. Para a escolha $v = 1$ o comportamento se repete, exceto se $v = 1$ e $d = 20$, pois o modelo identifica a partição $\rho = \{0, 100\}$ com probabilidade 0.65. Na Figura 3.17 é apresentado o intervalo com 95% de credibilidade para diferença $\sigma_k^2 - \sigma_{k-1}^2$ para a série 2 quando são consideradas as Situações 1 e 2. Nota-se que na Situação 1 se $v = 1$ e $d = 4$ ou $v = 1$ e $d = 6$ o modelo identificou uma mudança significativa na variância no instante 53. Na Situação 2, a variância apresenta mudança significativa no instante 53 se $v = 0.01$ e $d = 4$ ou $v = 0.01$ e $d = 6$. A situação 3 não mudança significativa (gráfico omitido). Para a série 1 não houve mudança significativas para as Situações 1, 2 e 3, por isso os gráficos foram omitidos.

As estimativas para covariância pode ser observada na Figura 3.18. Também pode-se notar o efeito da escolha do v na estimação. Para $\sigma_{(12)k}^2 - \sigma_{(12)k-1}^2$ não houveram mudanças significativas, indicando não haver mudanças ao longo do tempo (gráfico omitido).

De forma geral, percebe-se que o modelos com as especificações $\mathbf{D} = \begin{pmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{pmatrix}$, $v = 0.01$ e $d = 4$ foi o que apresentou um melhor resultado conjuntamente.

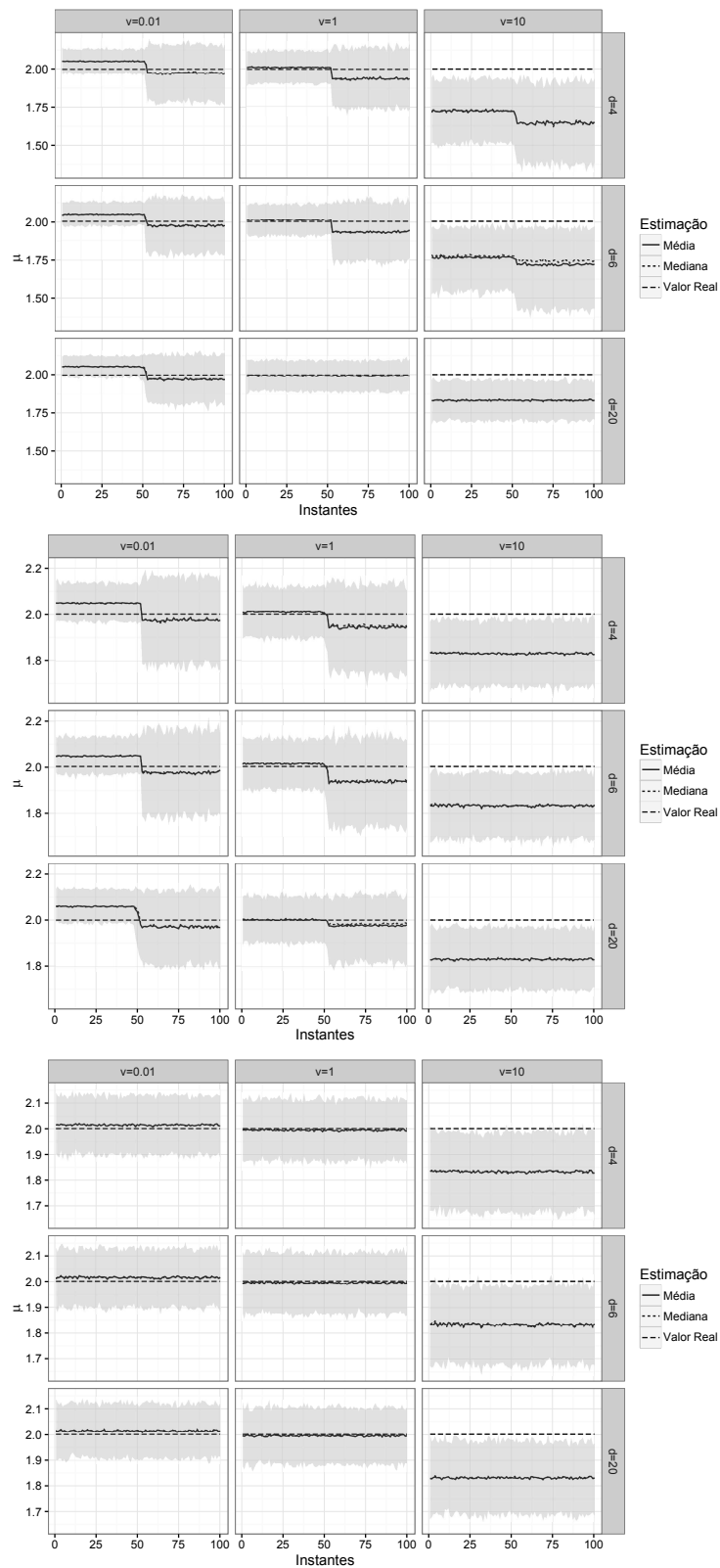


Figura 3.14: Estimativas *a posteriori* do parâmetro μ_k da série 2 para as Situações 1 (topo), 2 (meio) e 3 (base). A parte sombreada corresponde ao intervalo HPD com 95%.

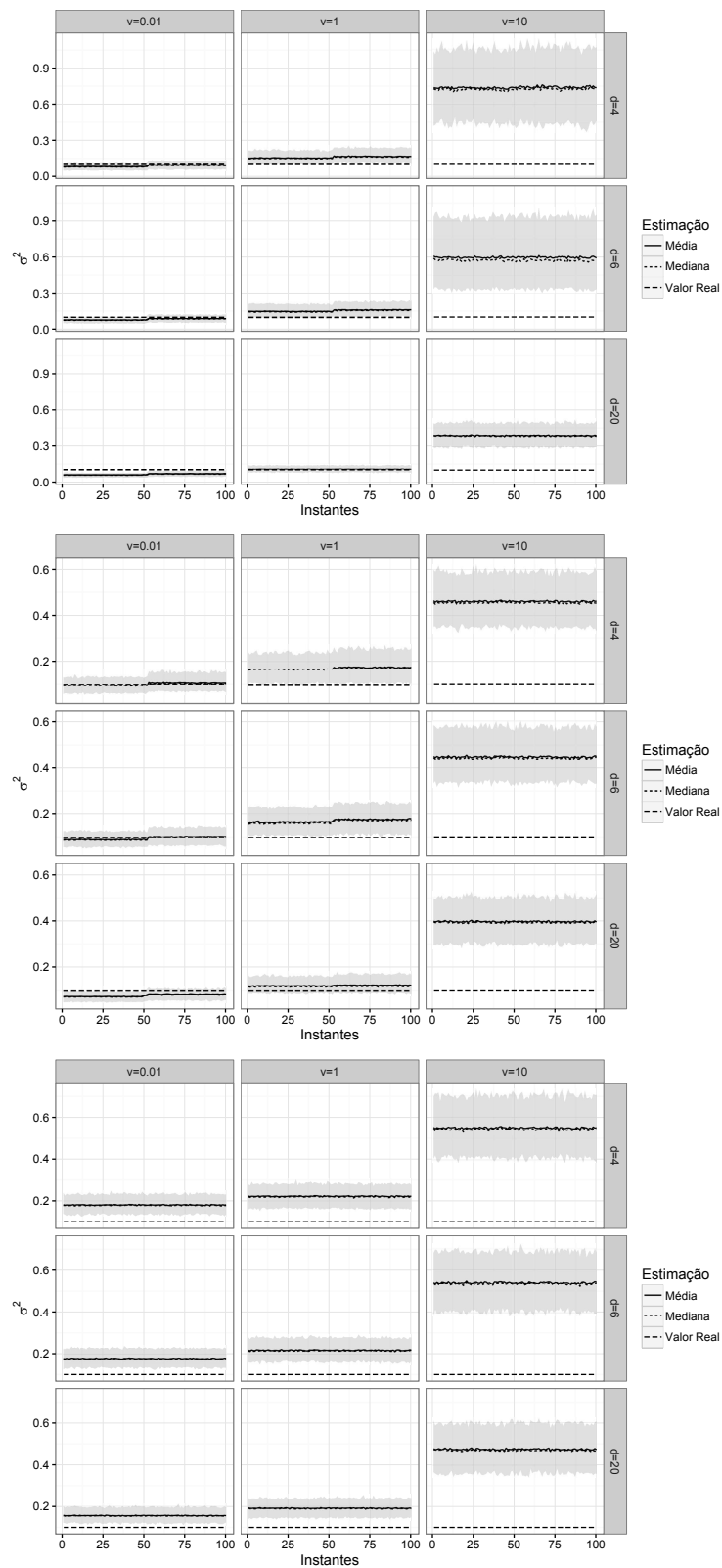


Figura 3.15: Estimativas do parâmetro σ_k^2 da série 1 para as Situações 1 (topo), 2 (meio) e 3 (base). A parte sombreada corresponde ao intervalo HPD de 95%.

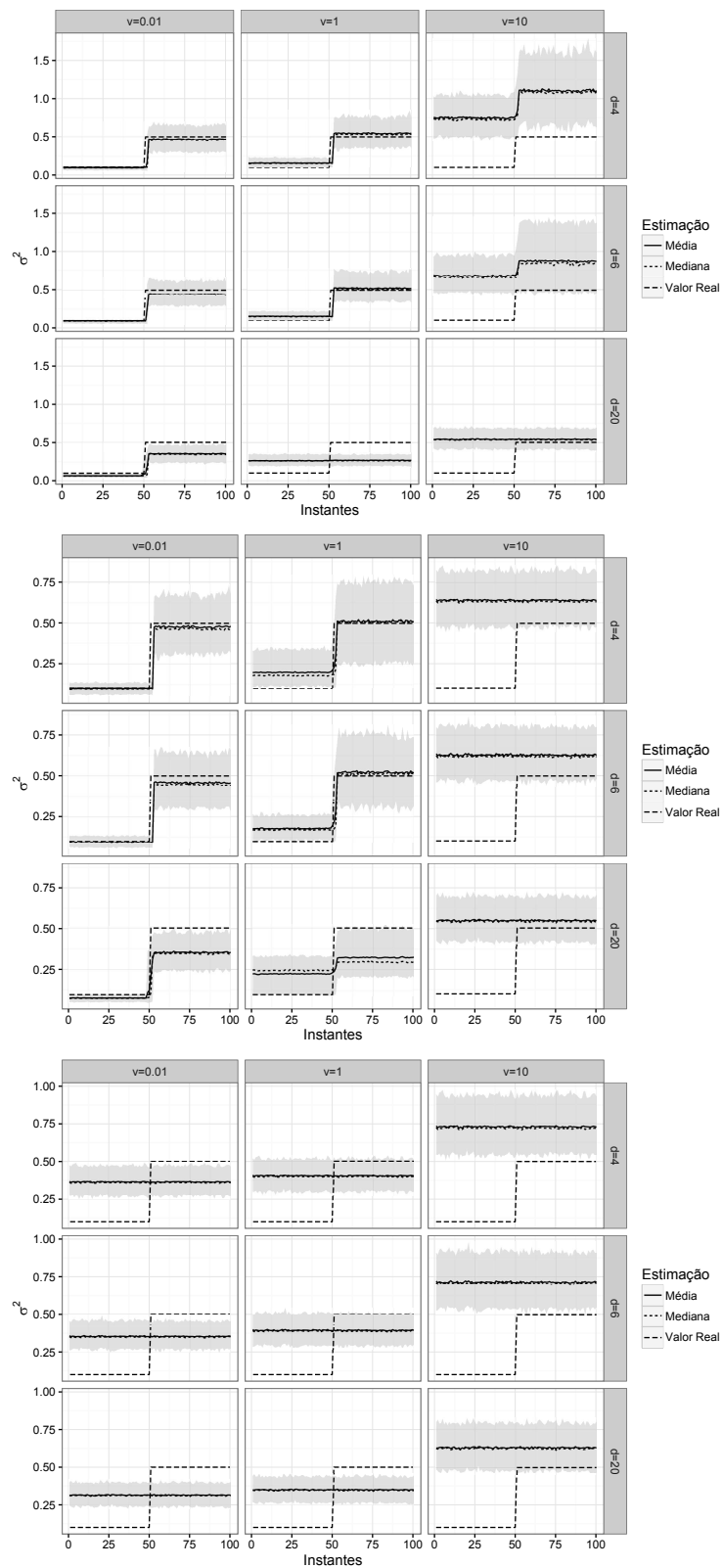


Figura 3.16: Estimativas do parâmetro σ_k^2 da série 2 para as Situações 1 (topo), 2 (meio) e 3 (base). A parte sombreada corresponde ao intervalo HPD de 95%.

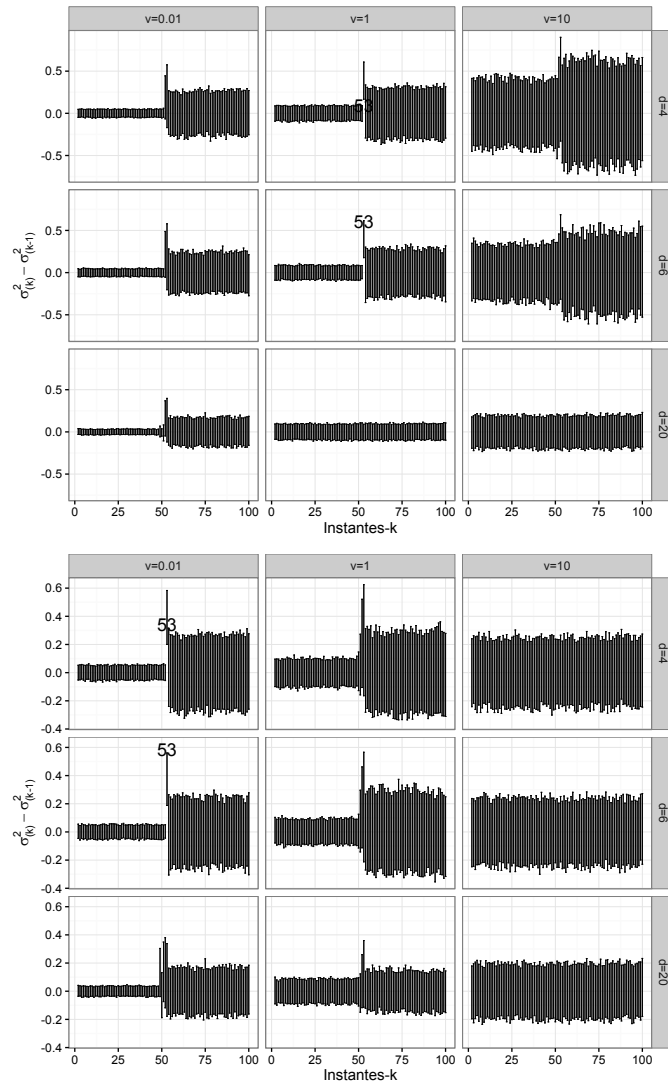


Figura 3.17: Intervalo HPD de 95% para $\sigma_k^2 - \sigma_{k-1}^2$ da série 2 para a Situação 1 (topo) e Situação 2 (base).

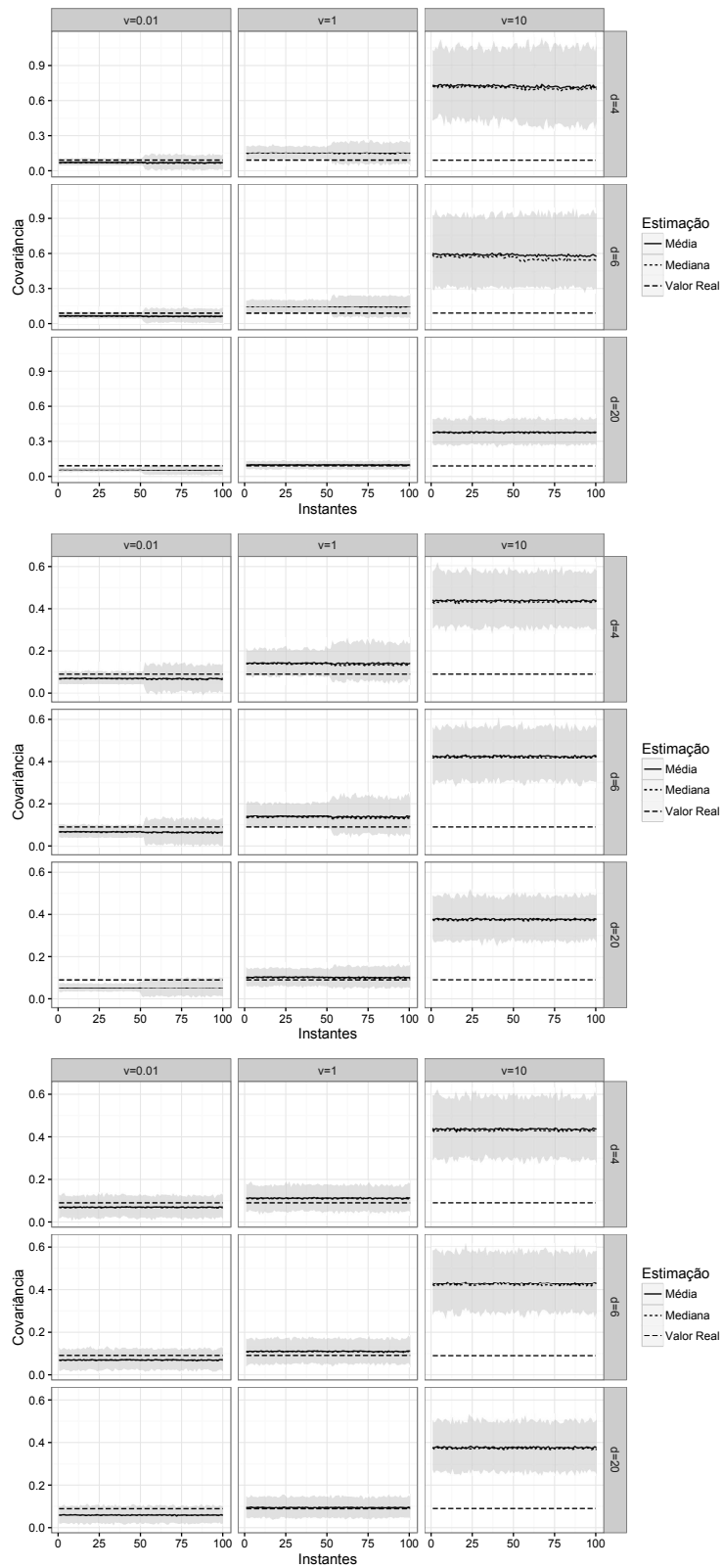


Figura 3.18: Estimativas do parâmetro $\sigma_{12(k)}^2$ para as Situações 1 (topo), 2 (meio) e 3 (base). A parte sombreada corresponde ao intervalo HPD de 95%.

3.3 Cenário 3: mudanças em todos os parâmetros

Este cenário foi o mesmo apresentado em Cheon & Kim [8]. Através dele, tem-se o objetivo avaliar o modelo quando há mudanças em todos os parâmetros. Foram geradas 1000 amostras com 8 pontos de mudanças da seguinte maneira:

$$\begin{aligned}
 \mathbf{y}_1, \dots, \mathbf{y}_{120} &\stackrel{iid}{\sim} N_2 \left[\begin{pmatrix} -3.5 \\ -3.5 \end{pmatrix}, \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1.3 \end{pmatrix} \right], \mathbf{y}_{121}, \dots, \mathbf{y}_{210} \stackrel{iid}{\sim} N_2 \left[\begin{pmatrix} 5.5 \\ 3 \end{pmatrix}, \begin{pmatrix} 1.2 & 0.2 \\ 0.2 & 1.2 \end{pmatrix} \right], \\
 \mathbf{y}_{211}, \dots, \mathbf{y}_{460} &\stackrel{iid}{\sim} N_2 \left[\begin{pmatrix} -4 \\ -7.5 \end{pmatrix}, \begin{pmatrix} 0.8 & -0.1 \\ -0.1 & 1.1 \end{pmatrix} \right], \mathbf{y}_{461}, \dots, \mathbf{y}_{530} \stackrel{iid}{\sim} N_2 \left[\begin{pmatrix} 4 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.1 & 0.1 \\ 0.1 & 1 \end{pmatrix} \right], \\
 \mathbf{y}_{531}, \dots, \mathbf{y}_{615} &\stackrel{iid}{\sim} N_2 \left[\begin{pmatrix} -5.5 \\ -7 \end{pmatrix}, \begin{pmatrix} 0.9 & 0.2 \\ 0.2 & 1.2 \end{pmatrix} \right], \mathbf{y}_{616}, \dots, \mathbf{y}_{710} \stackrel{iid}{\sim} N_2 \left[\begin{pmatrix} 2 \\ 5 \end{pmatrix}, \begin{pmatrix} 1.1 & -0.1 \\ -0.1 & 1.1 \end{pmatrix} \right], \\
 \mathbf{y}_{711}, \dots, \mathbf{y}_{800} &\stackrel{iid}{\sim} N_2 \left[\begin{pmatrix} -4 \\ -1 \end{pmatrix}, \begin{pmatrix} 1.3 & -0.1 \\ -0.1 & 1 \end{pmatrix} \right], \mathbf{y}_{801}, \dots, \mathbf{y}_{950} \stackrel{iid}{\sim} N_2 \left[\begin{pmatrix} 4.5 \\ 7.5 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix} \right], \\
 \mathbf{y}_{951}, \dots, \mathbf{y}_{1000} &\stackrel{iid}{\sim} N_2 \left[\begin{pmatrix} -3 \\ -1 \end{pmatrix}, \begin{pmatrix} 1.3 & -0.1 \\ -0.1 & 0.6 \end{pmatrix} \right].
 \end{aligned}$$

Os dados simulados podem ser vistos na Figura 3.19.

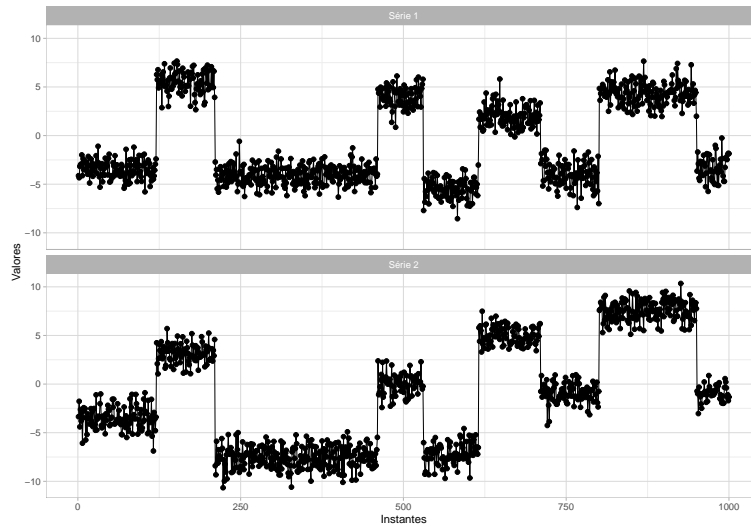


Figura 3.19: Valores das séries para o cenário 3.

Tabela 3.7: Média e variância *a priori* para $l = 1, 2$.

$E(\mu_l)$	$Var(\mu_l)$	$E(\sigma_l^2)$	$Var(\sigma_l^2)$	$E(\sigma_{12}^2)$	$Var(\sigma_{12}^2)$
0	10.0000	0.1000	-	0.01	-

Inferência sobre a partição, número de blocos e probabilidade de um instante ser mudança

Na Tabela 3.8 são apresentadas as dez partições *a posteriori* identificada pelo modelo. Nota-se que com probabilidade de 0.8517, o modelo conseguiu identificar a partição correta dos dados. Da Figura 3.20, nota-se que o modelo identificou 9 blocos, quantidade correta, com probabilidade 0.8517.

Tabela 3.8: Probabilidade a posteriori das partições.

ρ	Frequência de ρ	$P(\rho \mathbf{y}_1, \dots, \mathbf{y}_n)$
{0, 120, 210, 460, 530, 615, 710, 800, 950, 1000}	511	0.8517
{0, 120, 210, 460, 530, 615, 710, 800, 950, 997, 1000}	45	0.0750
{0, 120, 210, 460, 530, 615, 710, 800, 950, 996, 1000}	22	0.0367
{0, 120, 210, 460, 530, 615, 710, 800, 950, 998, 1000}	7	0.0117
{0, 120, 210, 460, 527, 530, 615, 710, 800, 950, 1000}	4	0.0067
{0, 120, 210, 460, 461, 530, 615, 710, 800, 950, 1000}	2	0.0033
{0, 120, 209, 210, 460, 530, 615, 710, 800, 950, 1000}	1	0.0017
{0, 120, 210, 459, 460, 530, 615, 710, 800, 950, 1000}	1	0.0017
{0, 120, 210, 460, 527, 530, 615, 710, 800, 950, 997, 1000}	1	0.0017
{0, 120, 210, 460, 528, 530, 615, 710, 800, 950, 997, 1000}	1	0.0017

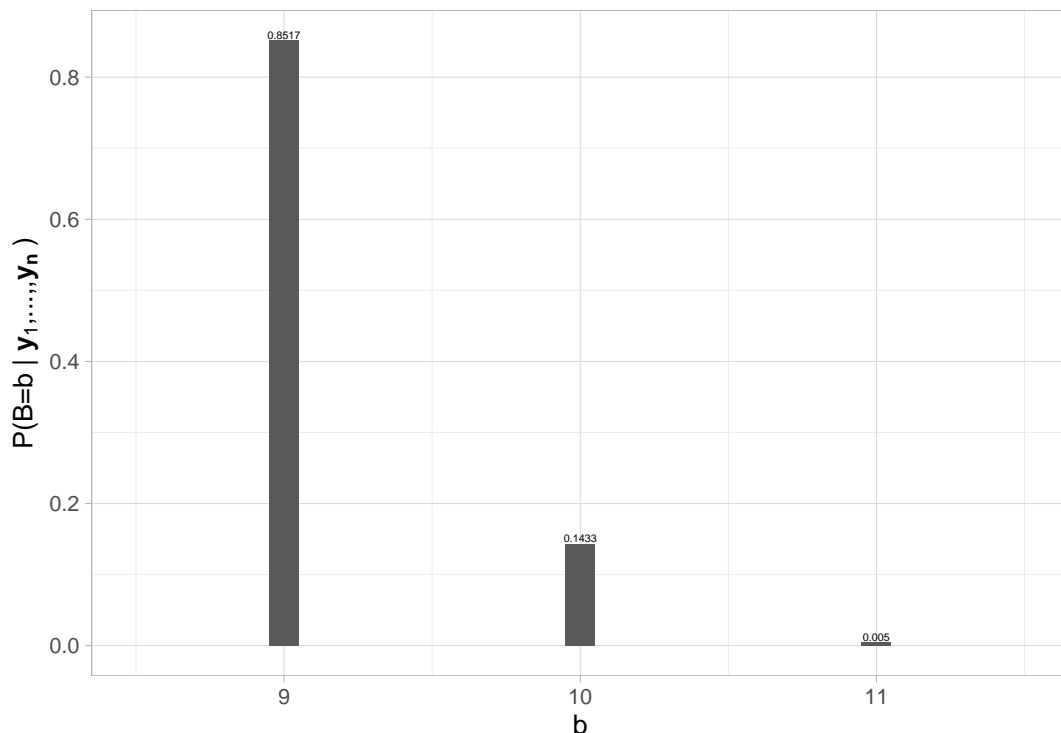


Figura 3.20: Probabilidade *a posteriori* do número de blocos.

Na Figura 3.21 é apresentada a distribuição *a posteriori* da probabilidade de mudança. Note que a distribuição está concentrada em valores pequenos, indicando

que a probabilidade de ter mudança é baixa. A média e mediana de p são, respectivamente, 0.00915 e 0.00882, enquanto o desvio padrão é de 0.0030

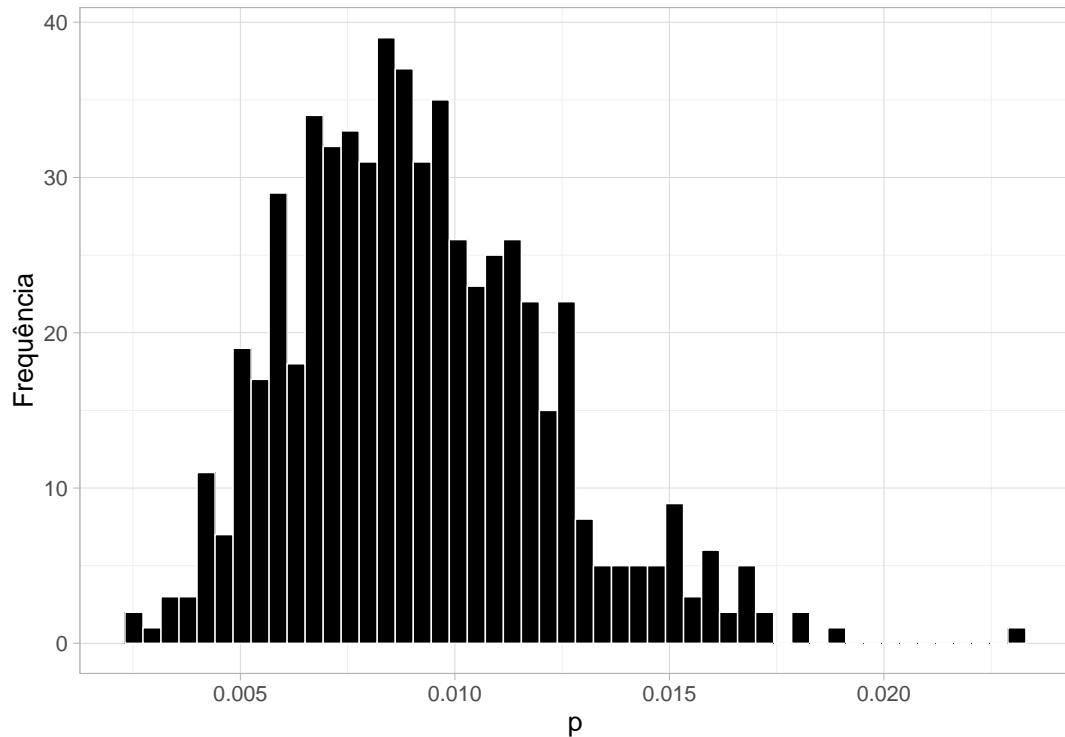


Figura 3.21: Distribuição *a posteriori* da probabilidade de mudança.

Na Figura 3.22 é apresentada a probabilidade *a posteriori* de cada instante ser um ponto de mudança. Nota-se que os instantes reais de mudança foram identificados com probabilidade igual a 1 e os demais instantes tiveram probabilidade pequena ou zero de serem pontos de mudança.

Cheon & Kim [8] apresentou dez partições identificado pelo modelo, identificando a partição correta dos dados através do *Bayesian Information Criterion* (BIC). Além disso, atribuiu alta probabilidade para os instantes reais de mudanças e consequentemente determinou de maneira correta o número de pontos de mudanças. No entanto, não realizou a estimação para a probabilidade de mudanças e nem para μ_k e Σ_k .

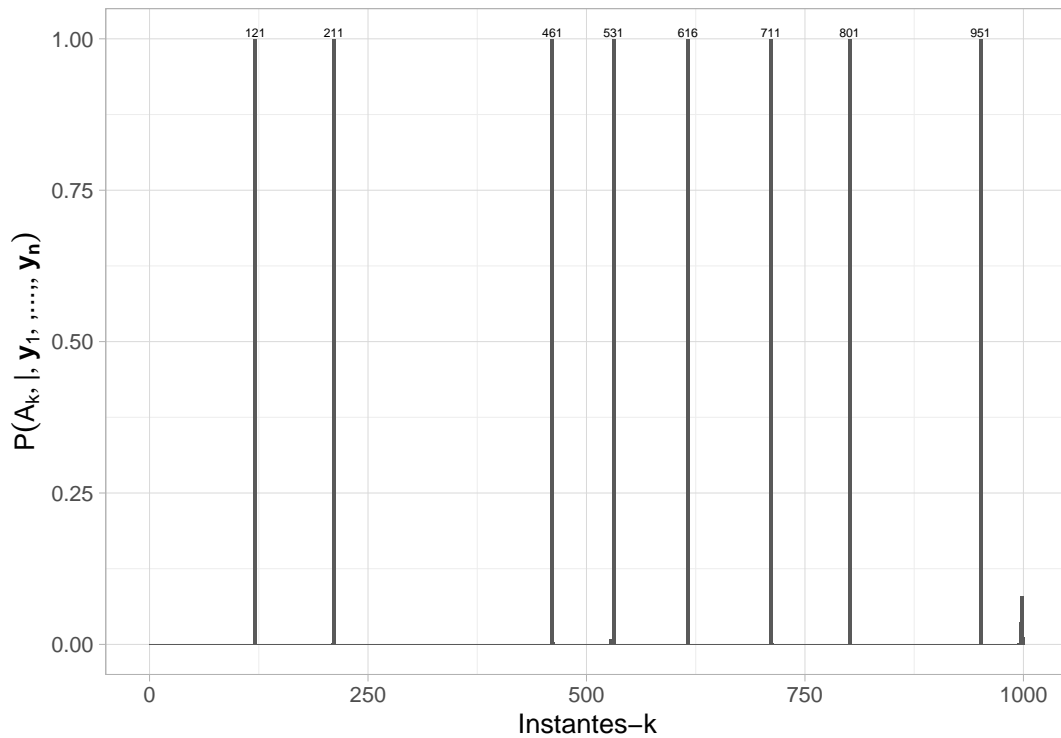


Figura 3.22: Probabilidade *a posteriori* de cada instante ser um ponto de mudança. Os números acima das barras representam os instantes

Analisando μ_k e Σ_k

A Figura 3.23 mostra as estimativas de μ_k da série 1 e 2. Nota-se que as estimativas estão bem próximas dos valores reais. Da Figura 3.24, nota-se os instantes das mudanças significativas no vetor de médias das séries 1 e 2. Percebe-se que houve oito mudanças significativas no vetor médias, sendo todas identificadas de acordo com a partição real dos dados.

Na Figura 3.25 é apresentada a estimação de σ_k^2 para as séries 1 e 2. Nota-se que para alguns instante houve uma subestimação ou superestimação dos valores. Em relação as mudanças significativas, o instante 121 é uma mudança significativa da variância da série 1 e o instante 951 é uma mudança significativa da variância da série 2.

A estimação da covariância entre as séries também é apresentada na Figura 3.25. É possível notar superestimação ou subestimação dos valores. O instante em que houve uma mudança significativa na covariância foi o 211.

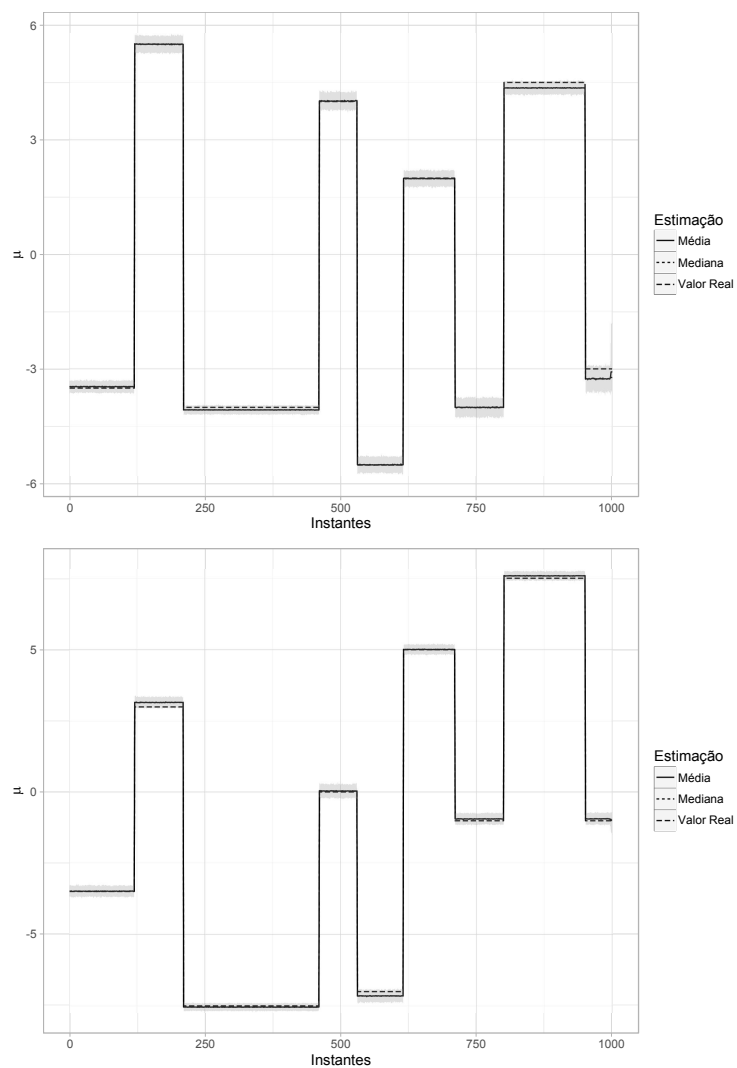


Figura 3.23: Estimativas do parâmetro μ_k das séries 1 (topo) e 2 (base). A parte sombreada corresponde ao intervalo HPD de 95%.

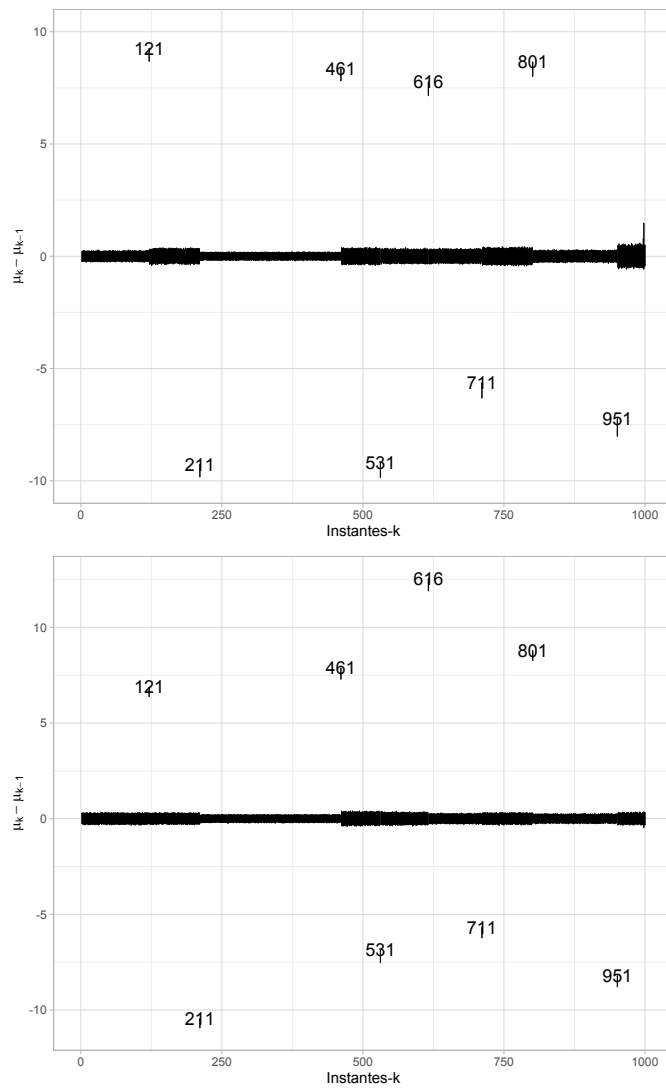


Figura 3.24: Intervalo HPD de 95% para $\mu_k - \mu_{k-1}$, séries 1 (topo) e 2 (base).

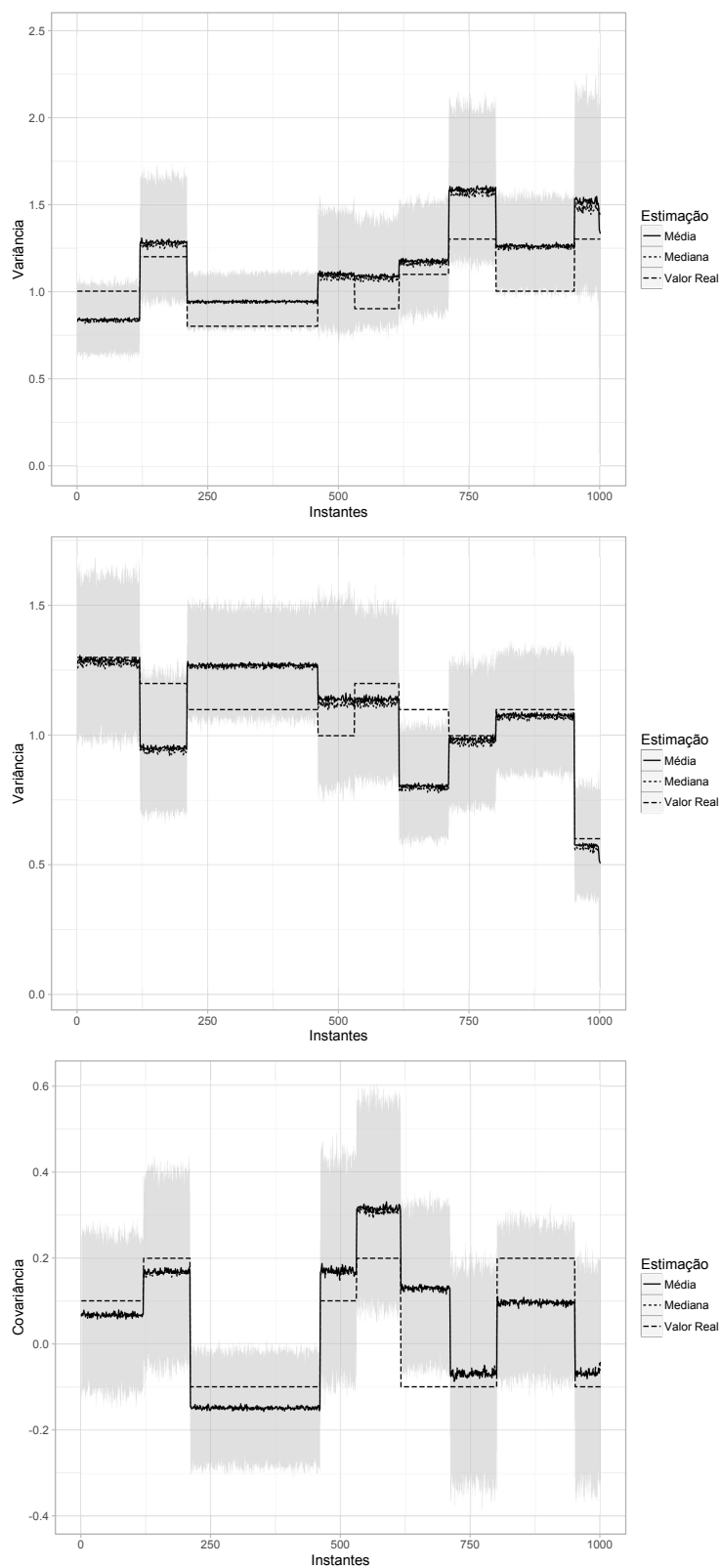


Figura 3.25: Estimativas do parâmetro σ_k^2 das séries 1 (topo) e 2 (meio) e do parâmetro $\sigma_k^2(12)$ (base). A parte sombreada corresponde ao intervalo HPD de 95%.

Capítulo 4

Avaliando o efeito do *Brexit* na economia

Em 23 de junho de 2016, os britânicos realizaram um plebiscito para decidir se o Reino Unido deveria ou não permanecer unido ao bloco de países europeus conhecido como União Europeia (UE). A UE é uma união econômica e política composta por 28 países a qual o Reino Unido aderiu em 1973. Entre outras coisas, este acordo conhecido como UE prevê a livre circulação de bens e pessoas entre países que o compõe. Durante o período que antecedeu a vitória do sim à separação, houve um período de grande instabilidade econômica devido as divergências sobre os efeitos econômicos oriundos desta especulação de separação.

Diante disso, este capítulo tem como objetivo avaliar efeito do *Brexit* na economia de oito países. Esse termo é a abreviação das palavras em inglês *Britain* (Grã-Bretanha) e *exit* (saída) e designa a saída do Reino Unido da União Europeia. A análise de séries multivariadas é de grande interesse quando tem-se o objetivo de compreender a relação existente entre as mesmas. No âmbito econômico, por exemplo, é interessante estudar como os diferentes índices financeiros se relacionam, pois isso possibilita averiguar as contaminações ou influências que um determinado índice impõe sobre os demais.

Para avaliar o efeito de tal acontecimento em importantes economias mundiais, será aplicado modelo partição produto para dados normal multivariada (MPPNM), desenvolvido no Capítulo 2, para a identificação de possíveis pontos de mudanças, ao longo do tempo, no vetor de médias e na estrutura de covariância de retornos de nove índices que serão apresentados na Seção 4.1. Escolheu-se os índices de economias fortes do bloco europeu como Alemanha, França e Suíça e também países que estiveram recentemente em crise como Espanha, Itália, Grécia e Portugal. Uma economia fora do bloco europeu também foi considerada, os Estados Unidos.

4.1 Banco de dados

O banco de dados a ser analisado é formado pelas seguintes séries de retornos: DAX (*Deutscher Aktienindex*) da Alemanha, IBEX 35 (*Iberia Index*) da Espanha, CAC 40 (*Cotation Assistée en Continu*) da França, ATG (*Athens General*) da Grécia, FTSE MIB (*Financial Times Stock Exchange Milano Indice di Borsa*) da Itália, PSI-20 (*Portuguese Stock Index*) de Portugal, FTSE 100 Reino Unido (*Financial Times Stock Exchange*), SMI (Swiss Market Index) da Suíça e S&P 500 (Standard & Poor's 500) dos Estados Unidos. Os dados estão disponíveis em <http://br.investing.com/>. Foi considerado um total de 200 observações que compreende os retornos diários (dias úteis) de tais índices observados no período entre 4 de janeiro de 2016 e 10 de outubro de 2016. A escolha do período citado deve-se ao fato de que nesse período houve o plebiscito no dia 23 de junho que definiu pela saída do Reino Unido da União Europeia(UE). Tem-se o interesse de avaliar os impactos que tal acontecimento proporcionou na economia dos demais países.

Define-se o retorno do índice l no instante k a variável $R_{lk} = \frac{P_{lk} - P_{l(k-1)}}{P_{l(k-1)}}$, em que P_{lk} é o valor atingido pelo índice l no fechamento do mercado no dia k . Assume-se que

$$\mathbf{Y}_k \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \sim N_9(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad \text{para } k = 1, \dots, 200,$$

onde \mathbf{Y}_k é um vetor, de ordem 9×1 , cujas componentes R_{lk} , para $l = 1, \dots, 9$, correspondem, respectivamente, aos retornos dos índices relacionados a Alemanha, Espanha, EUA, França, Grécia, Itália, Portugal, Reino Unido e Suíça. Os valores dos retornos são mostrados nas Figuras 4.1 e 4.2.

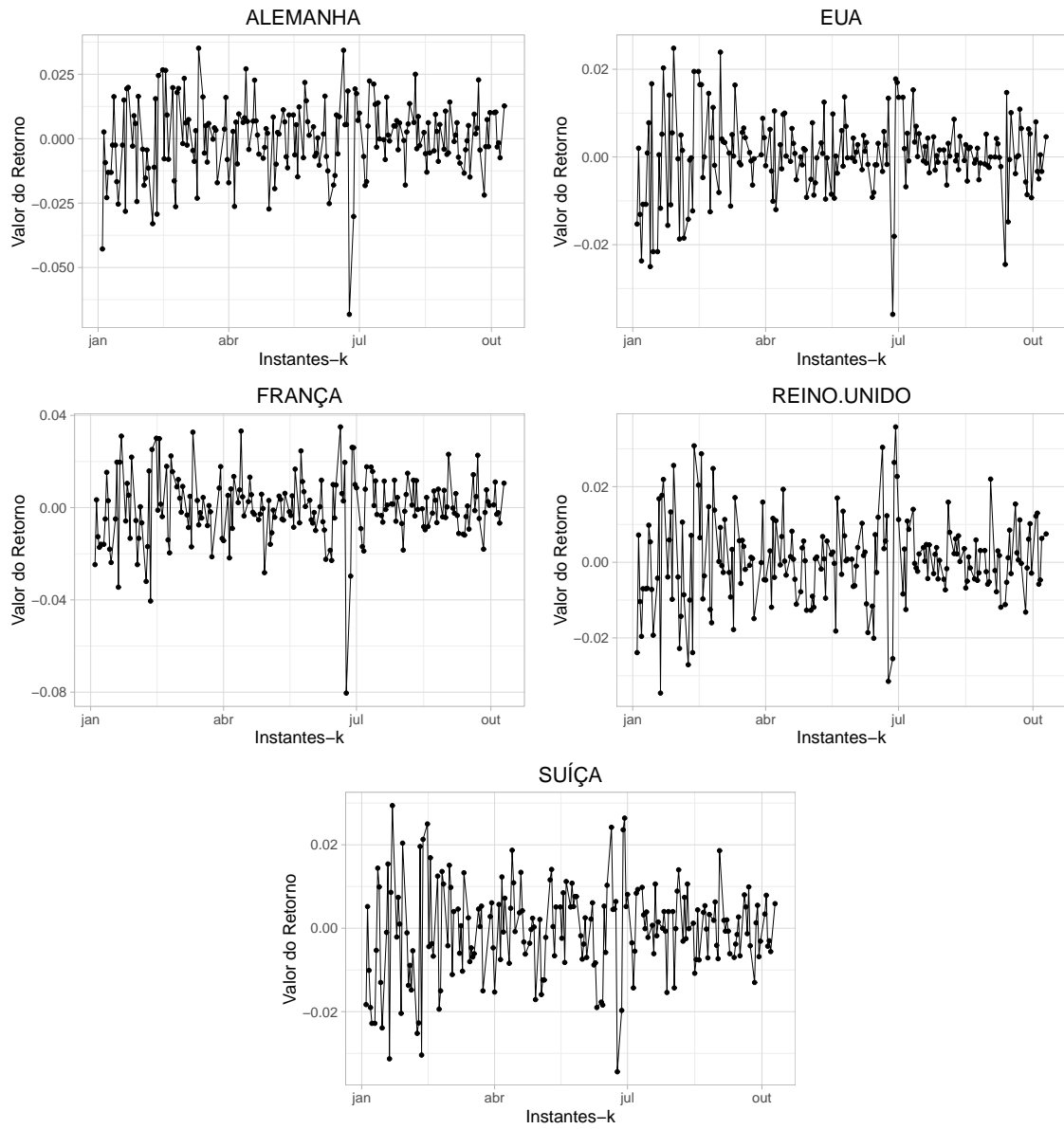


Figura 4.1: Retornos dos índices financeiros da Alemanha, EUA, França, Reino Unido e Suíça.

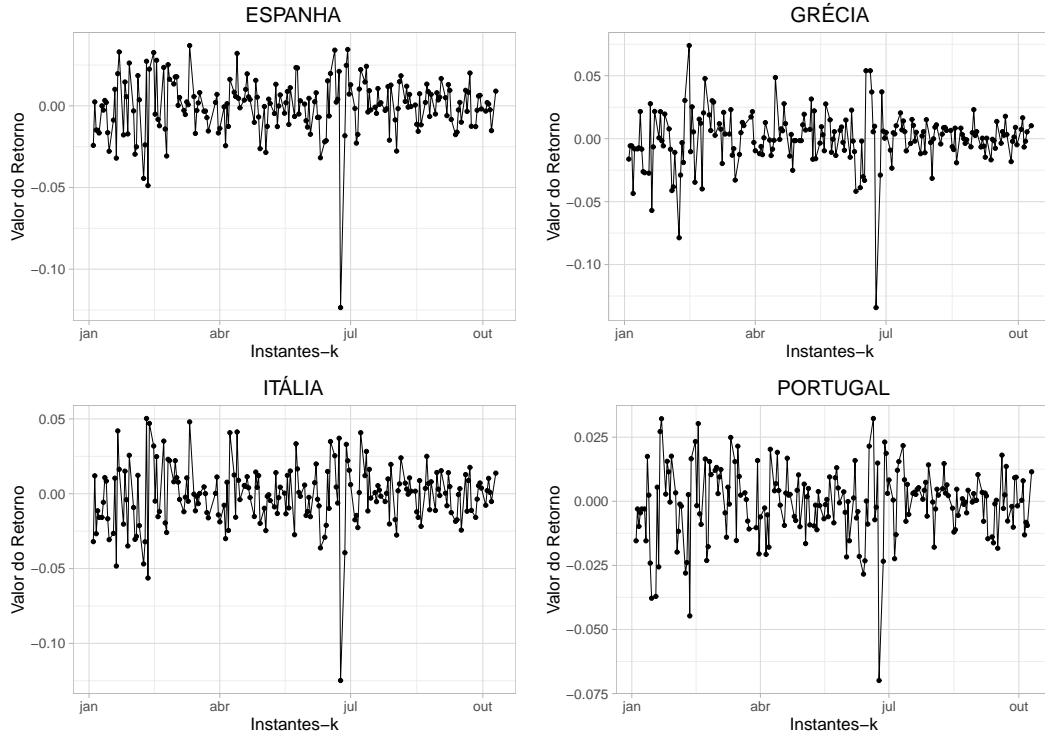


Figura 4.2: Retornos dos índices financeiros da Espanha, Grécia, Itália e Portugal.

4.2 Especificações *a priori*

Considerando as suposições do modelo MPPNM, apresentadas na seção 2.2, segue que as especificações *a priori* para os parâmetros comuns $(\boldsymbol{\mu}_{[ij]}, \boldsymbol{\Sigma}_{[ij]})$ são tais que

$$\boldsymbol{\mu}_{[ij]} | \boldsymbol{\Sigma}_{[ij]} \sim N_9 \left[\mathbf{m}, \frac{1}{0.001} \boldsymbol{\Sigma}_{[ij]} \right] \quad \text{e}$$

$$\boldsymbol{\Sigma}_{[ij]} \sim WI(\mathbf{D}, 13),$$

onde \mathbf{m} é um vetor de zeros com ordem 9×1 e \mathbf{D} é uma matriz diagonal de ordem 9×9 com a componentes 0.00001. Dessa forma, assume-se que, *a priori*, as variâncias dos retornos são, em média, muito pequenas e que os retornos são independentes. Estas especificações também revelam que, *a priori*, a média dos retornos estão em torno de zero, (ver Tabela 4.1).

Tabela 4.1: Média e variância dos parâmetros, para $l, t = 1, \dots, 9$

d	v	$E(\boldsymbol{\mu}_l)$	$Var(\boldsymbol{\mu}_l)$	$E(\sigma_l^2)$	$Var(\sigma_l^2)$	$E(\sigma_{lt}^2)$	$E(\sigma_{lt}^2)$	
1	13	0.001	0	0.0033	3e-06	2.2222e-11	0	8.333333e-12

Como as economias que estão sob análise são razoavelmente estáveis e nenhum

evento da magnitude do *Brexit* que pudesse afetar a economia de toda a Europa ocorreu no período em análise, espera-se que poucos ou nenhum ponto de mudança ocorra. Por causa disto, considera-se a distribuição *a priori* da probabilidade p de que ocorra uma mudança em um instante qualquer é uma distribuição Beta (2,198), isto é $p \sim \text{Beta}(2, 198)$, cuja, média e desvio padrão são, respectivamente, 0.01 e 0.0070181. Assumindo esta distribuição, o número esperado de blocos *a priori* é 2.99 com variância de 3.910796.

Para o amostrador de Gibbs, foram geradas 10000 amostras. Para a formação da amostra *a posteriori*, após a convergência ter sido atingida, foram descartadas as 4000 primeiras gerações e, assumindo *lag* 10, foram selecionadas 600 observações das 6000 restantes. O algoritmo foi implementado em R Core Team [39] e processado no computador Intel Core i7, 3.40GHz e 8 GB RAM. O tempo do CPU foi de 4.98 segundos. Para contornar o problema de *overflow* foi utilizada a função *mpfr* do pacote *Rmpfr*, Maechler [30]. As cadeias de alguns parâmetros podem ser vistas no Apêndice D.

4.3 Análise

Nesta seção serão apresentados os resultados obtidos a partir da utilização do MPPNV no banco de dados descritos na seção 4.1. Diante desses resultados, temos como objetivos avaliar o impacto da saída do Reino Unido da União Europeia através da identificação de possíveis pontos de mudança.

Inferência sobre a partição, número de blocos e probabilidade de um instante ser mudança

A Tabela 4.2 mostra as probabilidades *a posteriori* das partições identificadas pelo modelo. Note que, com probabilidade de 0.9817, a partição $\rho = \{0, 23/06, 24/06, 27/06, 10/10\}$ é mais provável. Perceba que com tal partição, os pontos de mudança identificados pelo modelo são 24/06, 27/06 e 28/06, datas próximas a ocorrência do *Brexit*. Tais pontos de mudanças também podem ser observados nas outras partições que ocorrem com menor probabilidade.

Tabela 4.2: Probabilidade *a posteriori* das partições.

ρ	Frequência de ρ	$P(\rho \mathbf{y}_1, \dots, \mathbf{y}_n)$
$\{0, 23/06, 24/06, 27/06, 10/10\}$	589	0.9817
$\{0, 23/06, 24/06, 10/10\}$	9	0.0150
$\{0, 1/4, 23/06, 24/06, 27/6, 10/10\}$	1	0.0017
$\{0, 23/06, 24/06, 27/06, 28/6, 10/10\}$	1	0.0017

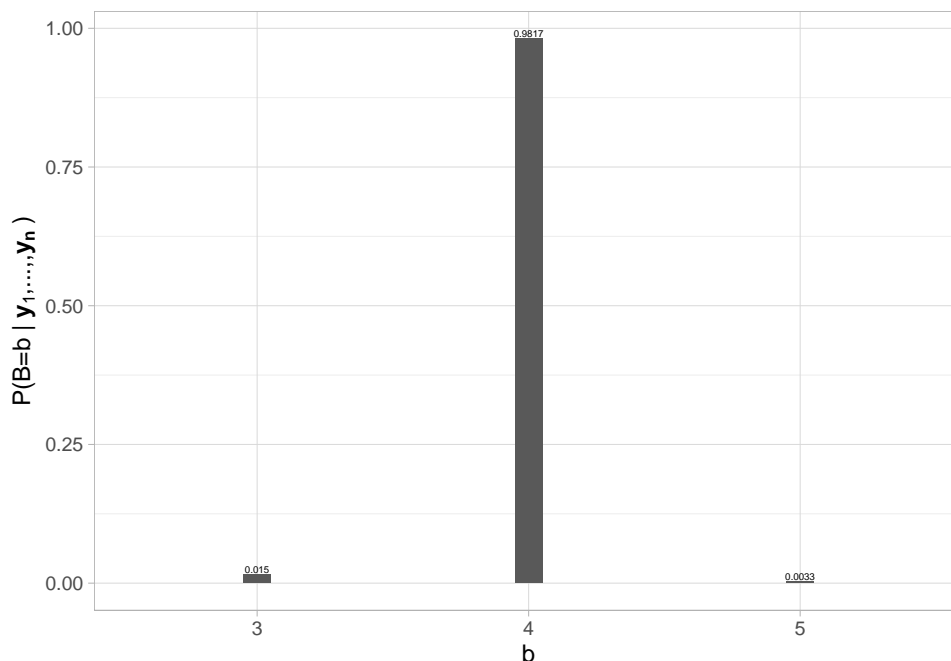


Figura 4.3: Probabilidade *a posteriori* dos blocos.

A Figura 4.3 mostra a probabilidade *a posteriori* do número de blocos observados na série multivariada. Com probabilidade *a posteriori* de 0.9817, são observados 4 blocos, indicando que, provavelmente, três mudanças correram na série.

Analisando Tabela 4.3, percebe-se que a média da distribuição *a posteriori* é de 0.01245, apenas um pouco maior do que foi assumido *a priori*. Na Figura 4.4 é mostrada a distribuição *a posteriori* da probabilidade de mudança p . Perceba que tal distribuição concentra massa probabilística em valores pequenos, logo, a probabilidade de haver mudança em algum instante do tempo é pequena.

A Figura 4.5 mostra as probabilidades *a posteriori* de cada instante ser um ponto de mudança. Perceba que os instantes 24/6 e 27/6 são identificados como pontos de mudança com probabilidade 1, enquanto 28/6 é ponto de mudança com probabilidade de 0.985. Tais pontos estão próximo do dia 23/6 em que houve a votação favorável à saída do Reino Unido da UE, logo, havendo indícios de que tal acontecimento possa ter afetado a economia dos países envolvidos. Diante disso, se faz necessário analisar como acontece essa contaminação entre os índices, ou seja, se tais pontos são decorridos de mudanças no vetor de médias ou na estrutura de covariância ou em ambas ao mesmo tempo e de como essas mudanças interferem nas diferentes séries em estudo.

Tabela 4.3: Estimativas *a posteriori* para a probabilidade de mudança.

Média	Mediana	Desvio padrão
0.01245	0.0117	0.0057

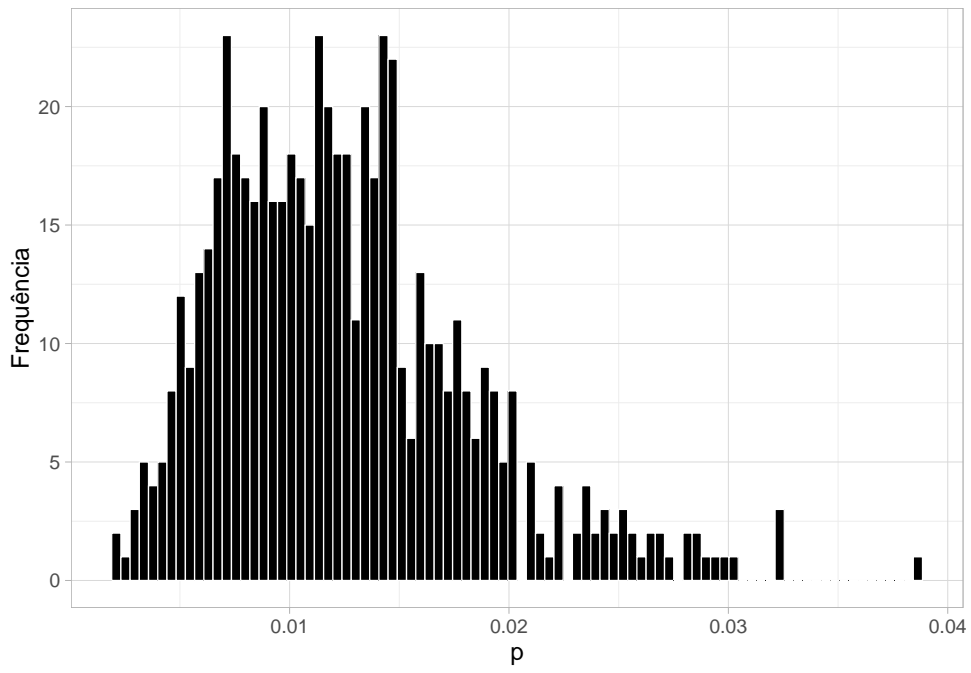


Figura 4.4: Distribuição *a posteriori* da probabilidade de mudança.

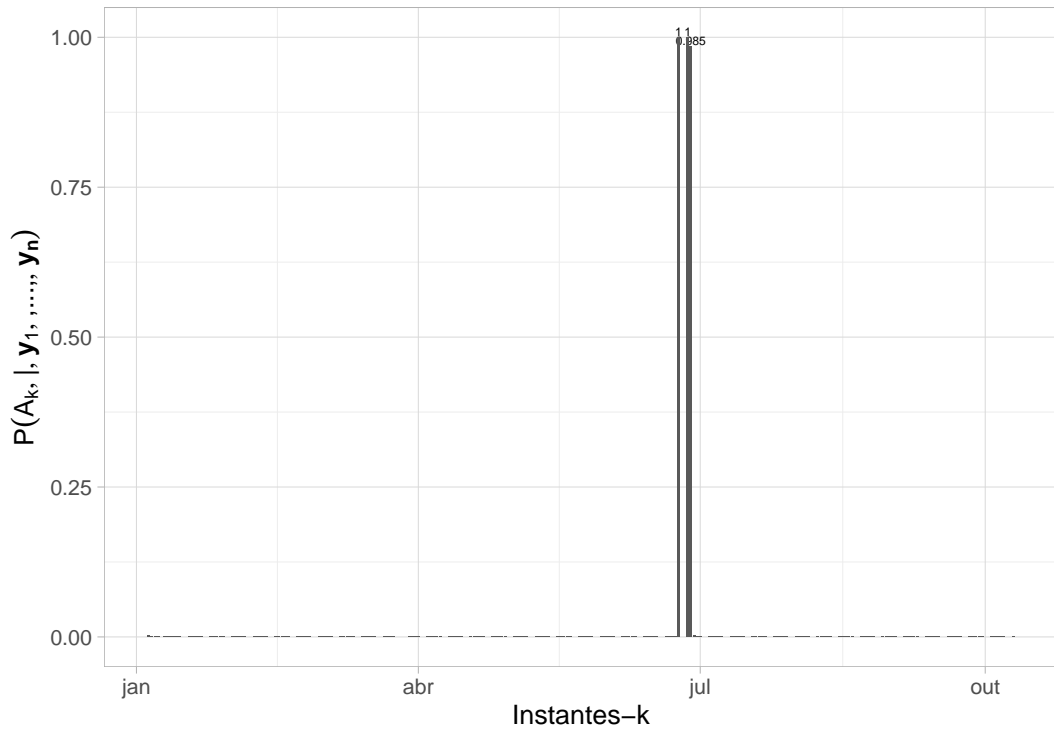


Figura 4.5: Probabilidade *a posteriori* de cada instante ser ponto de mudança.

Análise de μ_k e Σ_k

Das Figuras 4.6, 4.7 e 4.8, percebe-se que o retorno médio dos índices estão bem parecidos até o dia 23/06, onde os instantes 124, 125 e 126 correspondem as datas 24/06/2016, 27/06/2016 e 28/06/2016 respectivamente. Dentre tais índices, destaca-se o retorno do Reino Unido que possui um comportamento em torno de zero até o dia 23/06 e passando obter uma reação positiva após o dia 24/06, indicando um fortalecimento da sua economia após a votação favorável a saída da UE.

Das Figuras 4.9, 4.10 e 4.11, percebe-se que a variabilidade dos retornos são diferentes. O índice relacionado a Grécia, por exemplo, apresenta a maior variabilidade, enquanto o do EUA exibe a menor, seguido dos índices da Suíça e do Reino Unido, indicando que tais economias são bastante estáveis. Nota-se que após 28/06 as variâncias dos retornos registraram um decréscimo, consequentemente indicando uma maior estabilidade do mercado após a saída do Reino Unido.

As Figuras 4.12, 4.13 e 4.14 apresentam as covariâncias do índice do Reino Unido com os demais, as outras relações de covariância são apresentadas no Apêndice D. Nota-se que a maior covariância ocorre entre Reino Unido e Itália e a menor entre Reino Unido e EUA. Percebe-se também que, assim como na variância, houve um decréscimo na covariância entre os índices.

Mudanças significativas no vetor de médias foram detectadas nos instantes 24/6, 27/6 e 28/6. No entanto, nota-se que tais mudanças influenciam de maneira diferente a média dos índices. Para a média do retorno dos EUA, por exemplo, o dia 24/6 registrou um aumento, enquanto para os demais países houve uma queda na média, sendo o índice relacionado ao Reino Unido com a menor queda e o índice da Grécia com a maior queda. No entanto, o dia 27/6 registrou uma queda na média do índice relacionado ao EUA e para os demais houve um aumento. Após o dia 28/06 houve um crescimento na média dos retornos, sendo o índice relacionado ao Reino Unido com o maior aumento registrado.

As variâncias dos retornos apresentaram mudanças significativas nos dias 24/6 e 28/6. No dia 24/6 houve uma queda na variabilidade do retorno para todos os índices e no 28/6 houve um aumento também para todos. Na covariância, os instantes de mudanças significativas foram os mesmos registrados na variância. Tais instantes identificaram queda seguido de crescimento. Assim como na variância, também houve um decréscimo na covariância entre os índices após o dia 28/6.

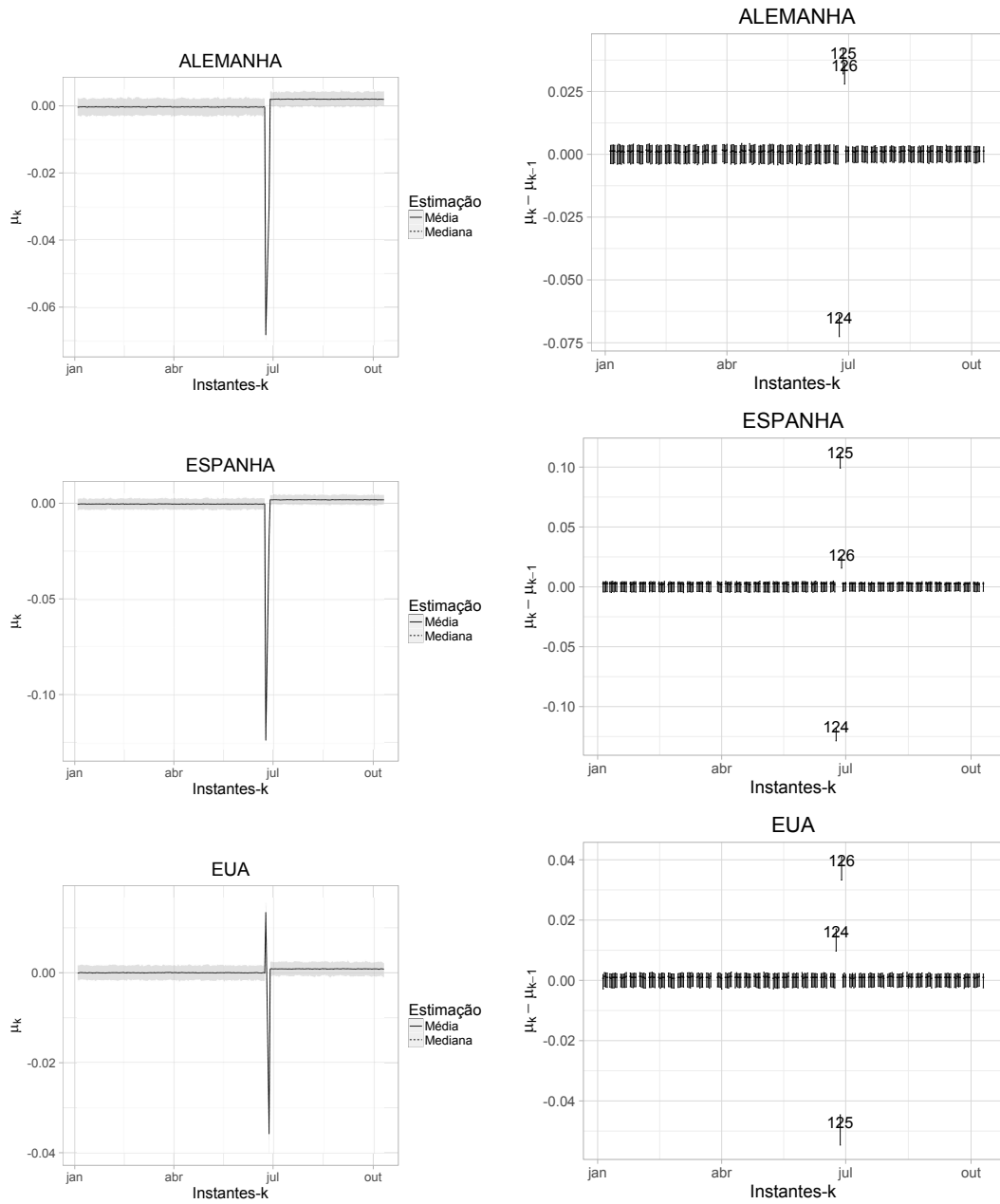


Figura 4.6: Estimativas μ_k (esquerda) e o intervalo HPD de 95% para $\mu_k - \mu_{k-1}$ (direita) para os índices da Alemanha, Espanha e Estados Unidos.

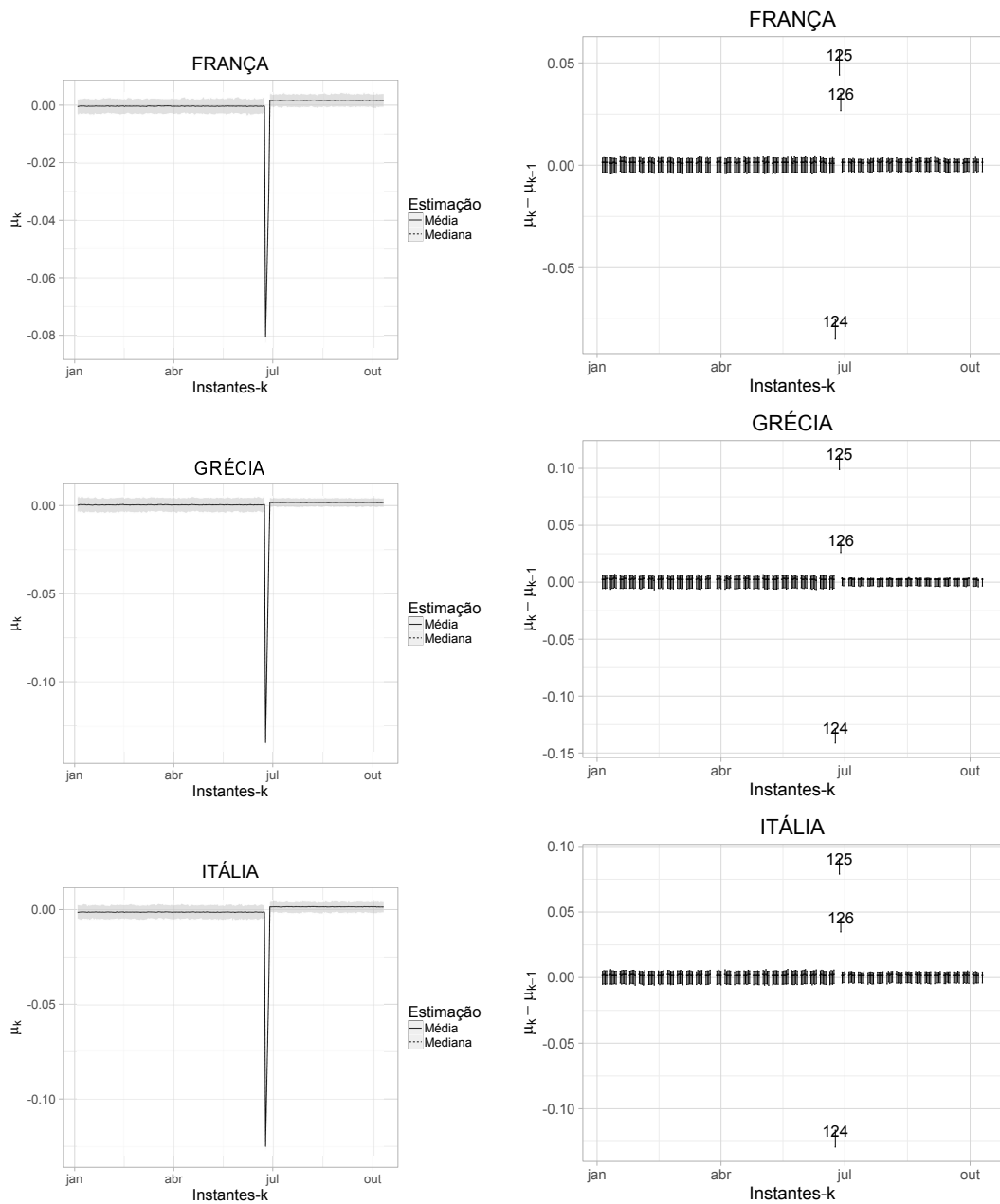


Figura 4.7: Estimativas μ_k (esquerda) e o intervalo HPD de 95% para $\mu_k - \mu_{k-1}$ (direita) para os índices da França, Grécia e Itália.

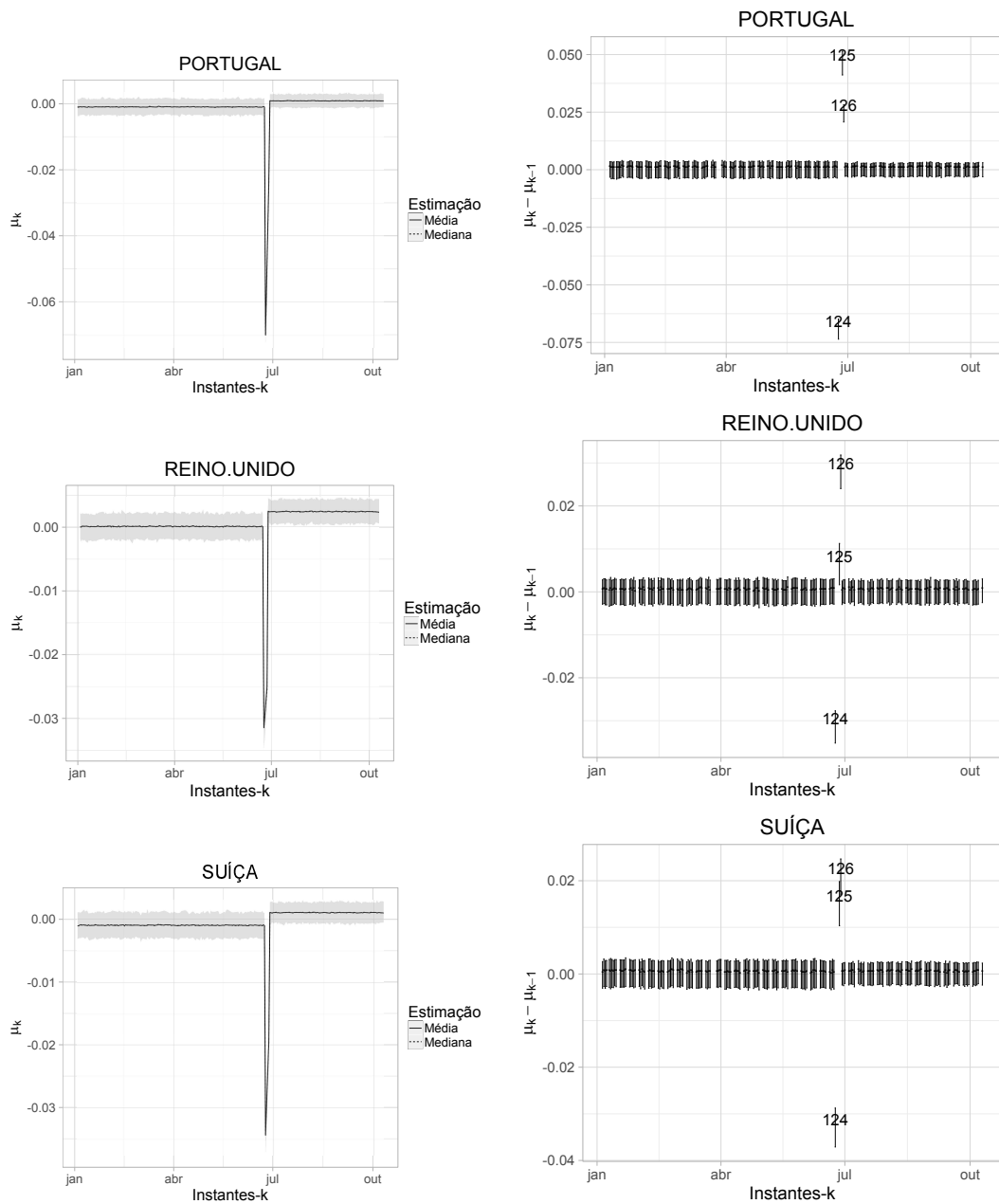


Figura 4.8: Estimativas μ_k (esquerda) e o intervalo HPD de 95% para $\mu_k - \mu_{k-1}$ (direita) para os índices de Portugal, Reino Unido e Suíça.

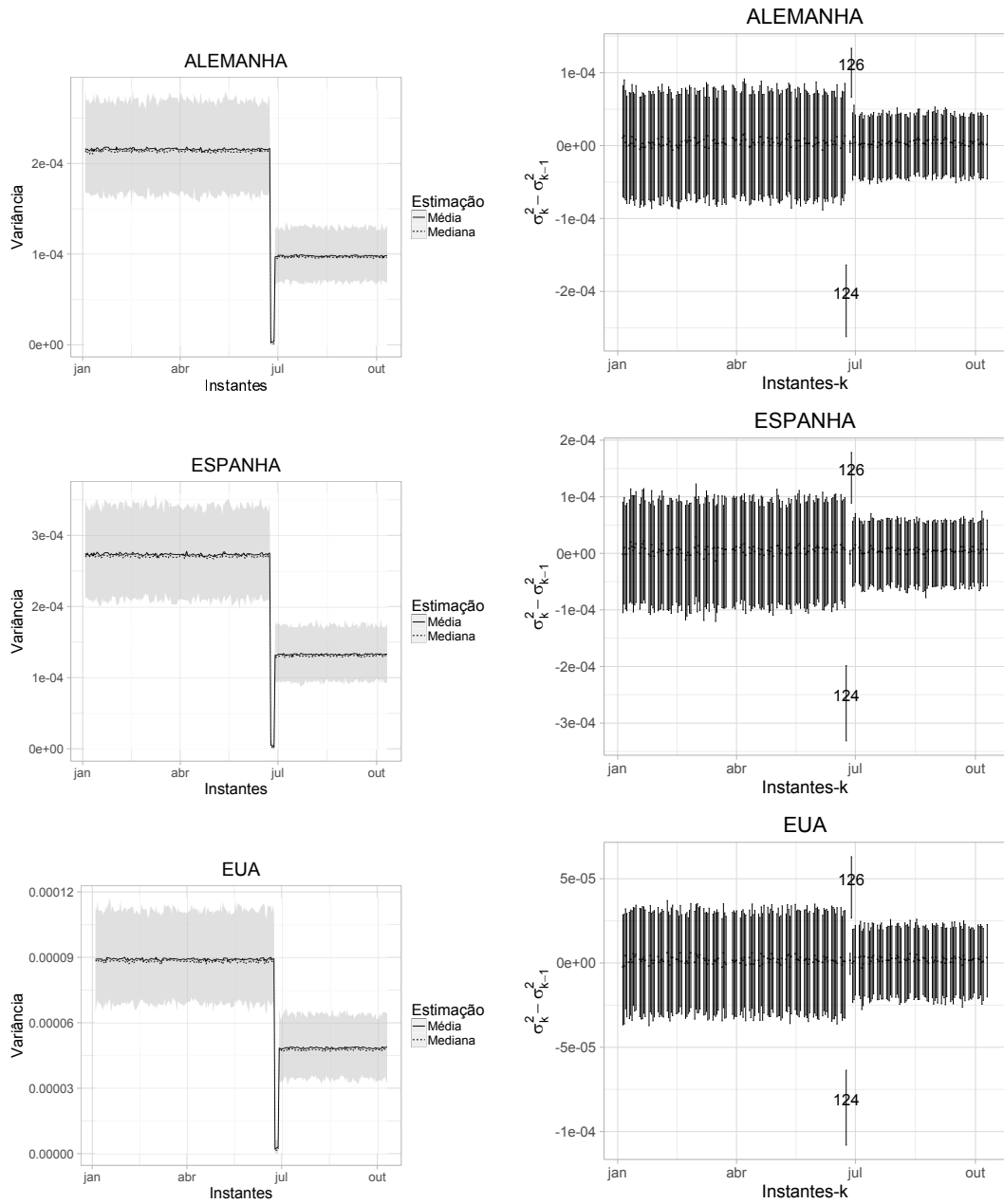


Figura 4.9: Estimativas de σ_k^2 (esquerda) e o intervalo HPD de 95% para $\sigma_k^2 - \sigma_{k-1}^2$ (direita) para os índices da Alemanha, Espanha e Estados Unidos.

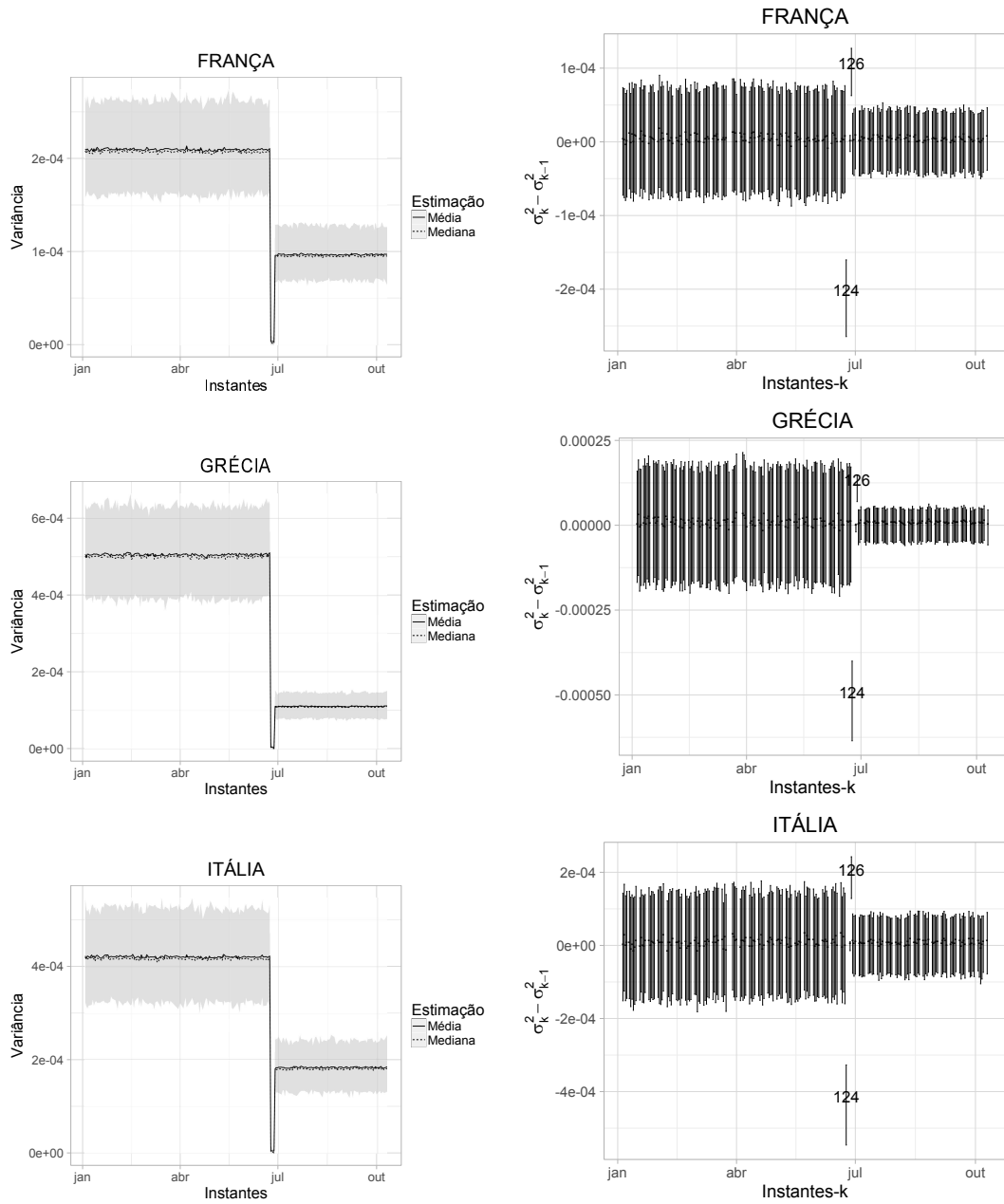


Figura 4.10: Estimativas de σ_k^2 (esquerda) e o intervalo HPD de 95% para $\sigma_k^2 - \sigma_{k-1}^2$ (direita) para os índices da França, Grécia e Itália.

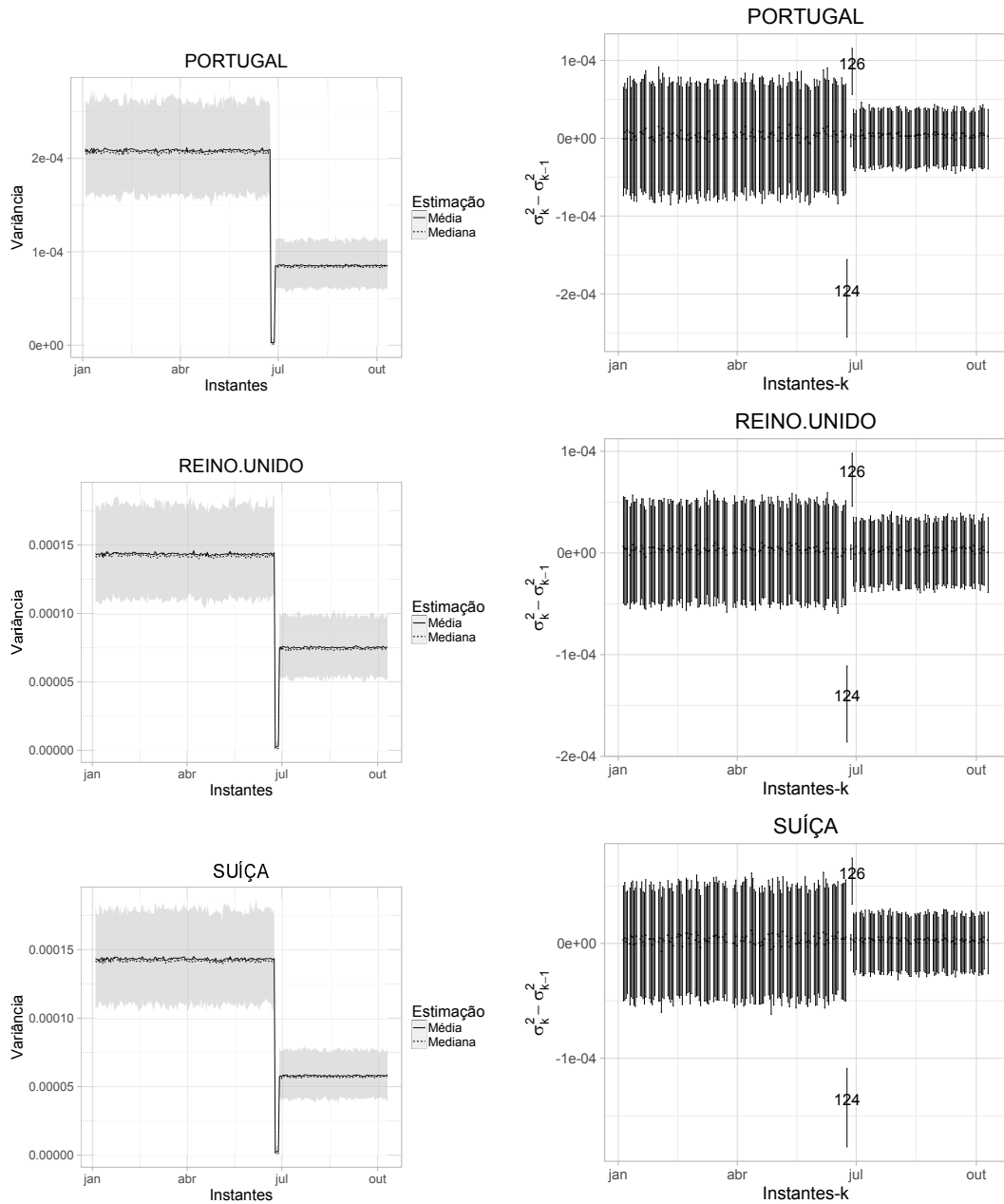


Figura 4.11: Estimativas de σ_k^2 (esquerda) e o intervalo HPD de 95% para $\sigma_k^2 - \sigma_{k-1}^2$ (direita) para os índices de Portugal, Reino Unido e Suíça.

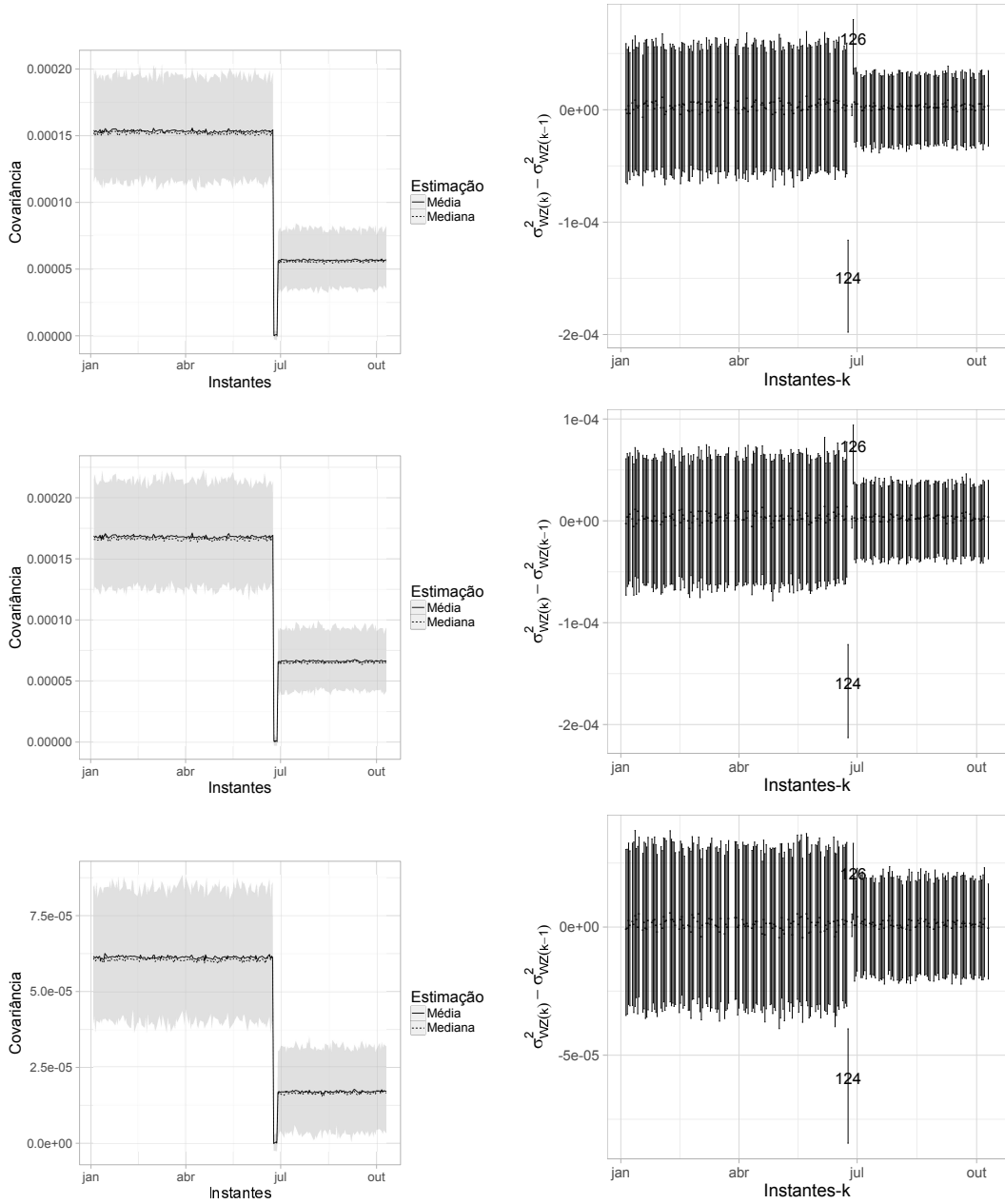


Figura 4.12: Estimativas de $\sigma_{WZ(k)}^2$ (esquerda) e intervalos HPD's de 95% para $\sigma_{(WZ)k}^2 - \sigma_{WZ(k-1)}^2$ (direita) entre os índices Reino Unido e Alemanha, Reino Unido e Espanha e Reino Unido e EUA.

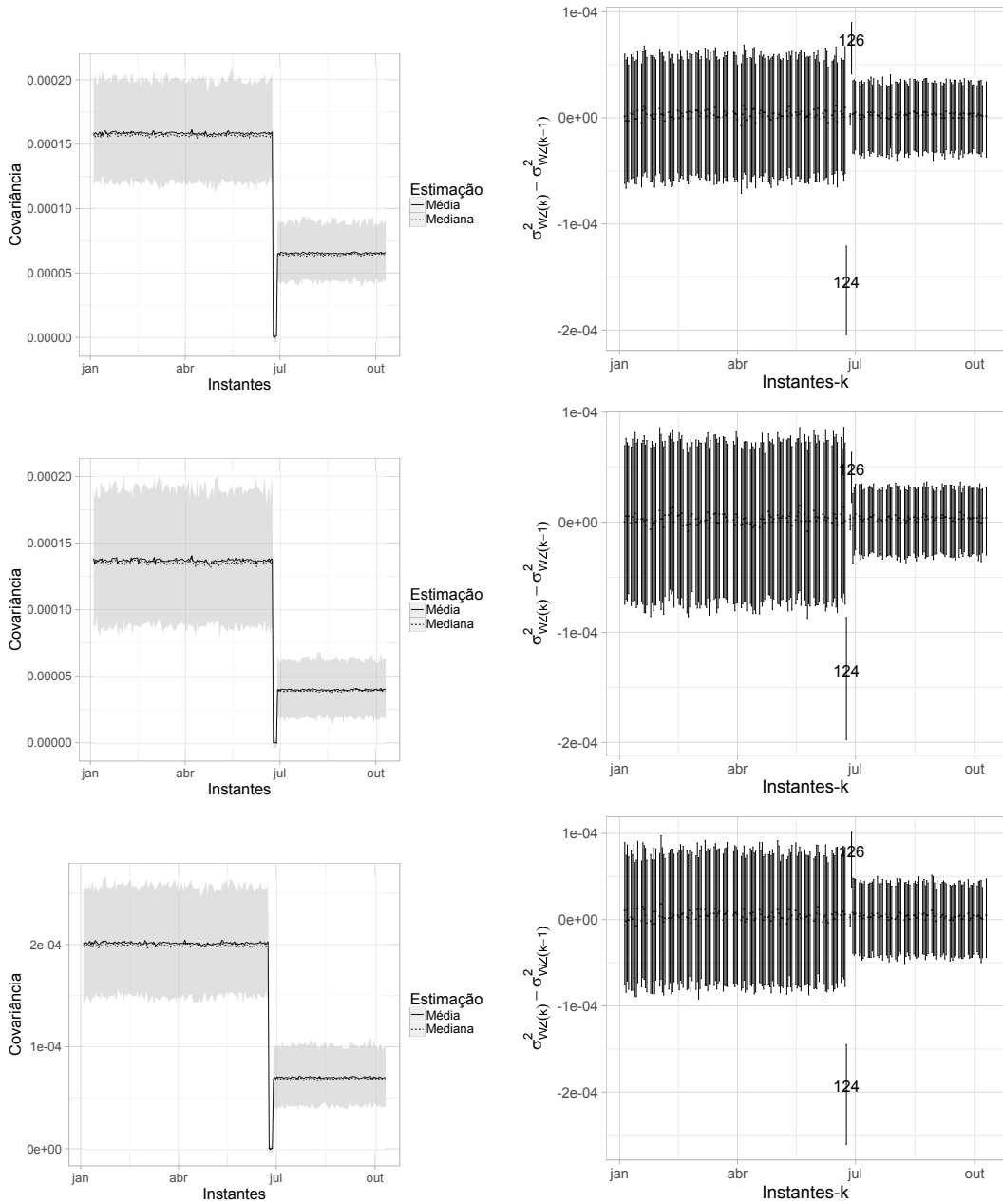


Figura 4.13: Estimativas de $\sigma_{WZ(k)}^2$ (esquerda) e intervalos HPD's de 95% para $\sigma_{(WZ)k}^2 - \sigma_{WZ(k-1)}^2$ (direita) entre os índices Reino Unido e França, Reino Unido e Grécia e Reino Unido e Itália.

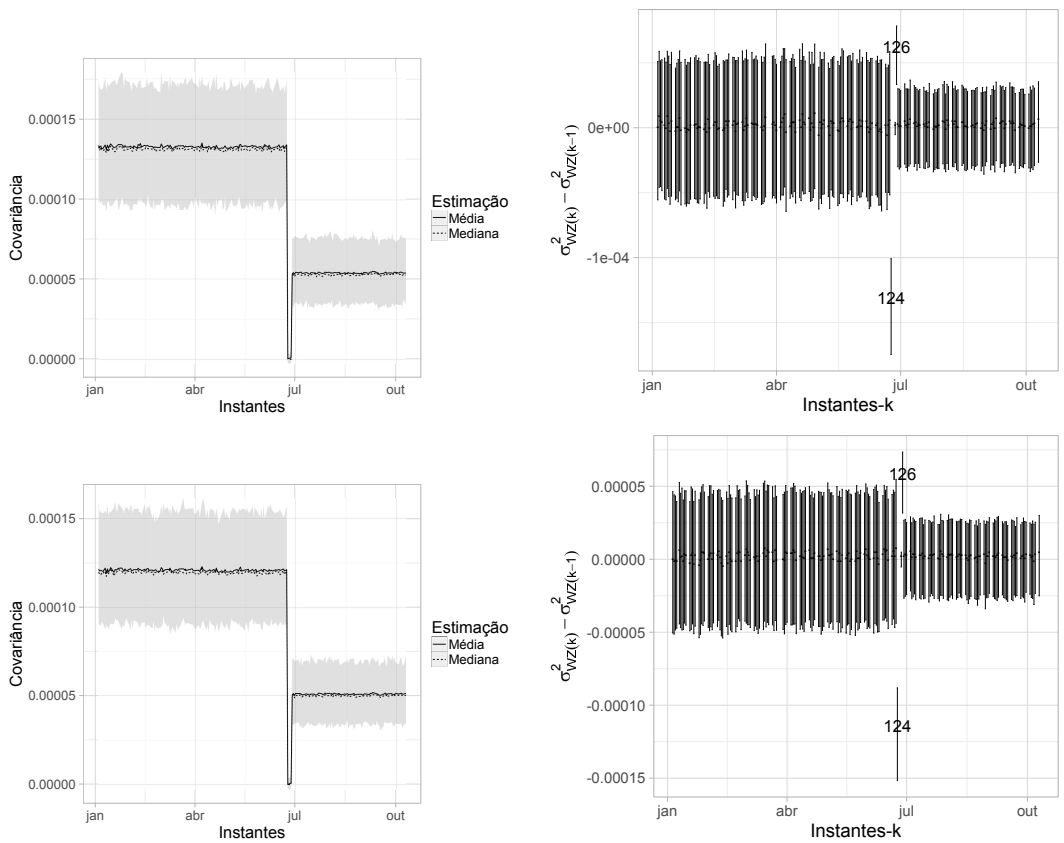


Figura 4.14: Estimativas de $\sigma_{WZ(k)}^2$ (esquerda) e intervalos HPD's de 95% para $\sigma_{(WZ)k}^2 - \sigma_{WZ(k-1)}^2$ (direita) entre os índices Reino Unido e Portugal e Reino Unido e Suíça.

Capítulo 5

Conclusão

Neste trabalho, considerou-se o modelo partição produto para detectar múltiplas mudanças no vetor de médias e na matriz de variância e covariância de uma sequência de dados com distribuição normal multivariada. Estendeu-se os resultados em Moura [33] por assumir o vetor de médias não nulo.

Diferentemente de Cheon & Kim [8], foram utilizadas distribuições *a priori* conjugadas para $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ tornando o modelo mais geral. Também foi proposto um procedimento para avaliar em quais parâmetros a mudança foi significativa, baseando-se em comparações dos parâmetros em instantes sucessivos do tempo.

A partir das análises de sensibilidade realizada no Capítulo 3, notou-se como as especificações *a priori* influenciam nas estimações. Do cenário 1, pequena mudança apenas na média de uma das séries, o modelo mostrou-se bastante satisfatório na identificação da partição correta dos dados, atribuindo alta probabilidade para os instantes reais de mudança e consequentemente estimando corretamente o número de blocos para as situações 1 e 2. Também percebeu-se o efeito das escolhas *a priori* de v , d e \mathbf{D} nas estimações. Foi constatado que quanto mais distante for o valor de \mathbf{D} da matriz de covariância real, o modelo tende a identificar partições do tipo $\rho = \{0, 100\}$. Também pôde-se notar que mesmo para escolha de \mathbf{D} igual a matriz de covariância real, os valores de v e d também podem influenciar na identificação das partições quando ambos apresentam valores altos. De forma geral, a inferência para o vetor de médias e para a matriz de covariância também foi bastante satisfatória, sendo identificado como ponto de mudança significativa apenas o instante real relacionado ao vetor de médias. No entanto, constatou-se que a combinação de v , d e \mathbf{D} influenciam de maneiras diferentes os parâmetros, isto é, com uma combinação de v e d o modelo pode identificar a partição correta com alta probabilidade, porém as estimativas para $\boldsymbol{\mu}_k$ e $\boldsymbol{\Sigma}_k$ podem não ser boas. De forma geral, percebeu-se que os modelos com as especificações $\mathbf{D} = \begin{pmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{pmatrix}$,

$v = 0.01$, $d = 4$ e $\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $v = 0.01$ e $d = 4$ foram os que apresentaram melhores resultados conjuntamente para os parâmetros.

Do cenário 2, mudança na variância de apenas em uma das séries, também pode-se observar o efeito das combinações de v , d e \mathbf{D} . Notou-se que o modelo identificou uma partição próxima da correta, identificando o ponto de mudança com atraso. Identificou de maneira correta o número de bloco na maioria das combinações de v , d e \mathbf{D} . Assim como para o cenário 1, a probabilidade de mudança em um instante qualquer também foi muito baixa. Constatou-se, assim como no cenário 1, o efeito que a escolha de v exerce sobre a estimação de $\boldsymbol{\mu}_k$ e $\boldsymbol{\Sigma}_k$, ressaltando que tal efeito sobre $\boldsymbol{\Sigma}_k$ é devido ao fato de que o parâmetro da distribuição *a posteriori* depender do vetor de médias amostrais, sendo o valor de v uma espécie de peso sobre tal vetores. Para este cenário, a inferência sobre $\boldsymbol{\mu}_k$ e $\boldsymbol{\Sigma}$ também foi bastante satisfatória, identificando uma mudança significativa apenas entre os instantes 50 e 51 da variância da série 2. De forma geral, percebeu-se que o modelos com as especificações $\mathbf{D} = \begin{pmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{pmatrix}$, $v = 0.01$ e $d = 4$ foi o que apresentou um melhor resultado conjuntamente.

No cenário 3, o modelo identificou dez partições diferentes, atribuindo alta probabilidade para a correta. Também notou-se que os instantes reais de mudanças foram identificados com probabilidade 1. O número correto de blocos também foi identificado pelo modelo com alta probabilidade. Para estimação do vetor de médias, o modelo foi bastante satisfatório. No entanto, para a matriz de variância e covariância o modelo superestimou ou subestimou em alguns instantes. Isso pode ter ocorrido devido a escolha de \mathbf{D} . A partir da realização dos testes sequencias realizados no cenário 3, pôde-se observar que os ponto de mudanças agem de forma diferente nos parâmetros $\boldsymbol{\mu}_k$ e $\boldsymbol{\Sigma}$, isto é, um determinado instante que é ponto de mudança na média pode não ser na matriz de variância e covariância.

A partir da aplicação, pôde-se notar que o modelo apontou de forma satisfatória o impacto econômico do *Brexit* na economia dos países envolvidos no estudo. O modelo identificou mudanças ocorridas próximas ao dia 23 de junho, data em que houve a votação que decidiu a saída do Reino Unido da União Europeia. Pôde-se notar que tal impacto influenciou de maneira diferente o retorno médio e a relação existente entre os índices. Para alguns índices, o dia 24/6, por exemplo, identificou uma queda e para outros um crescimento. Percebeu-se também, através dos testes sequenciais, que o dia 27/06 foi mudança significativa no vetor de médias, porém não foi na matriz de variância e covariância.

Em relação as comparações sequenciais, pôde-se perceber que foi uma alternativa satisfatória para auxiliar na identificação da mudanças no vetor de médias ou na

matriz de variância e covariância, pois possibilita analisar de maneira mais detalhada a origem dos pontos de mudanças identificadas pelo modelo

Para trabalhos futuros, pretende-se calcular a distribuição da diferença *a posteriori* entre $\theta_k - \theta_k$ e estender os resultados obtidos nesse trabalho considerando a hipótese das médias correlacionadas no tempo.

Apêndice A

Distribuições

Neste apêndice são apresentadas as funções densidades de probabilidade e algumas propriedades das distribuições utilizadas na construção do Modelo Partição Produto.

Distribuição Normal Multivariada

Considere o vetor aleatório $\mathbf{X} = (X_1, \dots, X_p)^t$ e $\mathbf{x} = (x_1, \dots, x_p)^t$ uma observação associada a esse vetor. Então, tem-se que \mathbf{X} possui uma distribuição normal p-variada, denotada por $\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N_p(\boldsymbol{\mu}_{p \times 1}, \boldsymbol{\Sigma}_{p \times p})$, onde $\boldsymbol{\mu}_{p \times 1}$ é o vetor de médias de dimensão $p \times 1$ e $\boldsymbol{\Sigma}_{p \times p}$ é a matriz de variância-covariância de dimensão $p \times p$, se sua função densidade é dada por

$$f_{\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{x}) = (2\pi)^{-(p/2)} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}. \quad (\text{A.1})$$

Neste caso, tem-se que a média e a variância são dadas, respectivamente, por

$$\begin{aligned} E(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \boldsymbol{\mu} \\ \text{Var}(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \boldsymbol{\Sigma}. \end{aligned}$$

Para mais detalhes sobre a distribuição normal p-variada, ver Johnson *et al.* [16].

Distribuição t-Student matriz variada

Sejam \mathbf{X} uma matriz de dimensão $n \times p$, $\boldsymbol{\Omega}$ uma matriz positiva definida de dimensão $n \times n$, $\boldsymbol{\Sigma}$ uma matriz positiva definida de dimensão $p \times p$, \mathbf{M} uma matriz de locação de dimensão $n \times p$ e ν o grau de liberdade. Diz-se que \mathbf{X} possui uma distribuição t-Student matriz-variada, denotada por $\mathbf{X}|\mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \nu \sim \mathcal{T}_{n \times p}(\nu, \mathbf{M}, \boldsymbol{\Omega}, \boldsymbol{\Sigma})$, se a sua

função de densidade é dada por

$$f_{\mathbf{X}|\mathbf{M},\mathbf{\Sigma},\mathbf{\Omega},\nu}(\mathbf{X}) = \frac{\Gamma_p\left(\frac{\nu+n+p-1}{2}\right) |\mathbf{\Omega}|^{-p/2} |\mathbf{\Sigma}|^{-n/2}}{\Gamma_p\left(\frac{\nu+p-1}{2}\right) (\pi)^{np/2}} \times |\mathbb{I}_n + \mathbf{\Omega}^{-1}(\mathbf{X} - \mathbf{M})\mathbf{\Sigma}^{-1}(\mathbf{X} - \mathbf{M})^t|^{-\frac{\nu+n+p-1}{2}}.$$

O valor esperado e a variância de \mathbf{X} são dadas, respectivamente por,

$$E(\mathbf{X}|\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Omega}, \nu) = \mathbf{M}$$

$$Var(\mathbf{X}|\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Omega}, \nu) = \frac{\mathbf{\Omega} \otimes \mathbf{\Sigma}}{\nu - 2} \quad \text{se } \nu > 2,$$

onde \otimes denota o produto de Kronecker. Para mais detalhes, ver Kotz & Nadarajah [18] e Gupta & Nagar [11].

Distribuição t-Student multivariada

Diz-se que \mathbf{X} possui uma distribuição t-Multivariada, denotada por $\mathbf{X} \sim t_\nu(\boldsymbol{\mu}, \mathbf{\Sigma})$, se sua função densidade é dada por

$$f_{\mathbf{X}|\boldsymbol{\mu},\mathbf{\Sigma},\nu}(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+p}{2}\right) |\mathbf{\Sigma}|^{-1/2} \left[1 + \frac{1}{\nu}(\mathbf{x} - \boldsymbol{\mu})^t \mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]^{-\frac{\nu+p}{2}}}{\Gamma\left(\frac{\nu}{2}\right) (\nu\pi)^{p/2}},$$

onde $\boldsymbol{\mu}$ é o vetor de locação de ordem $p \times 1$, $\mathbf{\Sigma}$ é a matriz $p \times p$, simétrica e positiva definida e ν é o grau de liberdade. Então, tem-se que o vetor de média e a matriz de covariância de \mathbf{X} são dadas, respectivamente, por

$$E(\mathbf{X}|\boldsymbol{\mu}, \mathbf{\Sigma}, \nu) = \boldsymbol{\mu},$$

$$Var(\mathbf{X}|\boldsymbol{\mu}, \mathbf{\Sigma}, \nu) = \frac{\nu}{\nu - 2} \mathbf{\Sigma} \quad \text{se, } \nu > 2.$$

Para mais detalhes sobre a distribuição t-Student p-variada, ver Kotz & Nadarajah [18].

Distribuição Wishart-inversa

Seja \mathbf{X} uma matriz aleatória de dimensão $p \times p$, simétrica e positiva definida. Considere \mathbf{D} uma matriz positiva definida e um escalar $d \geq p$. Então, tem-se que \mathbf{X} possui uma distribuição Wishart Invertida, denotada por $\mathbf{X}|\mathbf{D}, d \sim WI(\mathbf{D}, d)$, se

sua função de densidade é dada por

$$f_{\mathbf{X}|\mathbf{D},d}(\mathbf{X}) = \frac{|\mathbf{D}|^{d/2} |\mathbf{x}|^{-\frac{d+p+1}{2}} \exp \left\{ \frac{\text{tr}(\mathbf{X}^{-1}\mathbf{D})}{2} \right\}}{2^{dp/2} \Gamma_p \left(\frac{d}{2} \right)},$$

onde $\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma \left(a + \frac{1-i}{2} \right)$ e tr é denota a função traço. A esperança e moda de \mathbf{X} são dadas, respectivamente por

$$E(\mathbf{X}|\mathbf{D}, d) = \frac{\mathbf{D}}{d-p-1}, \quad \text{se } d > (p+1),$$

$$\text{Moda}(\mathbf{X}|\mathbf{D}, d) = \frac{\mathbf{D}}{d+p+1}.$$

A variância de cada elemento de \mathbf{x} é dada por

$$\text{Var}(x_{ij}) = \frac{(\nu-p+1)D_{ij}^2 + (\nu-p-1)D_{ii}D_{jj}}{(\nu-p)(\nu-p-1)^2(\nu-p-3)}, \quad \text{se } \nu > p+3.$$

Para mais detalhes sobre a distribuição Wishart invertida, ver O'Hagan & Forster [35] e Gupta & Nagar [11].

Apêndice B

Demonstração da Proposição 3.1.1

Neste apêndice, apresenta-se a prova dos resultados exibidos na Proposição 2.1.1 do Capítulo 2. Inicia-se desenvolvendo a função de verossimilhança de forma que seja mais conveniente os cálculos das distribuições *a posteriori* e preditiva *a priori*.

Cálculo da verossimilhança

Como $\mathbf{y}_1, \dots, \mathbf{y}_n \stackrel{ind}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a função verossimilhança é dada por

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}_1, \dots, \mathbf{y}_n) &= \prod_{k=1}^n (2\pi)^{(-p/2)} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_k - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{y}_k - \boldsymbol{\mu}) \right\} \\ &= (2\pi)^{(-np/2)} |\boldsymbol{\Sigma}|^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{k=1}^n (\mathbf{y}_k - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{y}_k - \boldsymbol{\mu}) \right\} \end{aligned} \quad (\text{B.1})$$

Denote por $\mathbf{S} = \frac{1}{n-1} \sum_{k=1}^n (\mathbf{y}_k - \bar{\mathbf{y}})(\mathbf{y}_k - \bar{\mathbf{y}})^t$ a matriz de covariância amostral de ordem $p \times p$ e $\bar{\mathbf{y}} = \frac{1}{n} \sum_{k=1}^n \mathbf{y}_k$ o vetor de médias amostrais de ordem $p \times 1$.

Considerando as propriedades 1 e 2 do Apêndice C, temos que

$$\begin{aligned}
\sum_{k=1}^n (\mathbf{y}_k - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{y}_k - \boldsymbol{\mu}) &= \sum_{k=1}^n \text{tr} \left[\boldsymbol{\Sigma}^{-1} (\mathbf{y}_k - \boldsymbol{\mu}) (\mathbf{y}_k - \boldsymbol{\mu})^t \right] \\
&= \text{tr} \left[\boldsymbol{\Sigma}^{-1} \sum_{k=1}^n (\mathbf{y}_k - \boldsymbol{\mu}) (\mathbf{y}_k - \boldsymbol{\mu})^t \right] \\
&= \text{tr} \left[\boldsymbol{\Sigma}^{-1} \sum_{k=1}^n [((\mathbf{y}_k - \bar{\mathbf{y}}) + (\bar{\mathbf{y}} - \boldsymbol{\mu})) ((\mathbf{y}_k - \bar{\mathbf{y}}) + (\bar{\mathbf{y}} - \boldsymbol{\mu}))^t] \right] \\
&= \text{tr} \left[\boldsymbol{\Sigma}^{-1} \left[\sum_{k=1}^n (\mathbf{y}_k - \bar{\mathbf{y}}) (\mathbf{y}_k - \bar{\mathbf{y}})^t + n (\bar{\mathbf{y}} - \boldsymbol{\mu}) (\bar{\mathbf{y}} - \boldsymbol{\mu})^t \right] \right] \\
&= \text{tr} \left[\boldsymbol{\Sigma}^{-1} [(n-1)\mathbf{S} + n(\bar{\mathbf{y}} - \boldsymbol{\mu})(\bar{\mathbf{y}} - \boldsymbol{\mu})^t] \right]
\end{aligned} \tag{B.2}$$

Substituindo o resultado obtido em (B.2) na expressão (B.1), segue que a função de verossimilhança assume a seguinte forma:

$$\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}_1, \dots, \mathbf{y}_n) &= (2\pi)^{(-np/2)} |\boldsymbol{\Sigma}|^{-n/2} \\
&\quad \times \exp \left\{ -\frac{1}{2} [(n-1) \text{tr} [\boldsymbol{\Sigma}^{-1} \mathbf{S}] + n(\bar{\mathbf{y}} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu})] \right\}.
\end{aligned} \tag{B.3}$$

Cálculo da distribuição *a posteriori*

Assumindo as distribuições *a priori* da Proposição 2.1.1 e considerando a função de verossimilhança em (B.3), pelo teorema de Bayes temos que a distribuição *a posteriori* conjunta de $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ é

$$\begin{aligned}
f(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}_1, \dots, \mathbf{y}_n) &\propto L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}_1, \dots, \mathbf{y}_n) f(\boldsymbol{\mu} | \boldsymbol{\Sigma}) f(\boldsymbol{\Sigma}) \\
&\propto |\boldsymbol{\Sigma}|^{(-n/2)} \exp \left\{ -\frac{1}{2} [(n-1) \text{tr} (\boldsymbol{\Sigma}^{-1} \mathbf{S}) + n(\bar{\mathbf{y}} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu})] \right\} |\boldsymbol{\Sigma}|^{-\frac{d+p+1}{2}} \\
&\quad \times \exp \left\{ -\frac{1}{2} \text{tr} (\boldsymbol{\Sigma}^{-1} \mathbf{D}) \right\} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} v (\boldsymbol{\mu} - \mathbf{m})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \mathbf{m}) \right\} \\
&\propto |\boldsymbol{\Sigma}|^{-\frac{d+n+p+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\boldsymbol{\Sigma}^{-1} ((n-1)\mathbf{S} + \mathbf{D})] \right\} \\
&\quad \times |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} [\text{tr} [\boldsymbol{\Sigma}^{-1} (n(\bar{\mathbf{y}} - \boldsymbol{\mu})(\bar{\mathbf{y}} - \boldsymbol{\mu})^t + v(\boldsymbol{\mu} - \mathbf{m})(\boldsymbol{\mu} - \mathbf{m})^t)]] \right\}.
\end{aligned}$$

Note que

$$\begin{aligned}
(n(\bar{\mathbf{y}} - \boldsymbol{\mu})(\bar{\mathbf{y}} - \boldsymbol{\mu})^t + v(\boldsymbol{\mu} - \mathbf{m})(\boldsymbol{\mu} - \mathbf{m})^t) &= (n+v)\boldsymbol{\mu}\boldsymbol{\mu}^t - \boldsymbol{\mu}(n\bar{\mathbf{y}}^t + v\mathbf{m}^t) - (n\bar{\mathbf{y}} + v\mathbf{m})\boldsymbol{\mu}^t \\
&\quad + n\bar{\mathbf{y}}\bar{\mathbf{y}}^t + v\mathbf{m}\mathbf{m}^t \\
&= (n+v) [\boldsymbol{\mu}\boldsymbol{\mu}^t - \boldsymbol{\mu}(n\bar{\mathbf{y}}^t + v\mathbf{m}^t)(n+v)^{-1} \\
&\quad - (n\bar{\mathbf{y}} + v\mathbf{m})\boldsymbol{\mu}^t(n+v)^{-1} \pm \mathbf{M}\mathbf{M}^t] + n\bar{\mathbf{y}}\bar{\mathbf{y}}^t + v\mathbf{m}\mathbf{m}^t \\
&= (n+v) [(\boldsymbol{\mu} - \mathbf{M})(\boldsymbol{\mu} - \mathbf{M})^t - \mathbf{M}\mathbf{M}^t] + n\bar{\mathbf{y}}\bar{\mathbf{y}}^t + v\mathbf{m}\mathbf{m}^t.
\end{aligned}$$

Consequentemente, segue que

$$\begin{aligned}
f(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}_1, \dots, \mathbf{y}_n) &\propto |\boldsymbol{\Sigma}|^{-\frac{d+n+p+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} (\mathbf{D} + (n-1)\mathbf{S} + n\bar{\mathbf{y}}\bar{\mathbf{y}}^t + v\mathbf{m}\mathbf{m}^t - (n+v)\mathbf{M}\mathbf{M}^t) \right] \right\} \\
&\quad \times |(n+v)^{-1}\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (n+v) (\boldsymbol{\mu} - \mathbf{M})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \mathbf{M}) \right\},
\end{aligned}$$

onde $\mathbf{M} = (n+v)^{-1}(n\bar{\mathbf{y}} + v\mathbf{m})$. Observe que

$$n\bar{\mathbf{y}}\bar{\mathbf{y}}^t + v\mathbf{m}\mathbf{m}^t - (n+v)\mathbf{M}\mathbf{M}^t = \frac{nv}{n+v} [\bar{\mathbf{y}}\bar{\mathbf{y}}^t + \mathbf{m}\mathbf{m}^t - \bar{\mathbf{y}}\mathbf{m}^t - \mathbf{m}\bar{\mathbf{y}}^t] = \frac{nv}{n+v} [(\bar{\mathbf{y}} - \mathbf{m})(\bar{\mathbf{y}} - \mathbf{m})^t].$$

Consequentemente, segue que

$$\begin{aligned}
f(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}_1, \dots, \mathbf{y}_n) &\propto |\boldsymbol{\Sigma}|^{-\frac{d+n+p+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \left(\mathbf{D} + (n-1)\mathbf{S} + \frac{nv}{n+v} [(\bar{\mathbf{y}} - \mathbf{m})(\bar{\mathbf{y}} - \mathbf{m})^t] \right) \right] \right\} \\
&\quad \times |(n+v)^{-1}\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu} - \mathbf{M})^t ((n+v)^{-1}\boldsymbol{\Sigma})^{-1} (\boldsymbol{\mu} - \mathbf{M}) \right\}.
\end{aligned} \tag{B.4}$$

Portanto, segue de (B.4) que

$$\boldsymbol{\mu} | \boldsymbol{\Sigma}, \mathbf{y}_1, \dots, \mathbf{y}_n \sim N_p(\mathbf{M}, (n+v)^{-1}\boldsymbol{\Sigma})$$

e

$$\boldsymbol{\Sigma} | \mathbf{y}_1, \dots, \mathbf{y}_n \sim IW(\mathbf{D}^*, d^*),$$

$$\text{onde } \mathbf{D}^* = \mathbf{D} + (n-1)\mathbf{S} + \frac{nv}{n+v} [(\bar{\mathbf{y}} - \mathbf{m})(\bar{\mathbf{y}} - \mathbf{m})^t] \quad \text{e} \quad d^* = d + n.$$

Cálculo da distribuição *a posteriori* $\boldsymbol{\mu}$

Considere as distribuições *a posteriori* de $\boldsymbol{\mu} | \boldsymbol{\Sigma}, \mathbf{y}_1, \dots, \mathbf{y}_n$ e $\boldsymbol{\Sigma} | \mathbf{y}_1, \dots, \mathbf{y}_n$ obtidas em B.4. Então, segue que a distribuição *a posteriori* de $\boldsymbol{\mu}$ é dada por

$$\begin{aligned}
f(\boldsymbol{\mu} \mid \mathbf{y}_1, \dots, \mathbf{y}_n) &= \int f(\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \mathbf{y}_1, \dots, \mathbf{y}_n) f(\boldsymbol{\Sigma} \mid \mathbf{y}_1, \dots, \mathbf{y}_n) d\boldsymbol{\Sigma} \\
&= \frac{(2\pi)^{-p/2} (n+v)^{p/2} |\mathbf{D}^*|^{d^*/2}}{2^{d^*p/2} \Gamma_p\left(\frac{d^*}{2}\right)} \int |\boldsymbol{\Sigma}|^{-\frac{d^*+1+p+1}{2}} \\
&\quad \times \exp\left\{-\frac{1}{2} \text{tr}\left[\boldsymbol{\Sigma}^{-1} (\mathbf{D}^* + (n+v)(\boldsymbol{\mu} - \mathbf{M})(\boldsymbol{\mu} - \mathbf{M})^t)\right]\right\} d\boldsymbol{\Sigma}.
\end{aligned}$$

Completando os termos faltantes da distribuição $WI(\mathbf{D}^* + (n+v)(\boldsymbol{\mu} - \mathbf{M})(\boldsymbol{\mu} - \mathbf{M})^t, d+n+1)$ e cancelando os termos semelhantes, segue que

$$f(\boldsymbol{\mu} \mid \mathbf{y}_1, \dots, \mathbf{y}_n) = \frac{\Gamma_p\left(\frac{d+n+1}{2}\right)}{\Gamma_p\left(\frac{d+n}{2}\right) \pi^{p/2}} (n+v)^{p/2} |\mathbf{D}^*|^{(d+n)/2} |\mathbf{D}^* + (n+v)(\boldsymbol{\mu} - \mathbf{M})(\boldsymbol{\mu} - \mathbf{M})^t|^{-\frac{d+n+1}{2}}$$

Note que

$$\frac{\Gamma_p\left(\frac{d+n+1}{2}\right)}{\Gamma_p\left(\frac{d+n}{2}\right)} = \frac{\pi^{\frac{p(p-1)}{4}} \prod_{j=1}^p \Gamma\left(\frac{d+n+1+1-j}{2}\right)}{\pi^{\frac{p(p-1)}{4}} \prod_{j=1}^p \Gamma\left(\frac{d+n+1-j}{2}\right)} = \frac{\Gamma\left(\frac{d+n+1}{2}\right)}{\Gamma\left(\frac{d+n+1-p}{2}\right)}.$$

Utilizando a Propriedade 4, Apêndice C, em $|\mathbf{D}^* + (n+v)(\boldsymbol{\mu} - \mathbf{M})(\boldsymbol{\mu} - \mathbf{M})^t|$, segue que

$$\begin{aligned}
f(\boldsymbol{\mu} \mid \mathbf{y}_1, \dots, \mathbf{y}_n) &= \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{(\pi\nu)^{p/2}} |\nu^{-1}(n+v)^{-1}\mathbf{D}^*|^{-1/2} \\
&\quad \times \left(1 + \frac{1}{\nu}(\boldsymbol{\mu} - \mathbf{M})^t(\nu^{-1}(n+v)^{-1}\mathbf{D}^*)^{-1}(\boldsymbol{\mu} - \mathbf{M})\right)^{-\frac{\nu+p}{2}}. \quad (\text{B.5})
\end{aligned}$$

Portanto, segue de B.5 que $\boldsymbol{\mu} \mid \mathbf{y}_1, \dots, \mathbf{y}_n \sim t_\nu(\mathbf{M}, \nu^{-1}(n+v)^{-1}\mathbf{D}^*)$ onde $\nu = d+n+1-p$.

Cálculo da Preditiva *a priori*

Considere a função de verossimilhança em (2.1) e as distribuições *a priori* da Proposição 2.1.1. Então, segue que a distribuição preditiva *a priori* de \mathbf{Y} é dada por

$$\begin{aligned}
f(\mathbf{Y}) &= \int \int L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{Y}) f(\boldsymbol{\mu} | \boldsymbol{\Sigma}) f(\boldsymbol{\Sigma}) d\boldsymbol{\mu} d\boldsymbol{\Sigma} \\
&= \frac{(2\pi)^{-np/2} |\mathbf{D}|^{d/2} v^{p/2}}{2^{dp/2} \Gamma_p\left(\frac{d}{2}\right)} \int |\boldsymbol{\Sigma}|^{-\frac{d+n+p+1}{2}} \exp\left\{-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{D})\right\} \left[\int (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \right. \\
&\quad \left. \times \exp\left\{-\frac{1}{2} \text{tr}\left(\boldsymbol{\Sigma}^{-1} \left[(\mathbf{Y} - \mathbf{1}\boldsymbol{\mu}^t)^t (\mathbf{Y} - \mathbf{1}\boldsymbol{\mu}^t) + v(\boldsymbol{\mu} - \mathbf{m})(\boldsymbol{\mu} - \mathbf{m})^t\right]\right)\right\} d\boldsymbol{\mu} \right] d\boldsymbol{\Sigma}.
\end{aligned}$$

Denote por $\mathbf{B} = (n + v)^{-1} (\mathbf{Y}^t \mathbf{1} + v\mathbf{m})$ e note que

$$\begin{aligned}
(\mathbf{Y} - \mathbf{1}\boldsymbol{\mu}^t)^t (\mathbf{Y} - \mathbf{1}\boldsymbol{\mu}^t) + v(\boldsymbol{\mu} - \mathbf{m})(\boldsymbol{\mu} - \mathbf{m})^t &= (n + v)\boldsymbol{\mu}\boldsymbol{\mu}^t - (\mathbf{Y}^t \mathbf{1} + v\mathbf{m})\boldsymbol{\mu}^t - \boldsymbol{\mu}(\mathbf{1}^t \mathbf{Y} + v\mathbf{m}^t) \\
&\quad + \mathbf{Y}^t \mathbf{Y} + v\mathbf{m}\mathbf{m}^t \\
&= (n + v) [\boldsymbol{\mu}\boldsymbol{\mu}^t - (\mathbf{Y}^t \mathbf{1} + v\mathbf{m})\boldsymbol{\mu}^t (n + v)^{-1} \\
&\quad - \boldsymbol{\mu}(\mathbf{1}^t \mathbf{Y} + v\mathbf{m}^t)(n + v)^{-1} \pm \mathbf{B}\mathbf{B}^t] \\
&\quad + \mathbf{Y}^t \mathbf{Y} + v\mathbf{m}\mathbf{m}^t \\
&= (n + v) [(\boldsymbol{\mu} - \mathbf{B})(\boldsymbol{\mu} - \mathbf{B})^t - \mathbf{B}\mathbf{B}^t] + \mathbf{Y}^t \mathbf{Y} + v\mathbf{m}\mathbf{m}^t.
\end{aligned}$$

Consequentemente, segue que

$$\begin{aligned}
f(\mathbf{Y}) &= \frac{(2\pi)^{-np/2} |\mathbf{D}|^{d/2} v^{p/2}}{2^{dp/2} \Gamma_p\left(\frac{d}{2}\right)} \\
&\quad \times \int |\boldsymbol{\Sigma}|^{-\frac{d+n+p+1}{2}} \exp\left\{-\frac{1}{2} \text{tr}\left(\boldsymbol{\Sigma}^{-1} [\mathbf{D} + \mathbf{Y}^t \mathbf{Y} + v\mathbf{m}\mathbf{m}^t - (n + v)\mathbf{B}\mathbf{B}^t]\right)\right\} \\
&\quad \times \left[\int (2\pi)^{-p/2} (n + v)^{-1} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2} (\boldsymbol{\mu} - \mathbf{B})^t ((n + v)^{-1} \boldsymbol{\Sigma})^{-1} (\boldsymbol{\mu} - \mathbf{B})\right\} d\boldsymbol{\mu} \right] \\
&\quad \left(\frac{1}{n + v}\right)^{p/2} d\boldsymbol{\Sigma}.
\end{aligned}$$

Completando os termos faltantes da distribuição $WI(\mathbf{D} + \mathbf{Y}^t \mathbf{Y} + v\mathbf{m}\mathbf{m}^t - (n + v)\mathbf{B}\mathbf{B}^t, d + n)$, logo a integral é 1. Note que $\boldsymbol{\mu} \sim N_p(\mathbf{B}, (n + v)^{-1} \boldsymbol{\Sigma})$, logo a integral também é 1. Então, segue que

$$f(\mathbf{Y}) = \frac{(2\pi)^{-np/2} |\mathbf{D}|^{d/2}}{2^{dp/2} \Gamma_p\left(\frac{d}{2}\right)} \left(\frac{v}{n + v}\right)^{p/2} \Gamma_p\left(\frac{d + n}{2}\right) 2^{\frac{(d+n)p}{2}} |\mathbf{D} + \mathbf{Y}^t \mathbf{Y} + v\mathbf{m}\mathbf{m}^t - (n + v)\mathbf{B}\mathbf{B}^t|^{-\frac{d+n}{2}}.$$

Observe que

$$\begin{aligned}
\mathbf{Y}^t \mathbf{Y} + v \mathbf{m} \mathbf{m}^t - (n+v) \mathbf{B} \mathbf{B}^t &= \mathbf{Y}^t \left[\mathbb{I}_n - \frac{\mathbf{1} \mathbf{1}^t}{n+v} \right] \mathbf{Y} - \frac{\mathbf{Y}^t \mathbf{1} \mathbf{m}^t v}{n+v} - \frac{v \mathbf{m} \mathbf{1}^t \mathbf{Y}}{n+v} + \frac{nv \mathbf{m} \mathbf{m}^t}{n+v} \\
&= \mathbf{Y}^t \mathbf{A} \mathbf{Y} - \mathbf{Y}^t \mathbf{A} \mathbf{A}^{-1} \mathbf{Z} - \mathbf{Z}^t \mathbf{A}^{-1} \mathbf{A} \mathbf{Y} \pm (\mathbf{A}^{-1} \mathbf{Z})^t \mathbf{A} \mathbf{A}^{-1} \mathbf{Z} + \frac{nv \mathbf{m} \mathbf{m}^t}{n+v} \\
&= (\mathbf{Y} - \mathbf{A}^{-1} \mathbf{Z})^t \mathbf{A} (\mathbf{Y} - \mathbf{A}^{-1} \mathbf{Z}) - (\mathbf{A}^{-1} \mathbf{Z})^t \mathbf{A} \mathbf{A}^{-1} \mathbf{Z} + \frac{nv \mathbf{m} \mathbf{m}^t}{n+v} \\
&= (\mathbf{Y} - \mathbf{1} \mathbf{m}^t)^t \left[\mathbb{I}_n - \frac{\mathbf{1} \mathbf{1}^t}{v} \right] (\mathbf{Y} - \mathbf{1} \mathbf{m}^t),
\end{aligned}$$

onde $\mathbf{A} = \left[\mathbb{I}_n - \frac{\mathbf{1} \mathbf{1}^t}{n+v} \right]$ e $\mathbf{Z} = \frac{\mathbf{1} \mathbf{m}^t v}{n+v}$. Utilizando a Propriedade 8, Apêndice C, segue que $\mathbf{A}^{-1} = \left[\mathbb{I}_n - \frac{\mathbf{1} \mathbf{1}^t}{v} \right]$. Portanto, $\mathbf{A}^{-1} \mathbf{Z} = \mathbf{1} \mathbf{m}^t$ e $-(\mathbf{A}^{-1} \mathbf{Z})^t \mathbf{A} \mathbf{A}^{-1} \mathbf{Z} + \frac{nv \mathbf{m} \mathbf{m}^t}{n+v} = 0$. Consequentemente, segue que

$$\begin{aligned}
f(\mathbf{Y}) &= \frac{\Gamma_p \left(\frac{d+n}{2} \right)}{\Gamma_p \left(\frac{d}{2} \right) \pi^{np/2}} \left(\frac{v}{n+v} \right)^{p/2} |\mathbf{D}|^{d/2} \left| \mathbf{D} + (\mathbf{Y} - \mathbf{1} \mathbf{m}^t)^t \left[\mathbb{I}_n - \frac{\mathbf{1} \mathbf{1}^t}{n+v} \right] (\mathbf{Y} - \mathbf{1} \mathbf{m}^t) \right|^{-\frac{d+n}{2}} \\
&= \frac{\Gamma_p \left(\frac{d+n}{2} \right)}{\Gamma_p \left(\frac{d}{2} \right) \pi^{np/2}} \left| \mathbb{I}_n + \frac{\mathbf{1} \mathbf{1}^t}{v} \right|^{-p/2} |\mathbf{D}|^{-n/2} \left| \mathbb{I}_p + \mathbf{D}^{-1} (\mathbf{Y} - \mathbf{1} \mathbf{m}^t)^t \left[\mathbb{I}_n - \frac{\mathbf{1} \mathbf{1}^t}{n+v} \right] (\mathbf{Y} - \mathbf{1} \mathbf{m}^t) \right|^{-\frac{d+n}{2}}.
\end{aligned}$$

Utilizando a Propriedade 7, Apêndice C, em

$$\left| \mathbb{I}_p + \mathbf{D}^{-1} (\mathbf{Y} - \mathbf{1} \mathbf{m}^t)^t \left[\mathbb{I}_n - \frac{\mathbf{1} \mathbf{1}^t}{n+v} \right] (\mathbf{Y} - \mathbf{1} \mathbf{m}^t) \right|,$$

segue que

$$\begin{aligned}
f(\mathbf{Y}) &= \frac{\Gamma_p \left(\frac{d+n}{2} \right)}{\Gamma_p \left(\frac{d}{2} \right) \pi^{np/2}} \left| \mathbb{I}_n + \frac{\mathbf{1} \mathbf{1}^t}{v} \right|^{-p/2} |\mathbf{D}|^{-n/2} \left| \mathbb{I}_n + \left[\mathbb{I}_n - \frac{\mathbf{1} \mathbf{1}^t}{n+v} \right] (\mathbf{Y} - \mathbf{1} \mathbf{m}^t) \mathbf{D}^{-1} (\mathbf{Y} - \mathbf{1} \mathbf{m}^t)^t \right|^{-\frac{d+n}{2}} \\
&= \frac{\Gamma_p \left(\frac{(d+1-p) + n + p - 1}{2} \right)}{\Gamma_p \left(\frac{(d+1-p) + p - 1}{2} \right) \pi^{np/2}} \left| \mathbb{I}_n + \frac{\mathbf{1} \mathbf{1}^t}{v} \right|^{-p/2} |\mathbf{D}|^{-n/2} \\
&\quad \times \left| \mathbb{I}_n + \left[\mathbb{I}_n - \frac{\mathbf{1} \mathbf{1}^t}{n+v} \right] (\mathbf{Y} - \mathbf{1} \mathbf{m}^t) \mathbf{D}^{-1} (\mathbf{Y} - \mathbf{1} \mathbf{m}^t)^t \right|^{-\frac{(d+1-p) + n + p - 1}{2}}. \tag{B.6}
\end{aligned}$$

Portanto, segue de (B.6) que $\mathbf{Y} \sim \mathcal{T}_{n \times p} \left(\nu = d + 1 - p; \mathbf{1m}^t; \mathbf{I} + \frac{\mathbf{11}^t}{v}; \mathbf{D} \right)$.

Apêndice C

Algumas propriedades matriciais

Neste apêndice apresenta-se algumas propriedades de matrizes que são utilizadas para a demonstração da preposição 2.1.1 do Capítulo 2.

Denote por \mathbf{A}^t e \mathbf{A}^{-1} as matrizes transposta e inversa de \mathbf{A} , respectivamente. Denote por $|\mathbf{A}|$ o determinante de \mathbf{A} .

Propriedade 1. Seja \mathbf{A} uma matriz simétrica de dimensão $p \times p$ e \mathbf{x} um vetor de dimensão $p \times 1$. Então, $\mathbf{x}^t \mathbf{A} \mathbf{x} = tr(\mathbf{x}^t \mathbf{A} \mathbf{x}) = tr(\mathbf{A} \mathbf{x} \mathbf{x}^t)$.

Propriedade 2. Sejam \mathbf{A} e \mathbf{B} matrizes de dimensões $p \times p$. Então, $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$.

Propriedade 3. Seja \mathbf{A} uma matriz de dimensão $p \times p$ e c uma constante. Então, $|c\mathbf{A}| = c^p |\mathbf{A}|$.

Propriedade 4. Seja \mathbf{A} uma matriz inversível de dimensão $m \times m$, c um vetor de dimensão $m \times 1$ e r um vetor de dimensão $1 \times m$. Então, $|\mathbf{A} + cr| = |\mathbf{A}|(1 + r\mathbf{A}^{-1}c)$.

Propriedade 5. Se \mathbf{A} é inversível, então $(\mathbf{A}^{-1})^t = (\mathbf{A}^t)^{-1}$.

Propriedade 6. Se \mathbf{A} é simétrica, então $\mathbf{A} = \mathbf{A}^t$.

Propriedade 7. Sejam \mathbf{A} e \mathbf{B} matrizes de dimensões $p \times q$ e $q \times p$, respectivamente. Então, $|\mathbb{I}_p + \mathbf{A}\mathbf{B}| = |\mathbb{I}_q + \mathbf{B}\mathbf{A}|$ onde \mathbb{I}_q e \mathbb{I}_p são matrizes identidades de ordem q e p respectivamente.

Propriedade 8. Sejam $\mathbf{A}_{p \times p}$, $\mathbf{B}_{q \times q}$, $\mathbf{C}_{p \times q}$ e $\mathbf{D}_{q \times p}$ matrizes. Então,

$$(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{D})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}\mathbf{B}(\mathbf{B} + \mathbf{B}\mathbf{D}\mathbf{A}^{-1}\mathbf{C}\mathbf{B})^{-1}\mathbf{B}\mathbf{D}\mathbf{A}^{-1}.$$

Apêndice D

Gráficos

Este apêndice mostra os gráficos referentes aos cenários apresentados no Capítulo 3 e a aplicação realizada no Capítulo 4.

Cenário 1

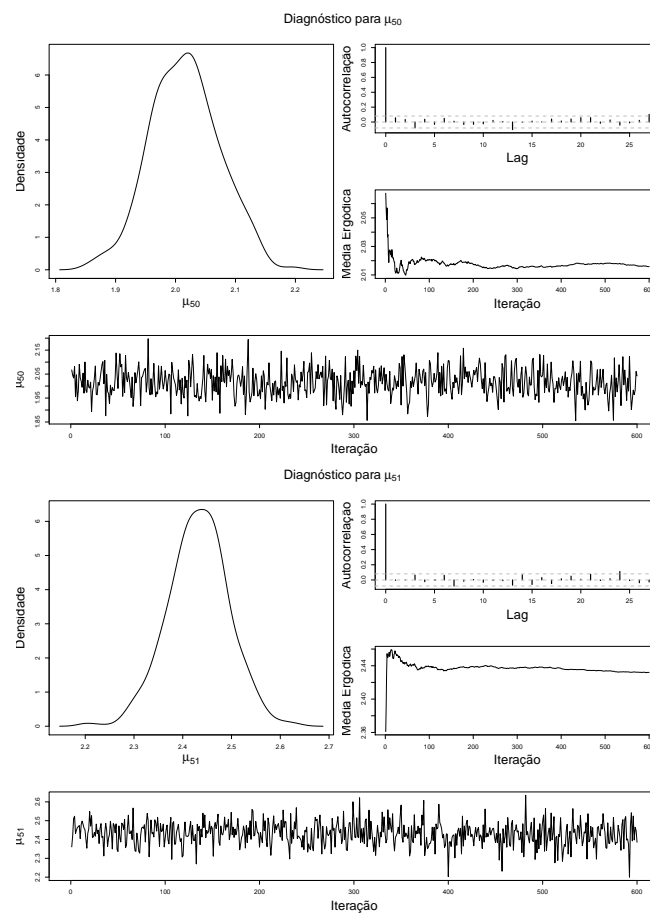


Figura D.1: Diagnóstico da cadeia de μ_{50} e μ_{51} da série 2, situação 2, $v = 1$ e $d = 6$.

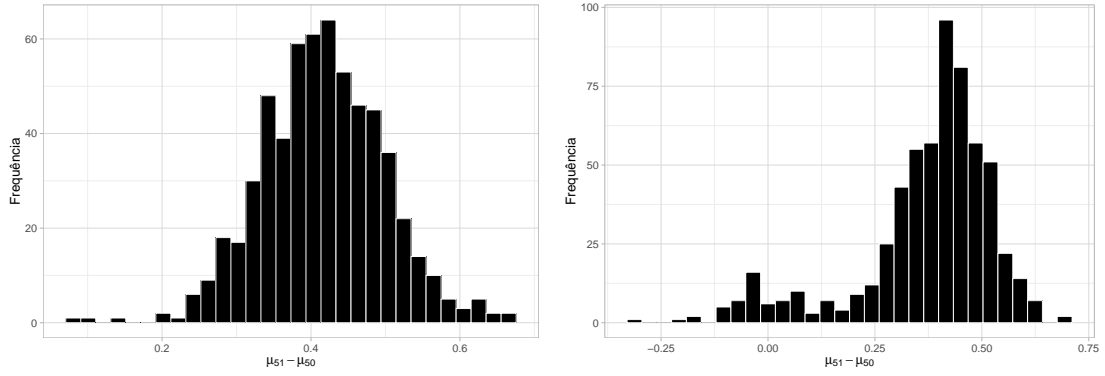


Figura D.2: O primeiro gráfico mostra a distribuição *a posteriori* de $\mu_{51} - \mu_{50}$ da série 2, situação 2, $v = 1$ e $d = 6$. O segundo gráfico mostra a distribuição *a posteriori* de $\mu_{51} - \mu_{50}$ da série 2, situação 2, $v = 1$ e $d = 4$.

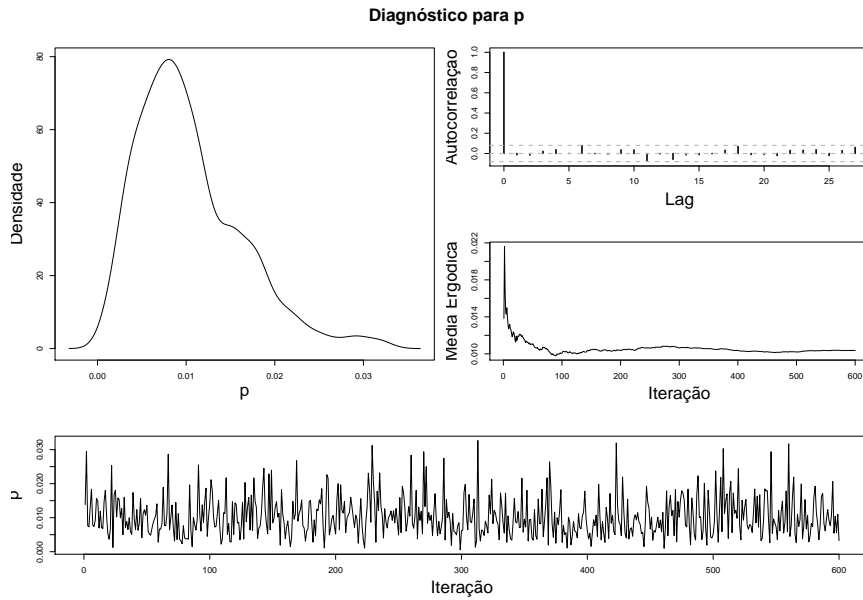


Figura D.3: Diagnóstico para p . Situação 2, $d = 4$ e $v = 0.01$.

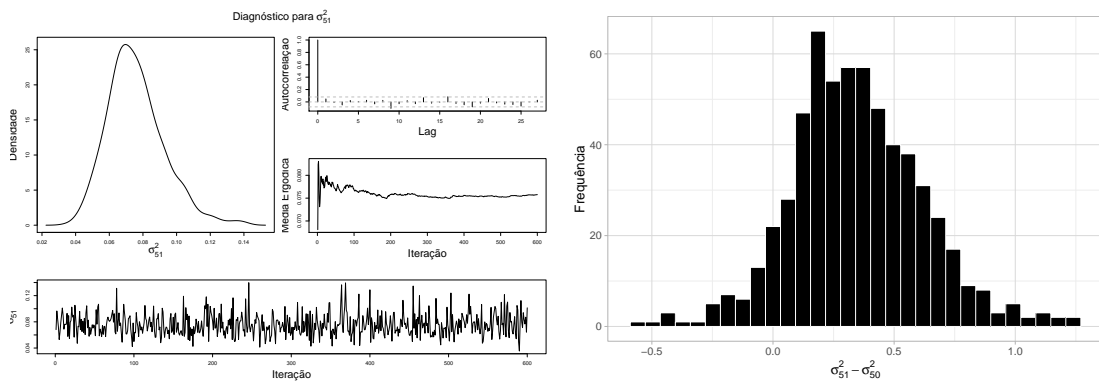


Figura D.4: O gráfico do lado esquerdo mostra o diagnóstico da cadeia de σ_{51}^2 da série 2, situação 2, $v = 0.01$ e $d = 4$. O gráfico do lado direito mostra a diferença entre $\sigma_{51}^2 - \sigma_{50}^2$ da série 2, situação 1, $v = 10$ e $d = 4$.

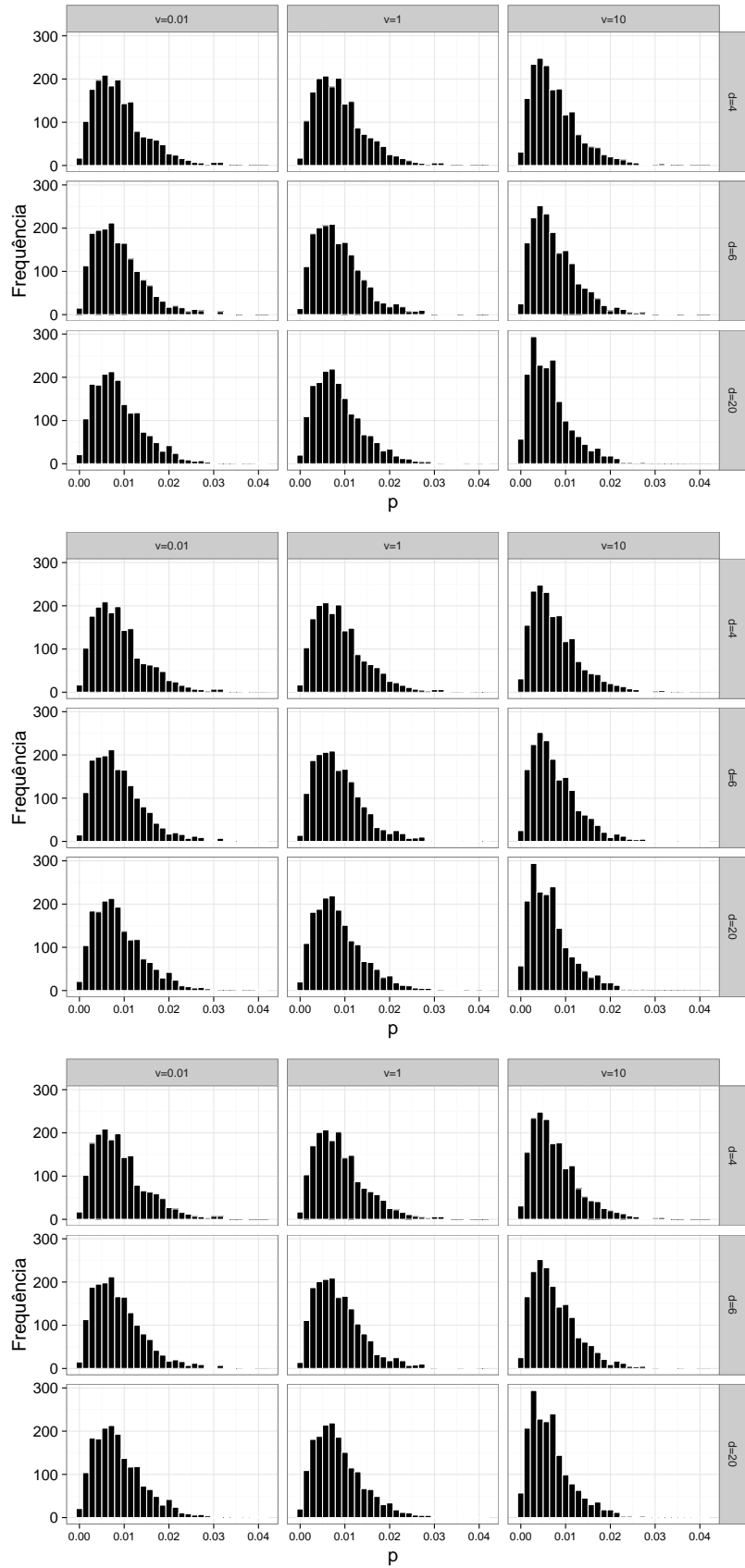


Figura D.5: Distribuição *a posteriori* da probabilidade de mudança para as Situações 1 (topo), 2 (meio) e 3 (base).

Cenário 2

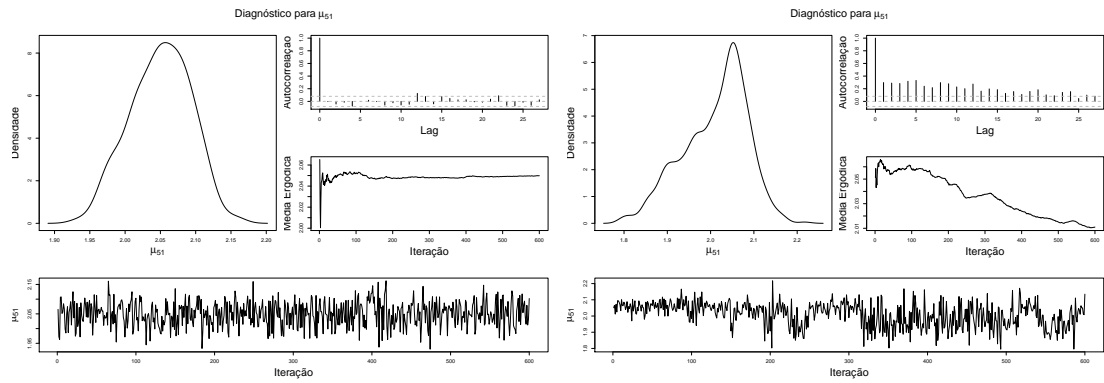


Figura D.6: O gráfico do lado esquerdo mostra o diagnóstico da cadeia de μ_{51} da série 1, para $d = 4$. O gráfico do lado direito mostra o diagnóstico da cadeia μ_{51} da série 1 para $d = 20$. Para ambos gráficos é considerada a situação 2 e $v = 0.01$.

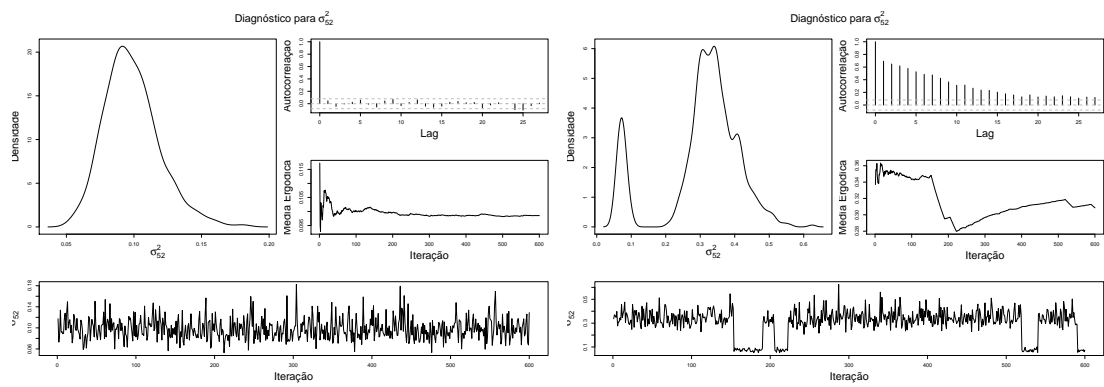


Figura D.7: O gráfico do lado esquerdo mostra o diagnóstico da cadeia de σ_{52}^2 da série 2, para $d = 4$. O gráfico do lado direito mostra o diagnóstico da cadeia σ_{52}^2 da série 2 para $d = 20$. Para ambos gráficos é considerada a situação 2 e $v = 0.01$.

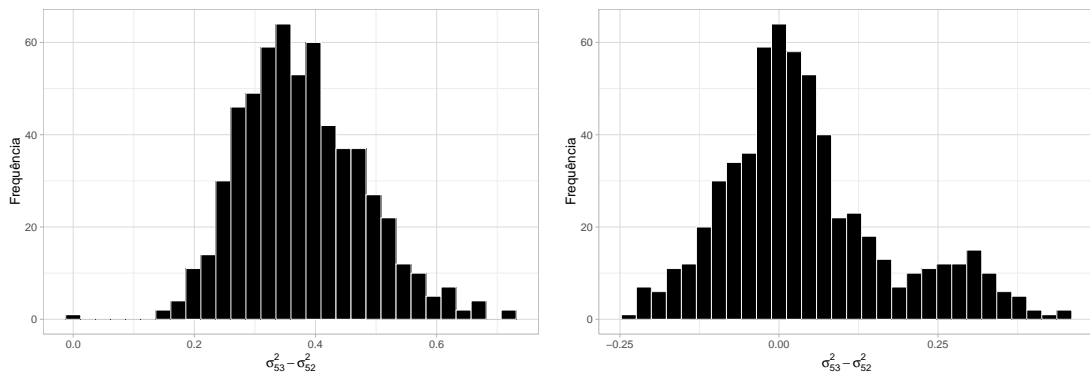


Figura D.8: O gráfico do lado esquerdo corresponde a diferença de $\sigma_{53}^2 - \sigma_{52}^2$ para a combinação de $v = 0.01$ e $d = 4$, enquanto o gráfico do lado direito corresponde a diferença de $\sigma_{53}^2 - \sigma_{52}^2$ para a combinação de $v = 0.01$ e $d = 20$. Para ambas as diferenças, foi considerada a situação 2.

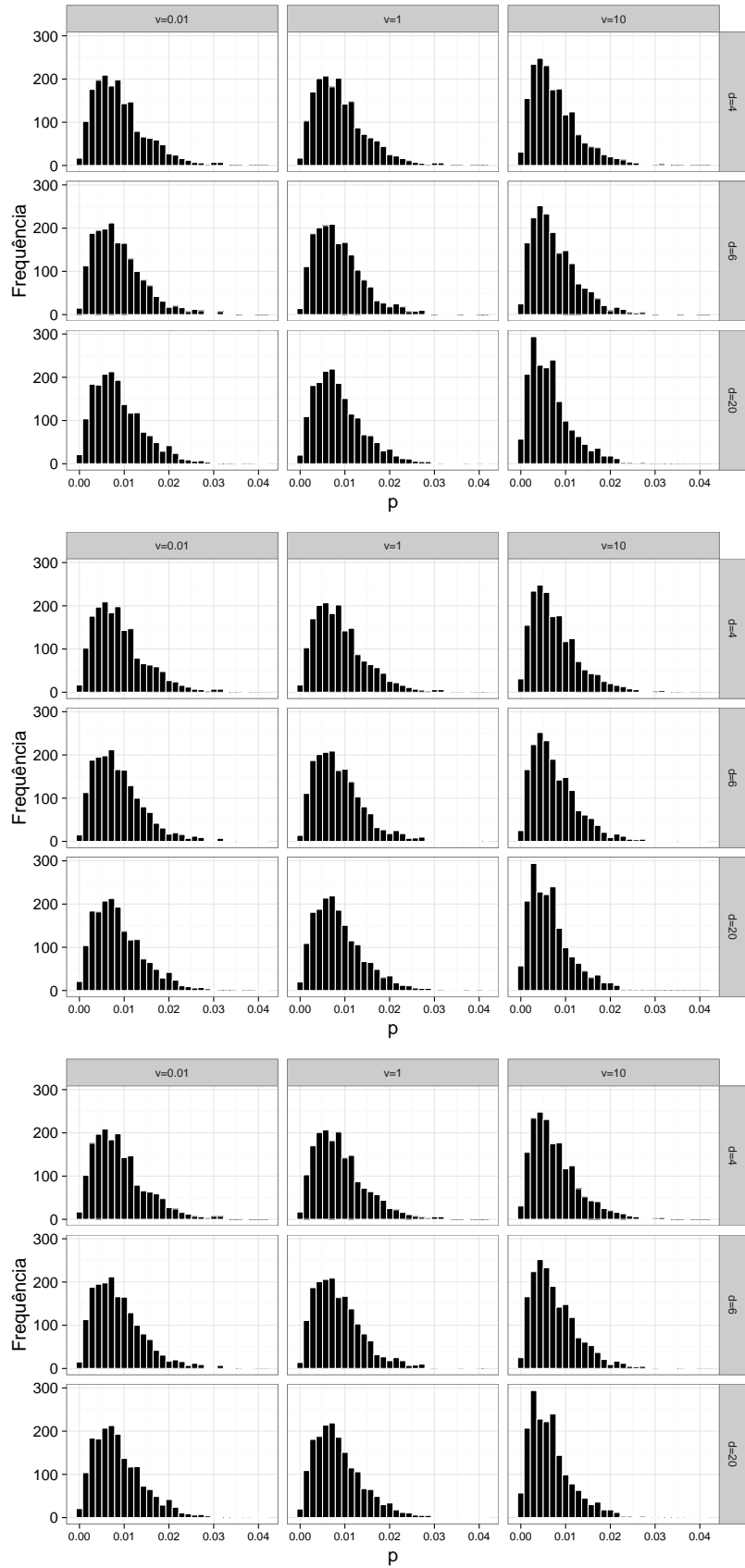


Figura D.9: Distribuição *a posteriori* da probabilidade de mudança para as Situações 1 (topo), 2 (meio) e 3 (base).

Aplicação

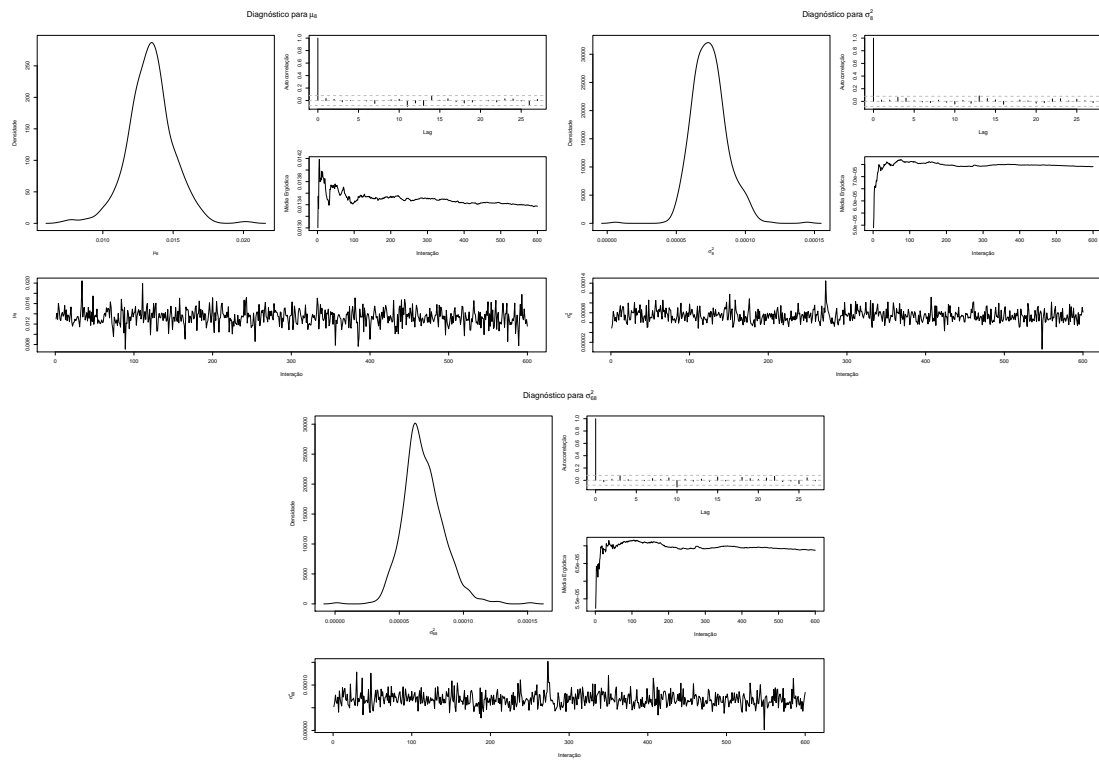


Figura D.10: Da esquerda para direita, diagnóstico da média e da variância do Reino Unido e diagnóstico da covariância entre Reino Unido e Itália. Em todos os gráficos foram considerados o dia 28/6.

Digóstico para p

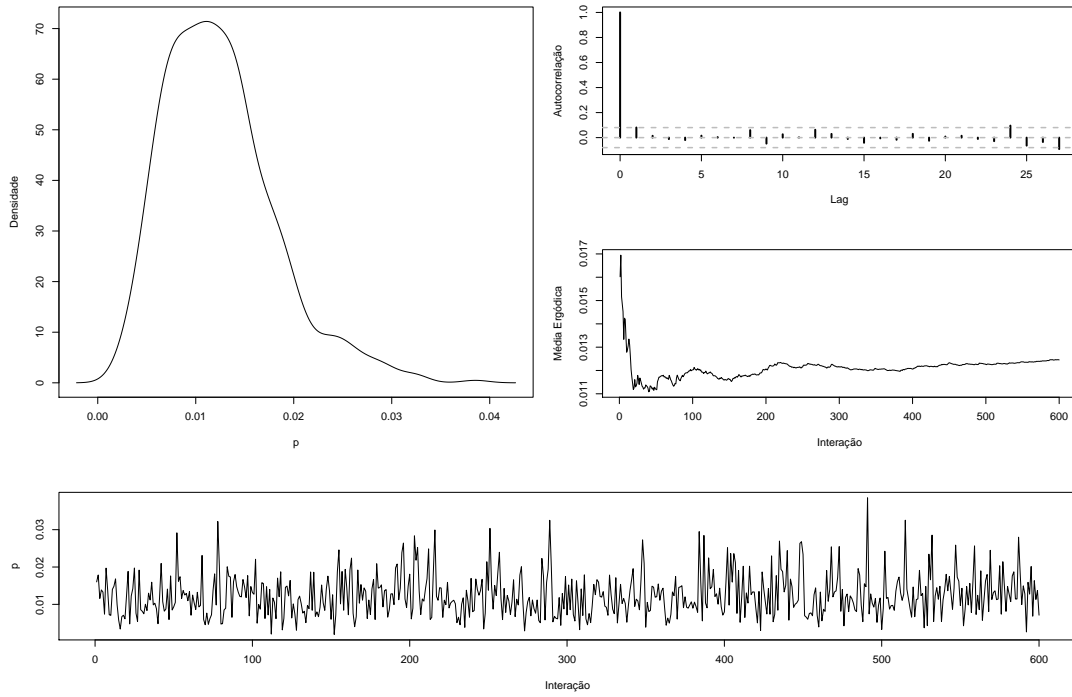


Figura D.11: Diagnóstico para p .

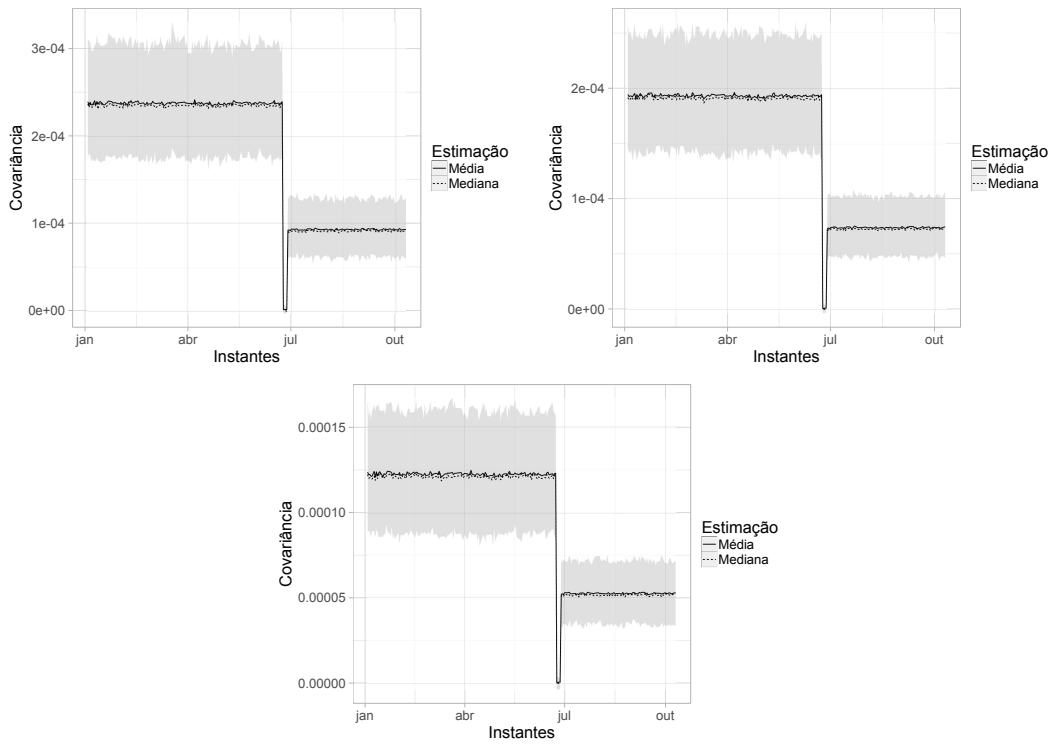


Figura D.12: Estimativas para a covariância, da esquerda para a direita, Itália e Portugal, Itália e Suíça e Portugal e Suíça

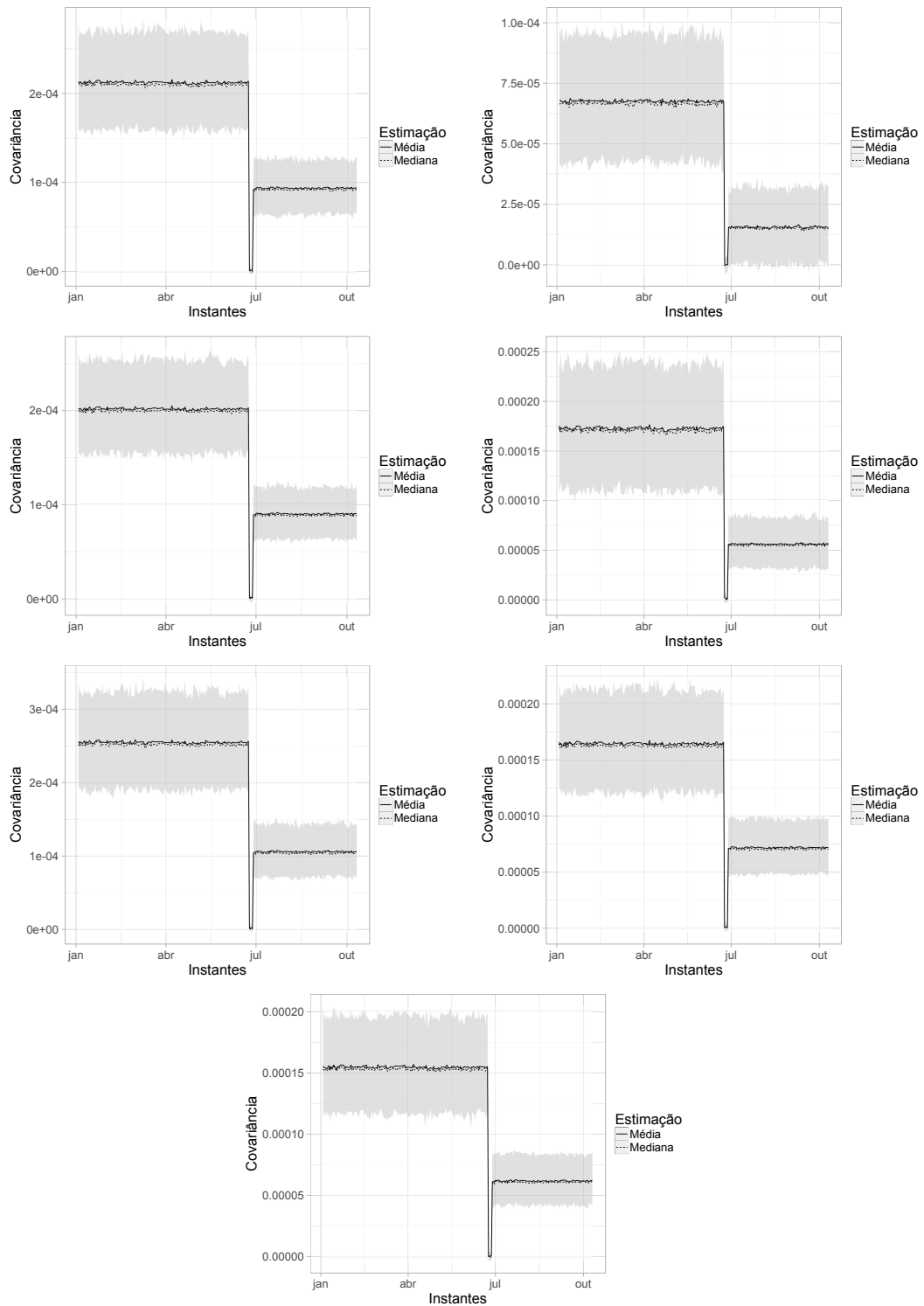


Figura D.13: Estimativas para a covariância, da esquerda para direita, entre Alemanha e Espanha, Alemanha e EUA, Alemanha e França, Alemanha e Grécia, Alemanha e Itália, Alemanha e Portugal, Alemanha e Suíça.

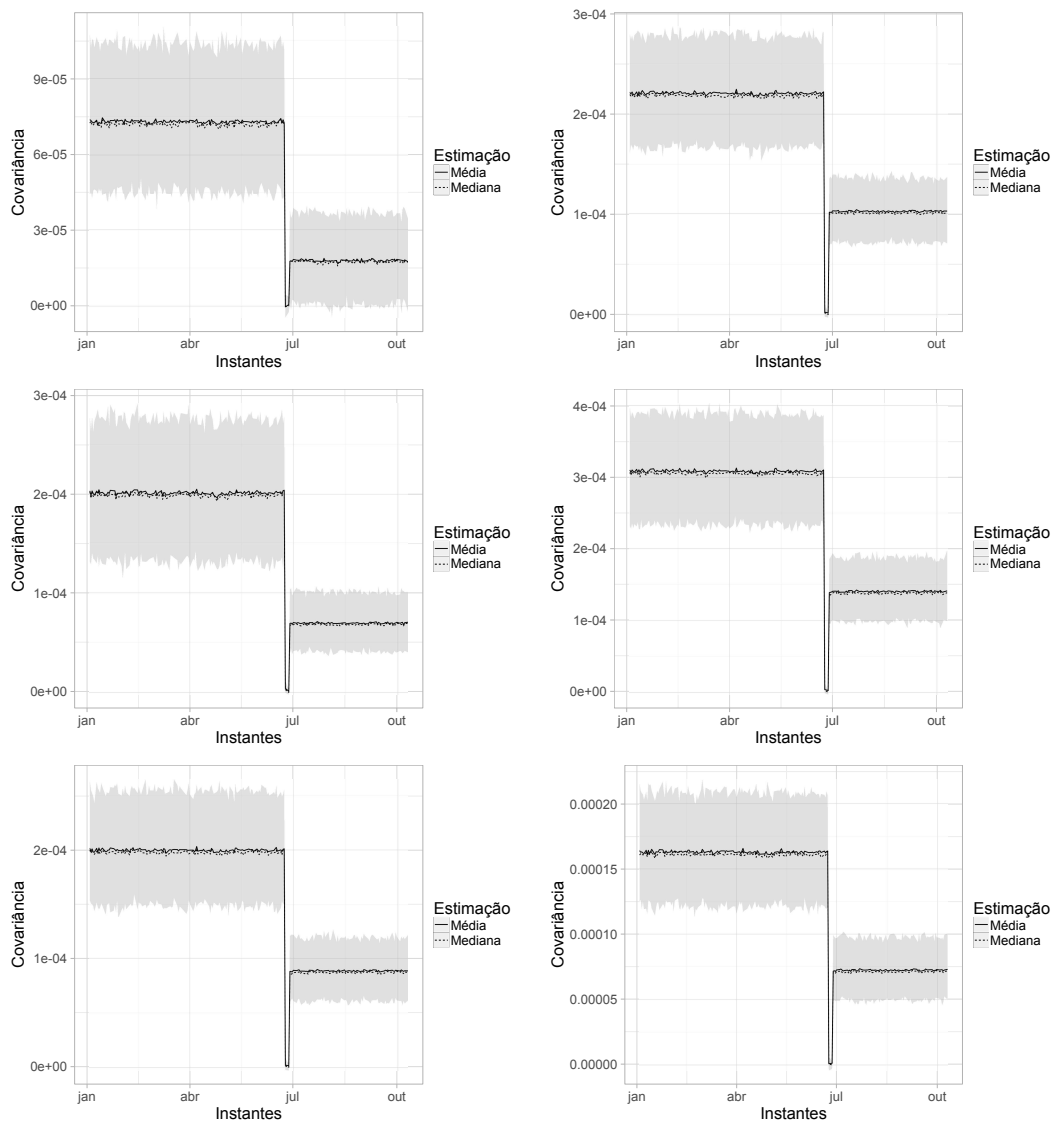


Figura D.14: Estimativas para a covariância, da esquerda para a direita, entre Espanha e EUA, Espanha e França, Espanha e Grécia, Espanha e Itália, Espanha e Portugal, Espanha e Suíça.

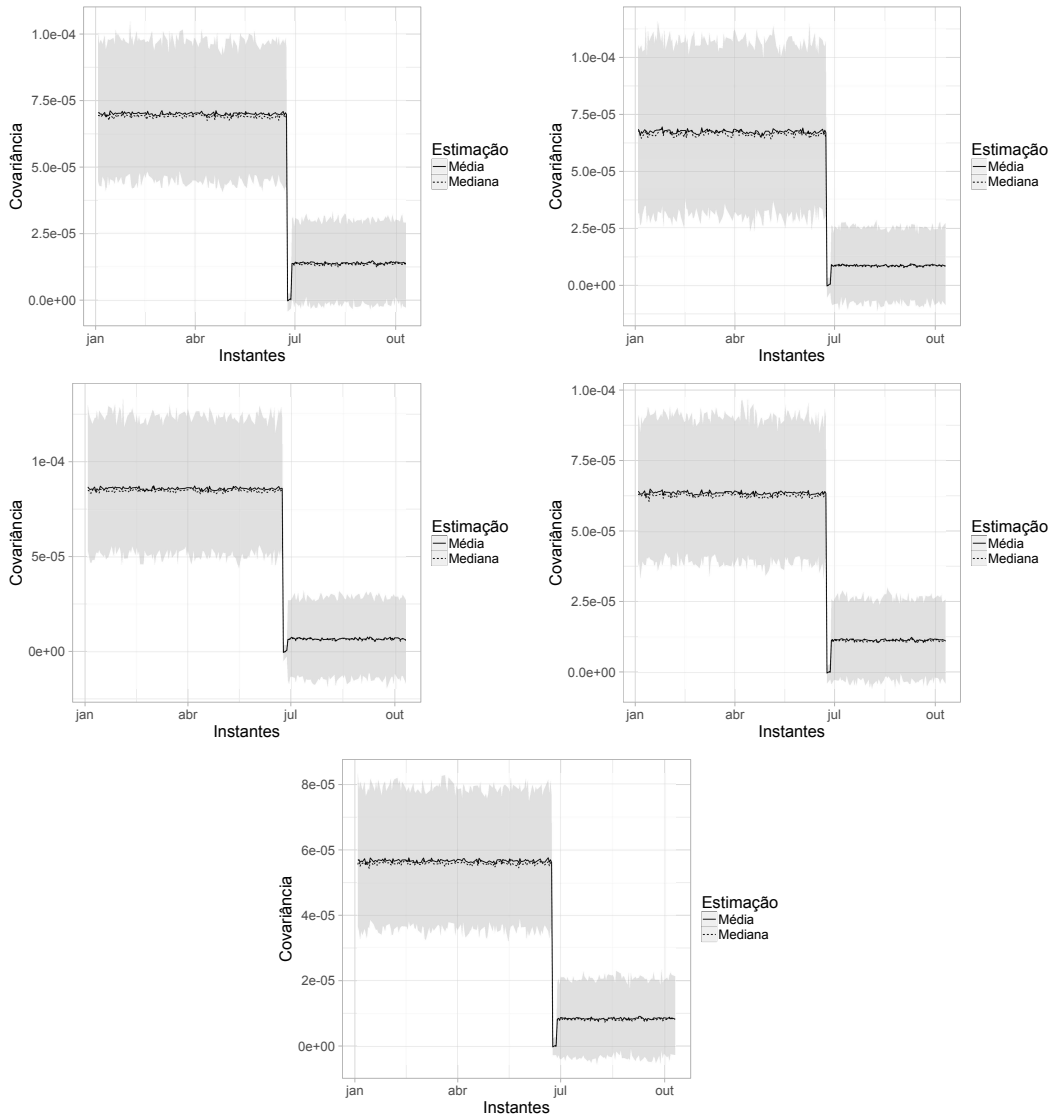


Figura D.15: Estimativas para a covariância, da esquerda para a direita, entre EUA e França, EUA e Grécia, EUA e Itália, EUA e Portugal e EUA e Suíça.

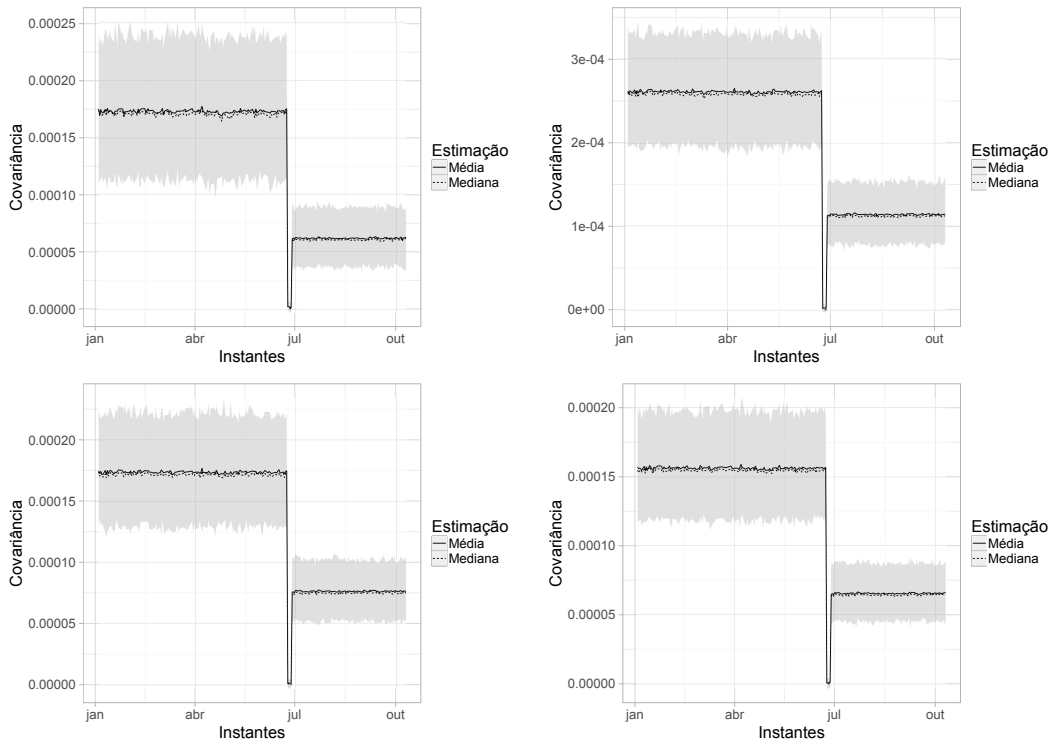


Figura D.16: Estimativas para a covariância, da esquerda para a direita, entre França e Grécia, França e Itália, França e Portugal e França e Suíça.

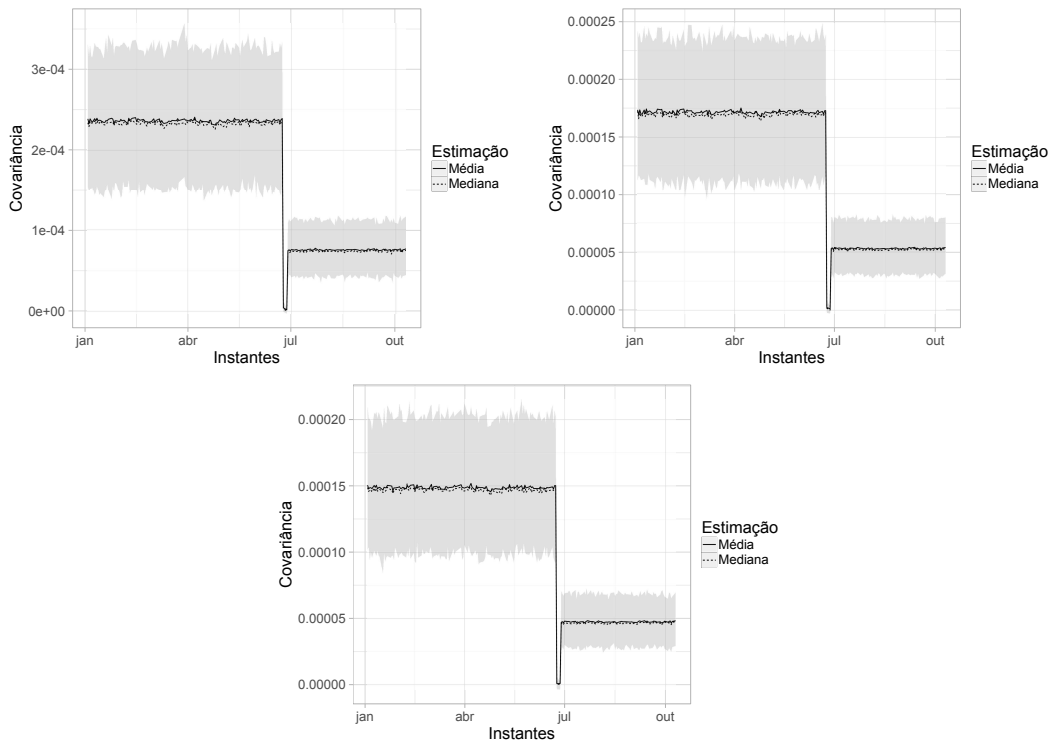


Figura D.17: Estimativas para a covariância, da esquerda para a direita, Grécia e Itália, Grécia e Portugal e Grécia e Suíça.

Referências Bibliográficas

- [1] Allen, D. E., McAleer, M., Powell, R. J. & Kumar-Singh, A. (2013). Nonparametric multiple change point analysis of the global financial crisis. *Available at SSRN 2270029*.
- [2] Arellano-Valle, R. B., Castro, L. M. & Loschi, R. H. (2013). Change point detection in the skew-normal model parameters. *Communications in Statistics-Theory and Methods*, **42**(4), 603–618.
- [3] Barry, D. & Hartigan, J. A. (1992). Product partition models for change point problems. *The Annals of Statistics*, pages 260–279.
- [4] Barry, D. & Hartigan, J. A. (1993). A bayesian analysis for change point problems. *Journal of the American Statistical Association*, **88**(421), 309–319.
- [5] Bhattacharya, P. K. (1987). Maximum likelihood estimation of a change-point in the distribution of independent random variables: general multiparameter case. *Journal of Multivariate Analysis*, **23**(2), 183–208.
- [6] Booth, N. & Smith, A. (1982). A Bayesian approach to retrospective identification of change-points. *Journal of Econometrics*, **19**(1), 7–22.
- [7] Bragança Pereira, C. A. & Stern, J. M. (1999). Evidence and credibility: full bayesian significance test for precise hypotheses. *Entropy*, **1**(4), 99–110.
- [8] Cheon, S. & Kim, J. (2010). Multiple change-point detection of multivariate mean vectors with the Bayesian approach. *Computational Statistics & Data Analysis*, **54**(2), 406–415.
- [9] Diaz, J. (1982). Bayesian detection of a change of scale parameter in sequences of independent gamma random variables. *Journal of Econometrics*, **19**(1), 23–29.
- [10] Ferreira, J. A., Loschi, R. H. & Costa, M. A. (2014). Detecting changes in time series: A product partition model with across-cluster correlation. *Signal Processing*, **96**, 212–227.

- [11] Gupta, A. K. & Nagar, D. K. (1999). *Matrix variate distributions*, volume 104. CRC Press.
- [12] Hartigan, J. A. (1990). Partition models. *Communications in Statistics-Theory and Methods*, **19**(8), 2745–2756.
- [13] Hinkley, D. V. (1971). Inference about the change-point from cumulative sum tests. *Biometrika*, **58**(3), 509–523.
- [14] Holbert, D. (1982). A bayesian analysis of a switching linear model. *Journal of Econometrics*, **19**(1), 77–87.
- [15] James, N. A. & Matteson, D. S. (2013). ecp: An r package for nonparametric multiple change point analysis of multivariate data. *arXiv preprint arXiv:1309.3295*.
- [16] Johnson, R. A., Wichern, D. W. et al. (2002). *Applied multivariate statistical analysis*, volume 5. Prentice hall Upper Saddle River, NJ.
- [17] Kehagias, A., Nicolaou, A., Petridis, V. & Fragkou, P. (2004). Text segmentation by product partition models and dynamic programming. *Mathematical and Computer Modelling*, **39**(2), 209–217.
- [18] Kotz, S. & Nadarajah, S. (2004). *Multivariate t-distributions and their applications*. Cambridge University Press.
- [19] Lavielle, M. & Teyssiere, G. (2006). Detection of multiple change-points in multivariate time series. *Lithuanian Mathematical Journal*, **46**(3), 287–306.
- [20] Lee, A. F. & Heghinian, S. M. (1977). A shift of the mean level in a sequence of independent normal random variables? a bayesian approach? *Technometrics*, **19**(4), 503–506.
- [21] Liu, P. (2010). *Maximum Likelihood Estimation of an Unknown Change-point in the Parameters of a Multivariate Gaussian Series with Applications to Environmental Monitoring*. Ph.D. thesis, Citeseer.
- [22] Loschi, R. & Cruz, F. (2002). An analysis of the influence of some prior specifications in the identification of change points via product partition model. *Computational Statistics & Data Analysis*, **39**(4), 477–501.
- [23] Loschi, R., Iglesias, P. & Arellano-Valle, R. (1999). Bayesian detection of change points in the chilean stock market. In *Proceedings of the Annual Meeting*

of American Statistical Association. Section on Bayesian Statistical Science, (Baltimore, MD, USA, 1999), pages 160–165.

- [24] Loschi, R., Cruz, F. & Arellano-Valle, R. (2005a). Multiple change point analysis for the regular exponential family using the product partition model. *Journal of Data Science*, **3**(3), 305–330.
- [25] Loschi, R. H. & Cruz, F. R. (2005). Extension to the product partition model: Computing the probability of a change. *Computational Statistics & Data Analysis*, **48**(2), 255–268.
- [26] Loschi, R. H., Cruz, F. R., Iglesias, P. L. & Arellano-Valle, R. B. (2003). A gibbs sampling scheme to the product partition model: an application to change-point problems. *Computers & Operations Research*, **30**(3), 463–482.
- [27] Loschi, R. H., Gonçalves, F. B. & Cruz, F. R. (2005b). Avaliação de uma medida de evidência de um ponto de mudança e sua utilização na identificação de mudanças na taxa de criminalidade em belo horizonte. *Pesquisa Operacional*, **25**(3), 449–463.
- [28] Loschi, R. H., Pontel, J. G. & Cruz, F. R. (2010). Multiple change-point analysis for linear regression models.
- [29] Mabaera, B., Naranbat, N., Katamba, A., Laticevschi, D., Lauritsen, J. M. & Rieder, H. L. (2009). Seasonal variation among tuberculosis suspects in four countries. *International health*, **1**(1), 53–60.
- [30] Maechler, M. (2015). *Rmpfr: R MPFR - Multiple Precision Floating-Point Reliable*. R package version 0.6-0.
- [31] Martínez, A. F., Mena, R. H. et al. (2014). On a nonparametric change point detection model in markovian regimes. *Bayesian Analysis*, **9**(4), 823–858.
- [32] Monteiro, J. V., Assunção, R. M., Loschi, R. H. et al. (2011). Product partition models with correlated parameters. *Bayesian Analysis*, **6**(4), 691–726.
- [33] Moura, C. R. (2004). *Extensão do Modelo Partição Produto para Dados Normais Multivariados: Uma Análise da Correlação de Índices Financeiros..* Dissertação de mestrado, Universidade Federal de Minas Gerais.
- [34] Müller, P., Quintana, F. & Rosner, G. L. (2012). A product partition model with regression on covariates. *Journal of Computational and Graphical Statistics*.

- [35] O'Hagan, A. & Forster, J. J. (2004). *Kendall's advanced theory of statistics, volume 2B: Bayesian inference*, volume 2. Arnold.
- [36] Perreault, L., Bernier, J., Bobée, B. & Parent, E. (2000a). Bayesian change-point analysis in hydrometeorological time series. part 1. the normal model revisited. *Journal of Hydrology*, **235**(3), 221–241.
- [37] Perreault, L., Parent, E., Bernier, J., Bobee, B. & Slivitzky, M. (2000b). Retrospective multivariate bayesian change-point analysis: a simultaneous single change in the mean of several hydrological sequences. *Stochastic Environmental Research and Risk Assessment*, **14**(4-5), 243–261.
- [38] Quintana, F. A. & Iglesias, P. L. (2003). Bayesian clustering and product partition models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**(2), 557–574.
- [39] R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [40] Ruggieri, E. (2013). A bayesian approach to detecting change points in climatic records. *International Journal of Climatology*, **33**(2), 520–528.
- [41] Salazar, D. (1982). Structural changes in time series models. *Journal of Econometrics*, **19**(1), 147–163.
- [42] Smith, A. (1975). A bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, **62**(2), 407–416.
- [43] Son, Y. S. & Kim, S. W. (2005). Bayesian single change point detection in a sequence of multivariate normal observations. *Statistics*, **39**(5), 373–387.
- [44] Yao, Y.-C. (1984). Estimation of a noisy discrete-time step function: Bayes and empirical bayes approaches. *The Annals of Statistics*, pages 1434–1447.