



MODELOS ESTOCÁSTICOS DE TAXAS DE MORTALIDADE E APLICAÇÕES

Sandra Maria Damásio Lagarto

Tese apresentada à Universidade de Évora
para obtenção do Grau de Doutor em Matemática
Especialidade: Estatística

ORIENTADOR: *Professor Doutor Carlos Alberto dos Santos Braumann*
CO-ORIENTADORA: *Professora Doutora Dulce Maria de Oliveira Gomes*

ÉVORA, JULHO DE 2014





MODELOS ESTOCÁSTICOS DE TAXAS DE MORTALIDADE E APLICAÇÕES

Sandra Maria Damásio Lagarto

Tese apresentada à Universidade de Évora
para obtenção do Grau de Doutor em Matemática
Especialidade: Estatística

ORIENTADOR: *Professor Doutor Carlos Alberto dos Santos Braumann*
CO-ORIENTADORA: *Professora Doutora Dulce Maria de Oliveira Gomes*

ÉVORA, JULHO DE 2014

INSTITUTO DE INVESTIGAÇÃO E FORMAÇÃO AVANÇADA

À memória do meu pai

Agradecimentos

Aos meus Orientadores, Professor Doutor Carlos Braumann e Professora Doutora Dulce Gomes, agradeço a generosidade na partilha do conhecimento e o constante entusiasmo. Sem a intervenção conjunta de ambos não teria sido possível concluir este processo.

Agradeço também aos Colegas e Professores da Universidade de Évora, com os quais considero ter sido igualmente um privilégio aprender nestes últimos anos, não só Matemática, mas sobretudo a sua utilidade para a vida. À Professora Doutora Sandra Vinagre, pela motivação que me fez regressar todos os dias à UÉ no primeiro ano do curso de licenciatura em Matemática Aplicada, e aos Professores Doutor Russel Alpizar e José Carlos Tiago de Oliveira, pelo apoio constante em todo este processo. Ao Dr. Nuno Brites e à Professora Doutora Patrícia Filipe, pela ajuda com o R na fase inicial deste trabalho. Ao Professor Doutor João Corte-Real, pela ajuda na obtenção e escolha da metodologia para o tratamento dos dados climáticos.

À Universidade de Évora faço ainda dois agradecimentos institucionais. À equipa do Projeto “PTDC/SDE/68126/2006 - O Futuro da População Portuguesa: a importância da estimativa da mortalidade e das migrações ao nível regional”, co-financiado pela FCT e pelo FEDER através do Programa POCI 2010, em que fui Bolseira de Investigação e à sua coordenadora, Professora Doutora Maria Filomena Mendes, por me ter introduzido no estudo dos fenómenos demográficos, e em particular da mortalidade, numa altura em que fiz as primeiras experiências que mais tarde viriam a conduzir ao Projeto de Tese. Ao Instituto de Investigação e Formação Avançada, responsável pelo processo de atribuição da Bolsa de Doutoramento no âmbito do Programa Bento de Jesus Caraça, sem a qual não teria sido possível frequentar este curso.

Agradeço ainda a todos aqueles que me acompanharam nesta aventura, que tem sido aprender matemática depois dos 30 (o que se revelou uma agradável surpresa), e que, em algum momento, me apoiaram ou me transmitiram algum tipo de motivação. À minha Mãe e à minha Irmã, porque são o meu suporte, um especial obrigada!

In precisely built mathematical structures, mathematicians find the same sort of beauty others find in enchanting pieces of music, or in magnificent architecture. There is, however, one great difference between the beauty of mathematical structures and that of great art. [...] Only mathematicians can read “musical scores” containing many numerical formulae, and play that “music” in their hearts. Accordingly, I once believed that without numerical formulae, I could never communicate the sweet melody played in my heart.

Kiyoshi Itô, *My Sixty Years in Studies of Probability Theory: acceptance speech of the Kyoto Prize in Basic Sciences* (1998)

Resumo

O prolongamento da vida humana é considerado atualmente um problema no plano socioeconómico. Os modelos probabilísticos para estudar a evolução das taxas de mortalidade têm, sobre os determinísticos, a vantagem de incorporar os efeitos aleatórios das variações ambientais (em sentido lato) e determinar o grau de incerteza das previsões.

Fazemos uma incursão na análise transversal da mortalidade ao longo do tempo, em alternativa à usual análise por coorte, desenvolvendo modelos de equações diferenciais estocásticas, que aplicamos à população portuguesa e que explicam a evolução temporal das taxas de mortalidade em todas as idades do arco de vida e de ambos os sexos.

Construímos modelos univariados separadamente para cada idade e sexo com fontes de ruído independentes, modelos bivariados por idade com correlações entre sexos e modelos multivariados com correlação entre idades e entre idades e sexos. Foi feito um estudo comparativo entre estes modelos e destes com modelos alternativos.

Abstract

STOCHASTIC DEATH RATES MODELS AND APPLICATIONS

The extension of human life is considered a very demanding social and economical issue. When we plan to study the evolution of death rates, stochastic models have some advantages compared to the deterministic ones, because we can input random environmental fluctuations and evaluate the uncertainty in predictions.

We propose a cross-sectional analysis of mortality, instead of the usual cohort analysis, by developing stochastic differential equations models, which we have applied to the Portuguese population, describing death rates trends for all ages of the life span of males and females.

We build univariate models separately for each age and sex with independent noise sources, bivariate models for each age with correlations between sexes, and multivariate models with correlations among ages and with correlations among ages and sexes. We compare these models with one another and with alternative ones.

Índice

1	Introdução	1
1.1	Motivação: o estudo da tendência evolutiva da mortalidade humana e os dados da população portuguesa	2
1.2	Breve síntese da literatura e porque modelar a mortalidade com equações diferenciais estocásticas	6
1.3	Objetivos e organização da tese	8
2	Breve introdução às equações diferenciais estocásticas	11
2.1	Introdução	12
2.2	Processos estocásticos	12
2.3	Equações diferenciais estocásticas	13
2.4	Fórmula de Itô	15
2.5	Exemplo: a lei de Gompertz	15
3	Modelos univariados de equações diferenciais estocásticas para taxas de mortalidade	19
3.1	Introdução	20
3.2	Movimento browniano geométrico	20
3.2.1	O MBG	21
3.2.2	Estimação	21
3.2.3	Resultados	24
3.3	Modelo de Gompertz estocástico	30
3.3.1	O MGE	30
3.3.2	Estimação	31
3.3.3	Resultados	34
3.4	Comparação dos resultados entre modelos	39
3.4.1	MBG vs MGE	40
3.4.2	Modelos de EDEs vs RNL	43
3.4.3	Modelos de EDEs vs ARIMA	48
3.5	Previsões longitudinais em modelos de EDEs	56
4	Modelos bivariados de equações diferenciais estocásticas para taxas de mortalidade	57
4.1	Introdução	58

4.2	Movimento browniano geométrico bidimensional com processos de Wiener correlacionados	59
4.2.1	O MBGB	59
4.2.2	Estimação	62
4.2.3	Resultados	64
4.3	Modelo de Gompertz estocástico bidimensional com processos de Wiener correlacionados	69
4.3.1	O MGEB	70
4.3.2	Estimação	72
4.3.3	Resultados	73
4.4	Comparação entre o movimento browniano geométrico bidimensional e um modelo de vetores autorregressivos e/ou de médias móveis	79
4.4.1	Modelos VARMA	80
4.4.2	Análise estrutural	82
4.4.3	Exemplo ilustrativo de um modelo VAR(p)	83
4.4.4	MBGB vs VAR(1)	94
5	Modelos multivariados de equações diferenciais estocásticas para taxas de mortalidade	97
5.1	Introdução	98
5.2	Movimento browniano geométrico multidimensional com processos de Wiener correlacionados	100
5.2.1	O MBGM	100
5.2.2	Estimação	103
5.2.3	Aspetos práticos relacionados com a implementação do modelo	105
5.2.4	Resultados do ajustamento	109
5.2.5	Resultados das previsões	121
6	Conclusões e trabalho futuro	137
	Apêndice A Código do R: Aspetos gerais	143
	Apêndice B Código do R: MBG	145
	Apêndice C Código do R: MGE	153
	Apêndice D Código do R: MGEB	163
	Apêndice E Código do R: MBGM	169
	Apêndice F Modelos ARIMA ajustados aos dados da mortalidade da população portuguesa	188
	Referências bibliográficas	197

Lista de Figuras

1.1	TBMs da população portuguesa: representação longitudinal vs transversal ao longo do tempo, sexo feminino	3
1.2	TBMs da população portuguesa de idades várias e sexos diferentes	4
1.3	TBM vs força da mortalidade, idade 63, sexo masculino	5
1.4	Diagrama metodológico	8
3.1	MBG: estimativas de \hat{R} com $IC_{95\%}$ exatos e assintóticos, por idade e por sexo	24
3.2	MBG: estimativas de \hat{V} com $IC_{95\%}$ exatos e assintóticos, por idade e por sexo	25
3.3	MBG: ajustamento com previsões a LP (25 anos: 2000 a 2024) e ampliação das previsões (2000 a 2009) PP e a LP com $IC_{95\%}$ assintóticos, idade 8, sexo masculino	26
3.4	MBG: réplicas simuladas da TBM, idade 8, sexo masculino	27
3.5	MBG: EQM do ajustamento às TBMs, por idade e por sexo	28
3.6	MBG: EQM das previsões a LP das TBMs (2000 a 2009), por idade e por sexo	28
3.7	MBG: ajustamento com previsões a LP (2000 a 2009), idades 29 e 98, sexo masculino	29
3.8	MGE: estimativas dos parâmetros, por idade e por sexo	35
3.9	MGE: ajustamento com previsões a LP (25 anos: 2000 a 2024) e ampliação das previsões (2000 a 2009) PP e a LP com $IC_{95\%}$ assintóticos, idade 39, sexo feminino	37
3.10	MGE: EQM do ajustamento às TBMs, por idade e por sexo	38
3.11	MGE: EQM das previsões a LP das TBMs (2000 a 2009), por idade e por sexo	39
3.12	MBG vs MGE: previsões a LP (25 anos: 2000 a 2024), idade 19, por sexo	40
3.13	MBG vs MGE: diferença entre os EQMs do ajustamento às TBMs, por idade e por sexo	41
3.14	MBG vs MGE: diferença entre os EQMs das previsões PP das TBMs (2000 a 2009), por idade e por sexo	42
3.15	MBG vs MGE: diferença entre os EQMs das previsões a LP das TBMs (2000 a 2009), por idade e por sexo	42
3.16	MBG vs MGE: EQM das previsões a LP das TBMs (2000 a 2009), idades 1 a 15, sexo feminino	43
3.17	MGE vs RNL: ajustamento e previsões (2000 a 2009) PP e a LP, idades 0 do sexo feminino e 50 e 84 do masculino	45
3.18	MGE vs RNL: diferença entre os EQMs do ajustamento às TBMs, por idade e por sexo	47

3.19 MGE vs RNL: diferença entre os EQMs das previsões a LP das TBMs (2000 a 2009), por idade e por sexo	47
3.20 MBG vs MGE vs ARIMA: ajustamento e previsões a LP (2000 a 2009), idades 6 e 25 do sexo feminino e 70 do masculino	50
3.21 MBG vs ARIMA: diferença entre os EQMs do ajustamento às TBMs, por idade e por sexo	52
3.22 MBG vs ARIMA: diferença entre os EQMs das previsões a LP das TBMs (2000 a 2009), por idade e por sexo	52
3.23 MGE vs ARIMA: diferença entre os EQMs do ajustamento às TBMs, por idade e por sexo	53
3.24 MGE vs ARIMA: diferença entre os EQMs das previsões a LP das TBMs (2000 a 2009), por idade e por sexo	53
3.25 MBG vs MGE vs RNL vs ARIMA: previsões a LP (25 anos: 2000 a 2024), idades 4 e 15 do sexo masculino e 29 e 95 do feminino	55
3.26 MBG: representação das previsões a 10 anos por coorte, por sexo	56
4.1 TBMs da população portuguesa: gráfico de superfície, todas as idades, por sexo	58
4.2 TBMs da população portuguesa: idades 2, 13, 17, 37, 72 e 83, por sexo	60
4.3 MBGB: estimativas dos parâmetros, por idade e por sexo	64
4.4 MBGB: ajustamento com previsões a LP (25 anos: 2000 a 2024) e ampliação das previsões (2000 a 2009) PP e a LP, idade 1, sexo feminino	66
4.5 MBG vs MBGB: diferença entre os EQMs do ajustamento às TBMs, por idade e por sexo	68
4.6 MBG vs MBGB: diferença entre os EQMs das previsões a LP das TBMs (2000 a 2009), por idade e por sexo	68
4.7 MGE: estimativas dos parâmetros, por idade e por sexo	74
4.8 MGE: ajustamento com previsões a LP (25 anos: 2000 a 2024) e ampliação das previsões (2000 a 2009) PP e a LP, idade 0, sexo feminino	75
4.9 MGE vs MGE: diferença entre os EQMs do ajustamento às TBMs, por idade e por sexo	77
4.10 MGE vs MGE: diferença entre os EQMs das previsões a LP das TBMs (2000 a 2009), por idade e por sexo	77
4.11 MBGB vs MGE: diferença entre os EQMs do ajustamento às TBMs, por idade e por sexo	78
4.12 MBGB vs MGE: diferença entre os EQMs das previsões a LP das TBMs (2000 a 2009), por idade e por sexo	79
4.13 Logaritmos dos retornos das TBMs da população portuguesa: FACs e FACPs, idade 2, por sexo	83
4.14 Logaritmos dos retornos das TBMs da população portuguesa: FCC, idade 2 do sexo feminino e masculino	86
4.15 VAR(1): resíduos, idade 2, sexo feminino	89

4.16 VAR(1): resíduos, idade 2, sexo masculino	90
4.17 VAR(1): carta de controlo tipo <i>CUSUM</i>	90
4.18 VAR(1): ajustamento, idade 2, por sexo	91
4.19 VAR(1): previsões a LP dos logaritmos dos retornos das TBMs (2000 a 2009) com <i>IC</i> _{95%} , idade 2, por sexo	91
4.20 VAR(1): funções de impulso-resposta, idade 2	93
4.21 VAR(1): decomposição do erro das previsões, idade 2	94
4.22 MBGB vs VAR(1): previsões a LP (2000 a 2009), idade 2, por sexo	95
5.1 TBMs da população portuguesa: todas as idades, por sexo	99
5.2 TBMs da população portuguesa: várias idades, por sexo	101
5.3 MBGM: etapas de implementação do modelo	107
5.4 MBGM: ajustamento de funções polinomiais às médias e logaritmos das variâncias empíricas dos logaritmos dos retornos das TBMs, todas as idades, ambos os sexos . . .	110
5.5 MBGM: transformada de Fisher das correlações empíricas, por sexo, e limites de <i>IC</i> _{95%} para a distribuição normal	111
5.6 MBGM: representação simultânea da transformada de Fisher das correlações empíricas e curvas da resposta média em função da idade, para o sexo feminino	112
5.7 MBGM: representação 3D da transformada de Fisher das correlações empíricas e plano de regressão, sexo feminino	113
5.8 MBGM: ajustamento da transformada de Fisher das correlações empíricas em função da idade, através de métodos de RNL, sexo feminino	115
5.9 MBGM: ajustamento da transformada de Fisher das correlações empíricas em função da idade, através de métodos de RNL, sexo masculino	116
5.10 MBGM: ajustamento da transformada de Fisher das correlações empíricas em função da idade, através de métodos de RNL, entre sexos	117
5.11 MBGM: ajustamento, idade 1, ambos os sexos (versões 10, 12 e M do modelo)	121
5.12 MBGM: previsões a LP (2000 a 2009) dos logaritmos dos retornos das TBMs das primeiras 10 idades do sexo feminino (versão 12 do modelo)	122
5.13 MBGM: previsões a LP (2000 a 2009) das TBMs das primeiras 10 idades do sexo feminino (versão 12 do modelo)	123
5.14 MBGM: representação conjunta (em escala logarítmica) das previsões a LP (2000 a 2009), TBM observada, médias das simulações e <i>MC</i> _{95%} , idade 10, sexo masculino (versão 12 do modelo)	126
5.15 MBGM: representação conjunta da TBM observada, previsões a LP (2000 a 2009), médias das simulações e <i>MC</i> _{95%} , idade 10, sexo masculino (versão 12 do modelo) . . .	126
5.16 MBGM: representação conjunta das médias das simulações e da estimação pontual das previsões a LP, todas as idades, ano 2000, por sexo (versão 12 do modelo)	127

5.17 MBGM: representação conjunta das médias das simulações, a partir das previsões a LP, e das $MC_{95\%}$, todas as idades, ano 2000, por sexo (versão 12 do modelo)	127
5.18 MBGM: EQM das previsões a LP das TBMs (2000 a 2009), por sexo (diferentes versões do modelo)	129
5.19 MBGM: previsões a LP (2000 a 2009), idades 13 a 15, por sexo (versão 11 do modelo)	131
5.20 MBGM: previsões a LP (2000 a 2009), idade 5, sexo masculino (versão 10 do modelo: sem e com correlações entre idades de sexos diferentes)	133
5.21 MBGM: previsões a LP (2000 a 2009), idade 17, sexo feminino (versão 3 do modelo: sem e com correlações entre idades de sexos diferentes)	133
5.22 MBGM: representação conjunta das previsões a LP (25 anos: 2000 a 2024) para o grupo etário 80-84 (versão 3 do modelo) com previsões do MBG e projeções do modelo PLC, para a idade 80, sexo feminino	135

Lista de Tabelas

3.1	MGE vs RNL: estimativas dos parâmetros, idades 0 do sexo feminino e 50 e 84 do masculino	46
3.2	AR(1): estimativas dos parâmetros, idades 6 e 25 do sexo feminino e 70 do masculino .	49
4.1	MBGB: estimativas dos parâmetros com $IC_{95\%}$ assintóticos, várias idades	65
4.2	MBGB vs MBG: teste de razão de verosimilhanças, por idade	67
4.3	MGEB: estimativas dos parâmetros com $IC_{95\%}$ assintóticos, várias idades	74
4.4	MGEB vs MGE: teste de razão de verosimilhanças, por idade	76
4.5	MBGB: estimativas dos parâmetros, idade 2	94
5.1	MBGM: teste de razão de verosimilhanças para versões sem e com correlações entre sexos	132
F.1	Modelos ARIMA: ajustamento, sexo feminino, por grupo etário	189
F.2	Modelos ARIMA: ajustamento, sexo masculino, por grupo etário	190

Lista de Caixas

3.1	MBG: estimativas do parâmetro R com limites dos $IC_{95\%}$ exatos e assintóticos, idades 0 a 9, sexo masculino	25
3.2	MBG: estimativas do parâmetro V com limites dos $IC_{95\%}$ exatos e assintóticos, idades 0 a 9, sexo masculino	26
3.3	MBG: previsões a LP das TBMs (2000 a 2009), idade 8, por sexo	27
3.4	MBG: EQM do ajustamento, previsões (2000 a 2009) PP e a LP das TBMs, idades 0 a 9, por sexo	29
3.5	MGE: estimativas dos parâmetros e margens de erro dos $IC_{95\%}$ assintóticos, idades 30 a 39, por sexo	36
3.6	MGE: previsões a LP das TBMs (2000 a 2009), idade 39, por sexo	37
3.7	MGE: EQM do ajustamento, previsões (2000 a 2009) PP e a LP das TBMs, idades 30 a 39, por sexo	38
4.1	VAR: teste da raiz unitária de Dickey Fuller (<i>trend</i>), idade 2, sexo feminino	84
4.2	VAR: teste da raiz unitária de Dickey Fuller (<i>drift</i>), idade 2, sexo feminino	85
4.3	VAR(p): determinação da ordem p , idade 2	87
4.4	VAR(1): ajustamento, idade 2	87
4.5	VAR(1): testes de diagnóstico aos resíduos, idade 2	88
4.6	VAR(1): previsões a LP dos logaritmos dos retornos das TBMs (2000 a 2009), idade 2, por sexo	92
4.7	VAR(1): teste à causalidade de Granger, idade 2	92
4.8	VAR(1) vs MBGB: previsões a LP dos logaritmos dos retornos das TBMs, idade 2, por sexo	95
5.1	MBGM: ajustamento de uma função polinomial às médias empíricas dos logaritmos dos retornos das TBMs, todas as idades do sexo masculino	110
5.2	MBGM: curvas exploratórias, para observar a variação de uma função média para a transformada de Fisher das correlações empíricas, sexo feminino	112
5.3	MBGM: representação tridimensional da transformada de Fisher das correlações empíricas com plano de regressão, sexo feminino	113
5.4	MBGM: EQM e AIC do ajustamento dos modelos por RNL para a transformada de Fisher das correlações empíricas	115

5.5	MBGM: <i>output</i> da aplicação do método de otimização	119
5.6	MBGM: estimativas dos parâmetros e margens de erro dos $IC_{95\%}$ assintóticos	120
5.7	MBGM: estimativas dos parâmetros para as diferentes versões do modelo e respetivos valores de log-verosimilhança e de AIC	128
5.8	MBGM: estimativas da média dos EQM das previsões a LP das TBMs (2000 a 2009) segundo diferentes versões do modelo, por sexo	130

Lista de Abreviaturas

- ARIMA** autorregressivos integrados e de médias móveis
- EDE** equação diferencial estocástica
- EQM** erro quadrático médio
- FAC** função de autocorrelação
- FACP** função de autocorrelação parcial
- FCC** função de correlação cruzada
- f.d.p.** função densidade de probabilidade
- LP** longo prazo
- MBG** movimento browniano geométrico
- MGE** modelo de Gompertz estocástico
- MGEB** modelo de Gompertz estocástico bidimensional
- MBGB** movimento browniano geométrico bidimensional
- MBGM** movimento browniano geométrico multidimensional
- MV** máxima verosimilhança
- p.e.** processo estocástico
- PP** passo-a-passo
- RNL** regressão não linear
- TBM** taxa bruta de mortalidade
- v.a.** variável aleatória
- VAR** vetores autorregressivos
- VARMA** vetores autorregressivos e de médias móveis
- WN** ruído branco
- $W(t)$ processo de Wiener

1

Introdução

1.1	Motivação: o estudo da tendência evolutiva da mortalidade humana e os dados da população portuguesa	2
1.2	Breve síntese da literatura e porque modelar a mortalidade com equações diferenciais estocásticas	6
1.3	Objetivos e organização da tese	8

1.1 Motivação: o estudo da tendência evolutiva da mortalidade humana e os dados da população portuguesa

Em Portugal, e na maioria dos países ocidentais, a estrutura etária da população tem vindo a modificar-se, acentuando-se cada vez mais o envelhecimento populacional devido à redução da natalidade e ao aumento da esperança de vida. De acordo com as estimativas da população residente de 2007 [41], a população idosa representava mais de 17% da população portuguesa. Os resultados definitivos do recenseamento de 2011 apontam para um aumento desse valor, com cerca de 19% de idosos [42], mas, segundo o Instituto Nacional de Estatística, este fenómeno tem tendência para se acentuar nas próximas décadas, podendo esse valor duplicar em algumas regiões do país [40].

Se é certo que o risco de mortalidade aumenta com a idade, as taxas de mortalidade têm vindo globalmente a diminuir, facto que tem conduzido ao estudo de fatores, intrínsecos e extrínsecos, suscetíveis de explicar essa evolução. Famílias de modelos, determinísticos ou, mais recentemente, estocásticos, têm vindo a ser testadas, dando origem, nomeadamente, a estudos comparativos para aferir qual o melhor modelo a aplicar neste contexto (veja-se, por exemplo, [5], [19] ou [33]).

Pelo exposto, e não obstante a mortalidade ser uma variável demográfica que tem vindo a ser estudada exaustivamente, o objetivo deste trabalho é desenvolver um modelo de equações diferenciais estocásticas (EDEs) que, através de uma análise transversal dos dados da mortalidade ao longo do tempo, nos permita inferir sobre a tendência futura do fenómeno de diminuição das taxas de mortalidade, para todos os grupos etários e por sexo, e fazer previsões a curto prazo ou passo-a-passo (PP) e também a médio/ longo prazo (LP).

Esta ideia surgiu numa fase inicial do plano de estudos do Programa de Doutoramento em Matemática, em que teve lugar uma análise exploratória exaustiva dos dados da mortalidade da população portuguesa. Os dados, obtidos através do *Human Mortality Database* [37], correspondem às taxas brutas de mortalidade (TBMs) e representam a razão entre o número de óbitos (total para o país num certo intervalo de tempo e para todas as causas de morte) e a uma estimativa da população residente que, por sua vez, corresponde à população exposta ao risco de óbito no mesmo intervalo idade-tempo. Para este estudo, seleccionámos, inicialmente, 200 séries temporais, com periodicidade anual, disponíveis entre 1940 e 2009, para 100 grupos etários anuais (que passamos a designar por idades) e por sexo, abrangendo o arco de vida dos 0 aos 99 anos. A idade 0 (que corresponde à primeira idade estudada) refere-se aos indivíduos que morreram no primeiro ano de vida, isto é, antes do primeiro ano completo, e assim sucessivamente até à idade 99; designámos, respetivamente, por F_0 e M_0 a TBM dos indivíduos do sexo feminino e do sexo masculino na idade 0 (e de modo análogo para as restantes idades).

Em demografia, é comum os dados estarem disponíveis por coorte (numa perspetiva longitudinal ao longo do tempo). Uma coorte representa um conjunto de indivíduos que nascem num mesmo ano e são acompanhados ao longo da vida. Não há nesse caso, em que se usa uma abordagem longitudinal ao longo do tempo, distinção entre idade e ano de calendário. Nesse contexto, é

muito difícil a modelação abarcar todas as idades do arco da vida humana, pois é necessário, geralmente, um número muito elevado de parâmetros para o efeito (com frequência superior a oito por cada coorte, por a trajetória da mortalidade ser muito irregular). A propósito desta abordagem, ver representação dos dados no gráfico da esquerda da figura 1.1. A curva descreve a evolução da mortalidade nas várias fases do arco da vida. Neste caso, foi fixado o ano de 1974, mas a forma, usualmente descrita na literatura como “curva em forma de banheira”, não se tem alterado significativamente ao longo do tempo – não obstante a redução da mortalidade infantil nas últimas décadas e a maior longevidade, esta forma caracteriza a mortalidade humana.

Em alternativa, a abordagem transversal que seguimos faz sentido, por considerarmos que existem fenómenos que, ao longo do tempo, afetam todas as idades. Entre outros, destacamos, pela positiva, as alterações nas condições de vida de natureza sócio-económica ou os avanços da medicina – como a introdução do plano nacional de saúde e do plano de vacinação (que, no caso das crianças, reduziu drasticamente a mortalidade). Também as alterações climáticas que geram fenómenos extremos ou outras situações de catástrofe, podem afetar globalmente uma população, neste caso aumentando a mortalidade. O fenómeno assim descrito tem uma forte tendência decrescente no período em análise e é mais fácil explicá-lo matematicamente com recurso a dois ou três parâmetros (ver gráfico da direita da figura 1.1). Em quase todas as idades, as TBMs são superiores nos indivíduos do sexo masculino relativamente aos do feminino, ainda que com uma evolução diferente em cada idade.

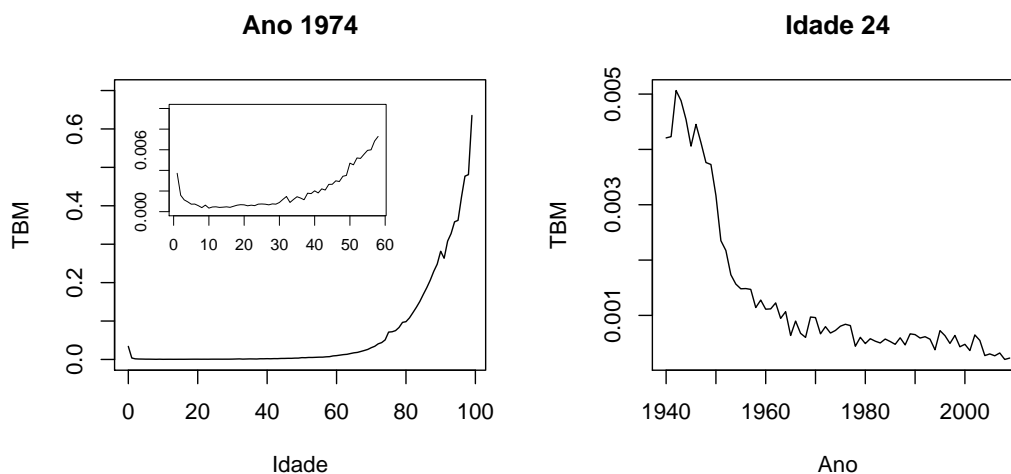


Figura 1.1: TBMs da população portuguesa (sexo feminino); representação longitudinal (idades 0 a 99) para o ano de 1974 (à esquerda, com ampliação das idades 2 a 59) e transversal (no período de 1940 a 2009) da idade 24 (à direita)

Os resultados e métodos são, pelo exposto, ilustrados através das TBMs da população portuguesa, que consideramos refletirem o comportamento da mortalidade nos países que já sofreram a transição demográfica (a propósito da evolução da mortalidade no contexto da transição demográfica, em Portugal e no mundo, ver, por exemplo, [61]). Na figura 1.2, mostramos o padrão

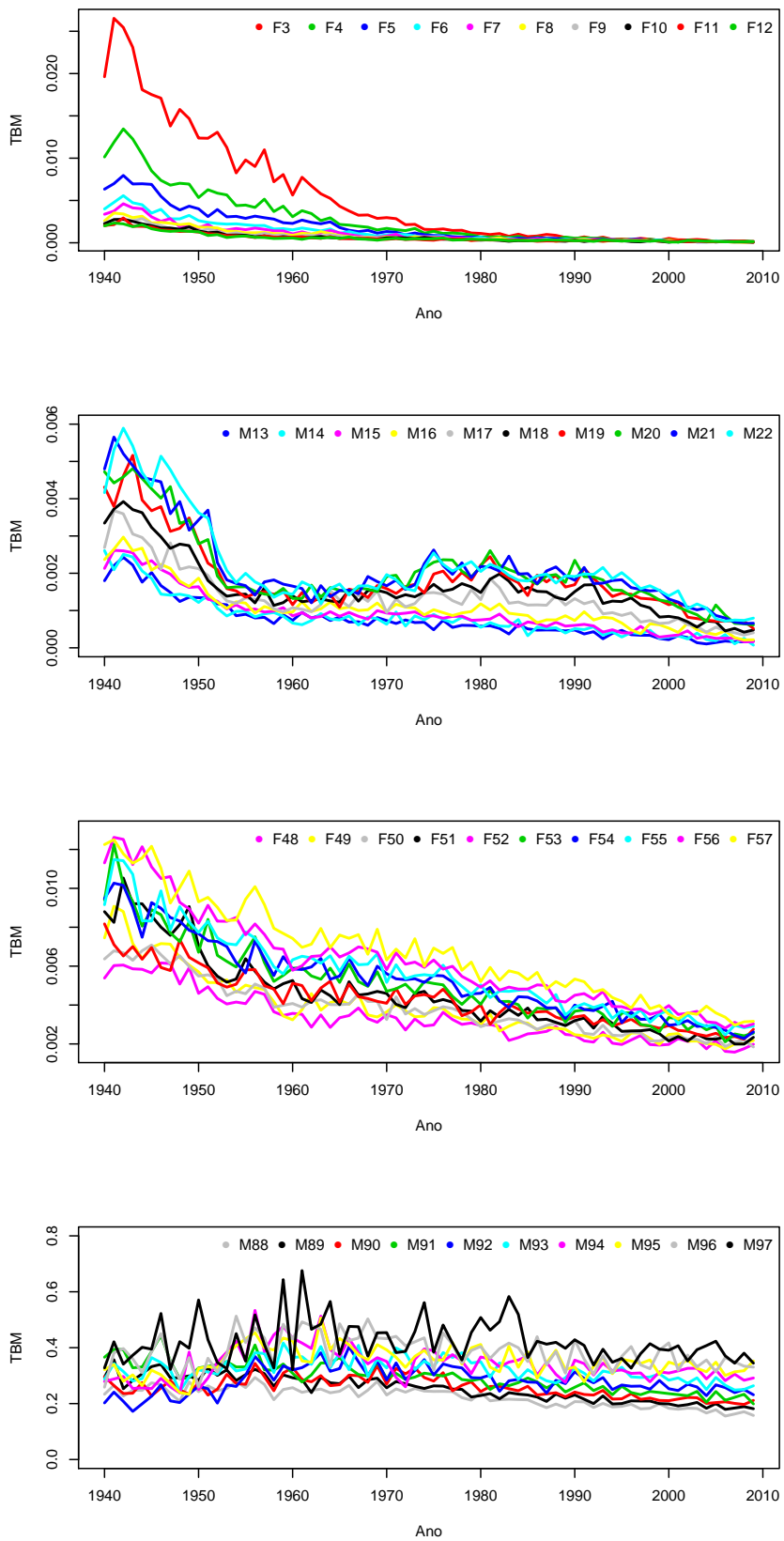


Figura 1.2: TBMs da população portuguesa das idades 3 a 12, 13 a 22, 48 a 57 e 88 a 97 (de cima para baixo, com F do sexo feminino e M do masculino), no período de 1940 a 2009

das TBMs em vários grupos de idades sucessivas (para ilustrar, genericamente, o comportamento dos dados). Representamos as séries que correspondem a idades de períodos diferentes do arco de vida dos dois sexos (no primeiro gráfico, acima, F3 a F12 representam as TBMs das idades 3 a 12 dos indivíduos do sexo feminino; analogamente, nos gráficos seguintes, com M13, por exemplo, a representar a TBM da idade 13 do sexo masculino). Com exceção das figuras que ilustram globalmente os dados (como o exemplo anterior), dividimos cada série temporal, que tem 70 observações (na realidade, são estimativas das TBMs anuais), em dois subconjuntos: observações de 1940 a 1999, para o ajustamento dos modelos, e de 2000 a 2009, para validar as previsões.

Antes de concluir esta secção, chamamos ainda a atenção para o facto de, em demografia, ser frequentemente objeto de estudo a variável força da mortalidade, habitualmente representada por μ . Sendo i uma certa idade, temos $\mu_i = -\ln(1 - q_i)$, com q_i a probabilidade de morte para um indivíduo com a idade i , de um dado sexo (estas questões são frequentes na construção de tábuas de mortalidade e estão descritas exaustivamente, por exemplo, em [65] ou [75]). Se considerarmos que essa probabilidade é constante entre as idades exatas i e $i + 1$ e num dado horizonte temporal (digamos num período anual), aproximamos q_i pela TBM correspondente à mesma idade e no mesmo período e podemos assim obter uma aproximação de μ (a propósito das medidas das estatísticas da mortalidade ver também [46] ou [57]). Na realidade, o desvio médio entre μ e as TBMs é bastante reduzido (estima-se que na ordem dos 10^{-5}) na maior parte das idades (no período em análise, apenas aumenta a partir da idade 85). Na figura 1.3, ilustramos essa diferença para a idade 63 do sexo masculino.

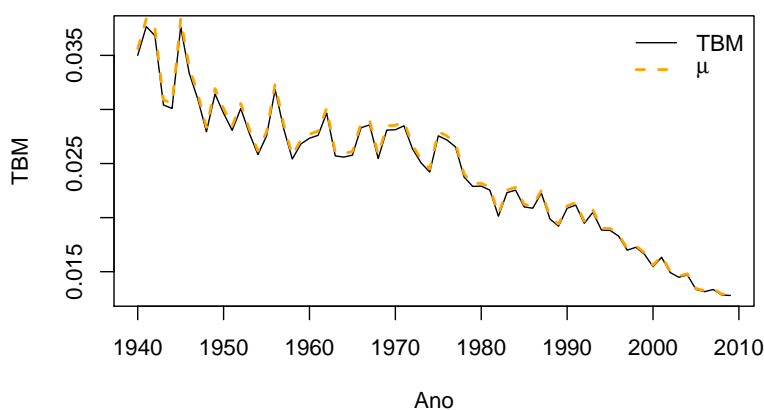


Figura 1.3: TBM vs força da mortalidade da idade 63 do sexo masculino, no período de 1940 a 2009

1.2 Breve síntese da literatura e porque modelar a mortalidade com equações diferenciais estocásticas

A evolução futura da esperança de vida é incerta, devido a fatores externos e à própria incerteza na tendência evolutiva das taxas de mortalidade enquanto fenómeno demográfico. Desde o século XIX, com os primeiros estudos de Gompertz, muito mudou na abordagem ao problema, que tem sido amplamente estudado. Inicialmente, os modelos não incorporavam incerteza - esta foi introduzida através da construção de tábuas de mortalidade [58, 59], estudando uma geração ou coorte, pelo que, apenas ao reconhecer-se essa necessidade foram desenvolvidos os primeiros modelos estocásticos (ou probabilísticos), que surgiram sobretudo desde a década de 90 do século passado, principalmente na perspetiva dos atuários, economistas e banca de investimento [36]. Destes, o *modelo de Lee-Carter*, de 1992 [51], é sem dúvida o mais conhecido, com muitas aplicações e variações (veja-se, por exemplo, [50], ou, do mesmo autor, um estudo do *Life Office Mortality Committee* [52], em que se faz um balanço de resultados da sua aplicação, ou ainda um outro [21] sobre as projeções da mortalidade em Portugal). Modelos paramétricos de séries temporais (como o ilustrado em [2]), ou outros como os de Cairns-Blake-Dowd, de 2006, Renshaw-Haberman, de 2003 ou de 2006), Olivier-Smith (Olivier-Jeffery, de 2004 e Smith, de 2005), modelos polinomiais ou de *p-splines* (a propósito destes modelos, ver [18, 23]) são exemplos de modelos estocásticos, todos eles procurando, para além do melhor ajustamento, sobretudo, as melhores previsões. Estes modelos, muito utilizados na análise longitudinal dos dados, são construídos a partir das leis da mortalidade, sendo que a componente estocástica que determina a evolução dos parâmetros é geralmente introduzida através de um termo que incorpora um processo estocástico (p.e.), cuja modelação é feita frequentemente por técnicas de análise de séries temporais (veja-se, por exemplo, as projeções para a população da Austrália, a partir dos dados entre 1921 a 2004, usando uma extensão do modelo de Hyndman e Ullah, de 2007 [39], em [38]).

Com uma longa aplicação no estudo do comportamento dos mercados financeiros, as EDEs, cujo *modelo de Black-Scholes* (anos 70 do século passado) veio impulsionar a investigação e o desenvolvimento de uma vasta teoria e aplicações a outras áreas da ciência, têm vindo a ser muito utilizadas na modelação do crescimento de populações (ver [8] e referências aí contidas e [25–28]). Recentemente, começaram a ser aplicados em Portugal modelos de EDEs variantes do *modelo de Ornstein-Uhlenbeck*, que incorporam também um termo com uma componente aleatória ambiental, a dados demográficos, designadamente no estudo longitudinal da mortalidade ou construção de tábuas de mortalidade dinâmicas - veja-se, por exemplo, a sua utilização na construção de tábuas de mortalidade prospetivas, aplicações atuariais e cobertura do risco de longevidade [12, 13] ou, no mesmo sentido, um estudo sobre tabelas dinâmicas, aplicado às taxas de mortalidade de Espanha [22]. Estes modelos permitem a introdução de aleatoriedade, que traduz os efeitos das variações ambientais nos coeficientes (logo, são mais realísticos), sendo possível, a partir da solução da equação, inferir sobre a sua distribuição de probabilidade. Das escassas referências que encontramos na literatura acerca da utilização de EDEs para modelar a mortalidade humana, e

ainda que numa perspectiva do estudo de coortes, destacamos o recente modelo de Jevtic, Luciano e Vigna, de 2013 [44], para uma superfície de mortalidade e com recurso a análise fatorial, também o modelo de Park de 2008, em que, para obter a probabilidade de sobrevivência, se estima a força de mortalidade através de um processo de difusão com saltos [69], e o modelo de Yashin *et al*, de 2007 [87], em que a mortalidade é função de fatores de risco, que se alteram com a idade e que são traduzidos por uma EDE com processos de Wiener ($W(t)$) independentes.

Embora o conjunto de modelos em tempo discreto do tipo Lee-Carter, que incorporam geralmente a componente estocástica num único termo, se revele bom no curto prazo (os parâmetros precisam geralmente de ser reajustados para projeções a médio-longo prazo), os novos modelos de EDEs em tempo contínuo trazem vantagens adicionais, uma vez que associam a incerteza à dinâmica do processo. A sua construção baseia-se nos modelos determinísticos de equações diferenciais ordinárias, incorporando o efeito da variabilidade ambiental na evolução das taxas.

Os modelos de EDEs que nos propomos construir e aplicar são modelos que se pretendem simples e flexíveis (embora com parâmetros diferenciados por idade). Admitindo que o sistema demográfico não evolui de modo independente dos sistemas económicos e ambientais [60], as taxas de mortalidade têm assim flutuações estocásticas em função do “ambiente”, em sentido lato (como já referimos), acrescentando que, para além da aleatoriedade ambiental (ou sistemática), as TBMs observadas têm também um erro amostral associado (aleatoriedade demográfica), que não é objeto de estudo nesta tese. Trata-se de um erro que, em termos relativos, é reduzido e por isso não é tratado, sendo que tem apenas alguma expressão nas idades mais avançadas (porque a “amostra”, isto é, a população em risco, é menor).

Em abordagem anterior, já referida [12], utilizaram-se modelos longitudinais de EDEs, de modo a explicar a evolução de uma coorte fictícia (idade e tempo a evoluir conjuntamente) e obtiveram-se bons resultados para idades avançadas. Contudo, a abordagem longitudinal tem limitações, pois, à partida, há que seleccionar um período de tempo/ idade restrito, dado o comportamento muito complexo da taxa de mortalidade com a idade quando se considera todo o arco da vida humana. A nossa abordagem, transversal ao longo do tempo, pelo contrário, modela a evolução da taxa de mortalidade de uma certa idade (fixa ao longo do tempo), que tem um comportamento relativamente regular.

Pelo exposto, como os dados mostram uma evolução dinâmica das TBMs ao longo do tempo (e não meramente amostral), faz sentido construir e aplicar modelos com uma componente ambiental aleatória, daí a utilização de modelos de EDEs. Acerca das potencialidades de utilização destes modelos, em que se procura explicar a variabilidade da mortalidade de uma forma simples e credível para fins de planeamento (por exemplo, pensões, poupanças, planos de saúde ou seguros), destaca-se o facto de podermos ainda converter posteriormente os resultados em termos de grandezas derivadas como a esperança de vida ou taxas de sobrevivência, variáveis aleatórias também dependentes das condições ambientais, estudando problemas complementares ou afins e introduzindo até variáveis explicativas exteriores ao sistema da mortalidade.

1.3 Objetivos e organização da tese

Considerando a problemática que constituiu o ponto de partida para a investigação (conforme exposto na secção 1.1), esta dissertação procura dar resposta à questão: qual a tendência futura, a curto/ médio prazo, de evolução das TBMs da população humana, por idade e por sexo e, em particular, como é que as previsões dadas por modelos de EDEs se comportam, se considerarmos o efeito de correlações nas TBMs entre sexos diferentes, para a mesma idade, e entre idades diferentes, quer entre sexos diferentes, quer dentro do mesmo sexo. Na figura 1.4, sintetizamos a abordagem metodológica que usamos para responder à problemática.

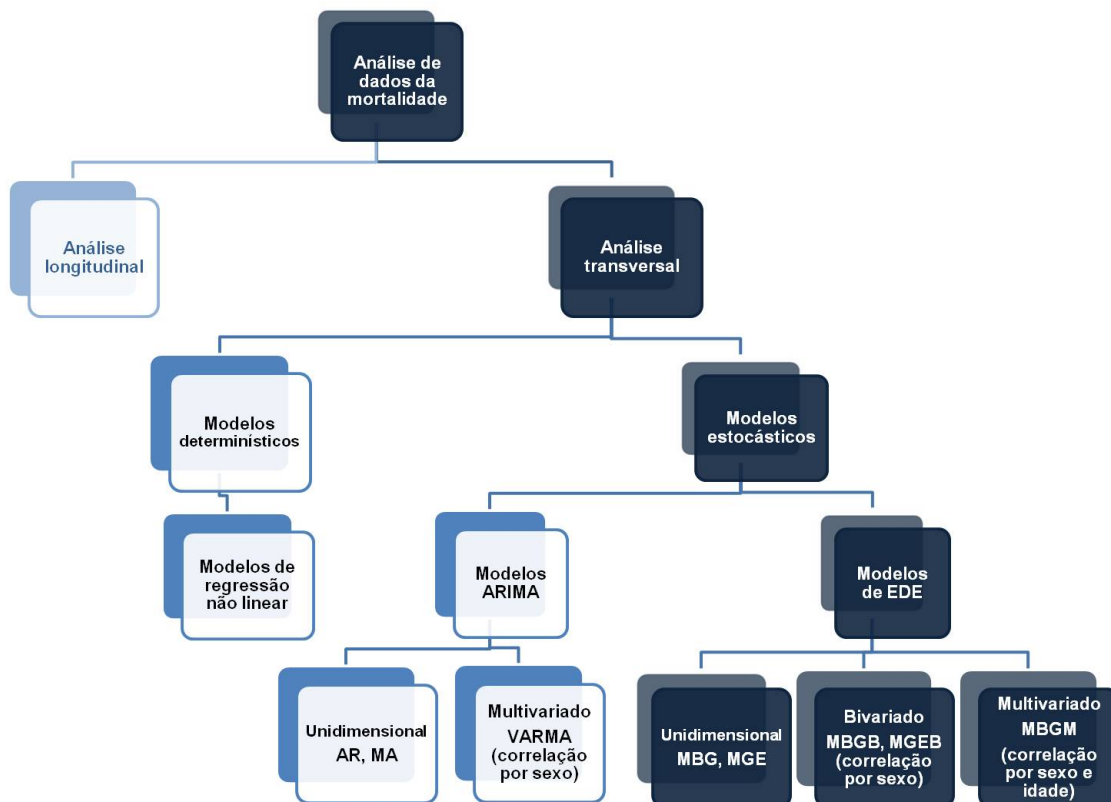


Figura 1.4: Diagrama metodológico

Esta dissertação é composta por 6 capítulos. No primeiro capítulo, apresentamos os aspetos motivacionais que conduziram à identificação do problema e seleção da metodologia e faz-se uma breve revisão da literatura sobre modelos de mortalidade, de modo a enquadrar a temática.

No segundo capítulo, fazemos uma breve exposição conceptual e metodológica acerca da teoria das EDEs necessária no desenvolvimento dos capítulos seguintes.

Os terceiro, quarto e quinto capítulos são os capítulos centrais da tese e referem-se à modelação das TBMs através de EDEs. No capítulo 3, aplicamos modelos de EDEs univariados aos dados. Faz-se o estudo analítico dos modelos apresentados e comparam-se os resultados obtidos através do movimento browniano geométrico (MBG) e do modelo de Gompertz estocástico (MGE) com os obtidos por outros métodos de análise transversal de dados, nomeadamente, baseados em

métodos de regressão não linear (RNL) e modelos clássicos de análise de séries temporais, concretamente, modelos autorregressivos integrados e de médias móveis (ARIMA), no sentido de justificar a consistência e vantagens da metodologia.

No capítulo 4, apresentamos dois novos modelos bivariados de EDEs, considerando que existe uma estrutura de correlação associada às taxas de mortalidade de sexos diferentes para a mesma idade. Designámos por movimento browniano geométrico bidimensional (MBGB) o primeiro dos modelos apresentados e por modelo de Gompertz estocástico bidimensional (MGEB) o segundo. Em ambos os casos, aplica-se um modelo de EDEs e compara-se o modelo completo, com correlação entre os processos de Wiener unidimensionais relativos a cada sexo, com o modelo sem correlação. À semelhança do estudo para os modelos unidimensionais, comparamos ainda os resultados da modelação das TBMs através dos modelos bivariados de EDEs (no caso, do MBGB) com outros modelos, designadamente modelos de vetores autorregressivos e de médias móveis (VARMA).

No capítulo 5, propomos um modelo multivariado de EDEs, considerando que existe uma estrutura de correlação, tendo em conta as dimensões sexo e idade. Aplicamos o novo modelo que designámos por movimento browniano geométrico multidimensional (MBGM) e em que consideramos processos de Wiener correlacionados entre idades dentro de cada sexo e também entre sexos diferentes. Comparam-se, neste caso, os resultados de diferentes versões do modelo proposto.

Nos capítulos 3, 4 e 5, serão tratados, com recurso a exemplos, os aspetos estatísticos de seleção dos modelos, estimação e previsão, bem como os respetivos intervalos de confiança. Quanto à calibração dos modelos de EDEs, na estimação dos parâmetros será utilizado o método de máxima verosimilhança (MV). Considerando a validação dos modelos, para além da comparação com modelos congéneres alternativos ou entre versões dos modelos, serão usadas medidas de avaliação de desempenho e o estudo da capacidade preditiva.

No capítulo 6, apresentamos um resumo das conclusões desta dissertação e fazemos algumas considerações sobre o trabalho futuro.

Por último, refira-se que, não obstante se terem utilizado ocasionalmente vários programas de computador para cálculo ou representação gráfica de dados e resultados, como o Mathematica, o Maple, o IBM SPSS Statistics ou o Microsoft Office Excel, o trabalho de programação foi desenvolvido no programa estatístico R (disponível, com acesso livre, em <http://www.r-project.org/>). Ao longo dos capítulos, apresentamos, sempre que se justificar, algumas caixas destacadas do texto com resultados da compilação do código R. Os algoritmos originais (ou excertos de algoritmos) escritos em puro código R, dos modelos de EDEs, são apresentados em apêndices que seguem a ordenação dos capítulos. Note-se que, pela sua extensão, salvo casos pontuais, não apresentamos o código relativo aos gráficos (para o efeito, utilizámos, entre outros, [56, 63] ou os portais <http://www.statmethods.net/graphs/index.html> e <http://research.stowers-institute.org/efg/R/>). No apoio geral à programação em R, usámos, para além das referências em áreas específicas que iremos mencionar ao longo dos capítulos, as referências [1, 11, 89].

2

Breve introdução às equações diferenciais estocásticas

2.1	Introdução	12
2.2	Processos estocásticos	12
2.3	Equações diferenciais estocásticas	13
2.4	Fórmula de Itô	15
2.5	Exemplo: a lei de Gompertz	15

2.1 Introdução

Apresentamos, de seguida, uma breve exposição de conceitos, propriedades e aspetos numéricos, relativos à teoria das equações diferenciais estocásticas (EDEs). Todos estes tópicos, alguns dos quais baseados na teoria das probabilidades ou com origem na análise matemática, estão exaustivamente enunciados e demonstrados na bibliografia de referência, nomeadamente em [3, 9, 12, 25, 36, 62, 66, 68, 81].

2.2 Processos estocásticos

À partida, consideramos que o fenómeno que vamos estudar não é puramente determinístico, pois ao observar, no nosso caso, as séries temporais das TBMs e as suas variações ao longo do tempo, constatamos que estas sofrem perturbações aleatórias que não conseguimos prever. As taxas de mortalidade têm assim aquilo que designamos por um comportamento estocástico. Estes processos podem ser modelados à custa de conjuntos de **variáveis aleatórias (v.a.)** que descrevam o sistema em estudo em cada instante de tempo, t , com $t \in T$ (normalmente, $T = \mathbb{R}^+$ ou $T = \mathbb{N}$, isto é, em modo contínuo ou discreto), e que dependem também do acaso, ω , com $\omega \in \Omega$, o conjunto de todos os resultados possíveis para um acontecimento (ou evento aleatório) ou estados possíveis da natureza (em sentido lato), suscetíveis de perturbar esse mesmo fenómeno. O nosso objetivo é pois introduzir uma fonte de ruído num modelo de modo a captar ou a explicar melhor as oscilações aleatórias de um dado processo ao longo do tempo. O fenómeno assim descrito, e que traduz a evolução temporal de um conjunto de v.a., $\{X(t)\}_{t \in T}$, é um **processo estocástico (p.e.)** indexado por T , que designamos apenas por $X(t)$ como abreviatura de $X(t, \omega)$ para simplificar a notação. A partir de agora, assumimos $T = [0, \infty[$, logo o p.e. é em tempo contínuo, e também a variável de estado é contínua (pois a variável pode mudar de valor em qualquer instante de tempo e pode tomar qualquer valor real). Um p.e. indexado por T é uma família de variáveis aleatórias, todas elas definidas sobre o mesmo **espaço de probabilidade** $(\Omega, \mathcal{F}, \mathcal{P})$, com \mathcal{P} a medida de probabilidade e \mathcal{F} uma álgebra- σ sobre Ω .

Existem várias classificações para os p.e., dependendo das características das v.a. que os definem, do conjunto T considerado e do próprio espaço de estados Ω . Refira-se, a este propósito, que todos os p.e. que vamos usar neste estudo, bem como as soluções das EDEs apresentadas, podem ser considerados processos de difusão e processos de Markov. O processo de Wiener ($W(t)$), fundamental para a construção de EDEs (pois pode traduzir o efeito acumulado das oscilações ambientais sobre um dado fenómeno, até um certo instante t considerado) é um processo de difusão e um processo de Markov homogéneo.

Seja B um *conjunto de Borel*, tal que $B \in \mathcal{B}$, com \mathcal{B} a álgebra- σ de Borel representando a mais pequena álgebra- σ que contém os intervalos contidos em T . $X(t)$ é um **processo de Markov** se, para todo o $s, t \in T$ com $s < t$ e para qualquer conjunto de Borel B ,

$$P[X(t) \in B | X(u), 0 \leq u \leq s] = P[X(t) \in B | X(s)].$$

Esta propriedade, conhecida por *propriedade de Markov*, diz-nos que, conhecendo o valor presente do processo, os seus valores futuros são independentes de valores passados. Se um processo de Markov tiver probabilidades de transição estacionárias (no tempo), isto é,

$$P[X(t + \tau) \in B | X(s + \tau) = x] = P[X(t) \in B | X(s) = x],$$

diz-se um processo de Markov homogéneo.

Um p.e. $X(t)$ com momentos de segunda ordem designa-se **processo de difusão** se se verifica a propriedade de Markov e se, adicionalmente, apresenta quase certamente (*q.c.*) trajetórias contínuas e existem, para $\epsilon > 0$, $x \in \mathbb{R}$ e $s \in [0, d]$, com convergências uniformes, os limites

$$\lim_{\Delta \rightarrow 0^+} \frac{1}{\Delta} P_{s,x} [|X(s + \Delta) - x| > \epsilon] = 0,$$

$$\lim_{\Delta \rightarrow 0^+} E_{s,x} \left[\frac{X(s + \Delta) - x}{\Delta} \right] = a(s, x)$$

e

$$\lim_{\Delta \rightarrow 0^+} E_{s,x} \left[\frac{(X(s + \Delta) - x)^2}{\Delta} \right] = b(s, x),$$

onde $P_{s,x}$ é a probabilidade condicional a $X(s) = x$ e $E_{s,x}$ representa a esperança matemática condicional a $X(s) = x$. A definição pode ser generalizada para processos de segunda ordem.

A $a(s, x)$ e $b(s, x)$, que correspondem, respetivamente, aos momentos infinitesimais de primeira e segunda ordem, chamam-se coeficiente de tendência ou média infinitesimal e coeficiente de difusão ou variância infinitesimal. Se esses coeficientes não dependerem de t , o processo de difusão diz-se homogéneo.

Um **processo de Wiener** (ou **processo de Wiener padrão**), $W(t)$, é um processo de difusão homogéneo que verifica as propriedades:

- $W(0) = 0$ *q.c.*;
- os incrementos $\Delta W(t) = W(t) - W(s)$ (com $s < t$) têm uma distribuição normal com média zero e variância $t - s$;
- os incrementos $W(t_i) - W(s_i)$ (com $i = 1, \dots, n$ e $0 \leq s_1 < t_1 \leq s_2 < t_2 \leq \dots \leq s_{n-1} < t_{n-1} \leq s_n < t_n$), em intervalos de tempo não sobrepostos, são independentes.

Note-se que $W(t)$ tem distribuição normal com média zero e variância t , isto é, $W(t) \sim \mathcal{N}(0, t)$, pois $W(t) - W(0) = W(t)$ é o incremento no intervalo $[0, t]$. Também se verifica que $Cov[W(s), W(t)] = E[W(s)W(t)] = \min(s, t)$.

2.3 Equações diferenciais estocásticas

Normalmente, obtemos uma **equação diferencial estocástica**, EDE, a partir de uma equação diferencial determinística a que adicionamos um termo de ruído com o objetivo de descrever as flutuações aleatórias que afetam o fenómeno em estudo. Admitindo que o valor acumulado até ao

instante t dessas flutuações aleatórias pode ser descrito por um processo de Wiener padrão, $W(t)$, a EDE toma a forma diferencial

$$dX(t) = f(t, X(t))dt + g(t, X(t))dW(t), \quad (2.1)$$

com condição inicial $X(0) = X_0$ que supomos ser uma v.a. independente de $W(t)$, onde f e g são funções reais. Uma solução $X(t) = X(t, \omega)$ da equação (2.1) é um p.e. que verifique a equação integral

$$X(t) = X(0) + \int_0^t f(s, X(s))ds + \int_0^t g(s, X(s))dW(s), \quad (2.2)$$

mais explicitamente

$$X(t, \omega) = X(0, \omega) + \int_0^t f(s, X(s, \omega))ds + \int_0^t g(s, X(s, \omega))dW(s, \omega),$$

com os integrais definidos como iremos descrever a seguir.

Seja $F(s, \omega) = f(s, X(s, \omega))$ e $G(s, \omega) = g(s, X(s, \omega))$. O integral $\int_0^t F(s, \omega)ds$ pode considerar-se, para cada ω fixo, um *integral de Riemann*. Já o integral $\int_0^t G(s, \omega)dW(s, \omega)$ não pode ser definido como um *integral de Riemann-Stieltjes* pois diferentes somas de Riemann-Stieltjes convergem para limites diferentes. Trabalhamos com funções $G(s, \omega)$ não-antecipativas com norma L^2 finita, isto é, $(\|G\|_2)^2 = E[\int_0^t |G(s, \omega)|^2 ds] < +\infty$. A função $G(s, \omega)$ diz-se não-antecipativa se for conjuntamente mensurável em s e ω e for independente dos incrementos futuros dos processos de Wiener. Para $G \in L^2$ usamos o integral de Itô, que se define como o limite em média quadrática das somas de Riemann-Stieltjes, isto é,

$$l.i.m._{n \rightarrow +\infty} \sum_{k=1}^n G(t_{k-1})(W(t_k) - W(t_{k-1})),$$

onde $0 = t_{0,n} \leq t_{1,n} \leq \dots \leq t_{n,n} = t$ ($n = 1, 2, \dots$) são decomposições do intervalo $[0, t]$ cujo diâmetro tende para 0 quando $n \rightarrow +\infty$. Note-se que as somas de Riemann-Stieltjes utilizam como ponto intermédio o ponto inicial de cada intervalo da decomposição. Outras escolhas de pontos intermédios dariam outros tipos de integral, mas a escolha feita (não antecipativa), que conduz ao integral de Itô, tem a vantagem de produzir propriedades bastante interessantes do integral. Esta definição pode estender-se à classe das funções G não antecipativas tais que $\int_0^t |G(s)|^2 ds < +\infty$ q.c..

O grande impulsionador, quer para as definições, quer para o que se passou a designar por cálculo estocástico, foi Kiyoshi Itô, matemático japonês que desenvolveu nos anos 40 do século passado as bases para a teoria das EDEs.

Identificando funções quase iguais, L^2 é um espaço de Hilbert. Das propriedades dos integrais estocásticos, destacamos as seguintes, considerando o intervalo de integração $[0, t]$, $a, b \in \mathbb{R}$ e $G, G_1, G_2 \in L^2$:

- $\int_0^t dW(s) = W(t) - W(0)$;
- $\int_0^t (aG_1(s) + bG_2(s))dW(s) = a \int_0^t G_1(s)dW(s) + b \int_0^t G_2(s)dW(s)$;
- $E[\int_0^t G(s)dW(s)] = 0$;

- $E[(\int_0^t G(s)dW(s))^2] = E[\int_0^t G^2(s)dt]$;
- $E[\int_0^t G_1(s)dW(s) \int_0^t G_2(s)dW(s)] = E[\int_0^t G_1(s)G_2(s)dt]$.

Voltando a (2.2), caso f e g satisfaçam as propriedades adequadas (ver, por exemplo, [9]), a solução existe e é única e é um processo de difusão com coeficiente de tendência $a(s, x) = f(s, x)$ e coeficiente de difusão $b(s, x) = g^2(s, x)$. Quando f e g não dependem do tempo, como sucede neste trabalho, a EDE diz-se autônoma e a sua solução é uma difusão de Itô.

2.4 Fórmula de Itô

Um processo $X(t)$,

$$X(t, \omega) = X(0, \omega) + \int_0^t F(s, \omega)ds + \int_0^t G(s, \omega)dW(s, \omega), \quad (2.3)$$

com $X(0) = X_0$ independente de $W(t)$ e F e G mensuráveis em s e ω , que verifiquem, *q.c.*, as propriedades:

$$\int_0^t G^2(s)ds < +\infty$$

e

$$\int_0^t |F(s)|ds < +\infty,$$

diz-se um **processo de Itô**.

Seja $X(t)$ um processo de Itô. Se $Y(t) = h(t, X(t))$, com $h(t, x)$ de classe $C^{1,2}$ (isto é, com derivada parcial de primeira ordem contínua em t e derivada parcial de segunda ordem contínua em x), então $Y(t) = Y(t, \omega)$ é ainda um processo de Itô. A fórmula de Itô (que se refere à *regra de diferenciação de uma função composta* ou *regra da cadeia*), pode ser dada, relativamente a $Y(t)$, na forma

$$dY(t) = \frac{\partial h(t, X(t))}{\partial t} dt + \frac{\partial h(t, X(t))}{\partial x} dX(t) + \frac{1}{2} \frac{\partial^2 h(t, X(t))}{\partial x^2} (dX(t))^2, \quad (2.4)$$

usando-se, no terceiro termo, as igualdades

$$dtdt = 0$$

$$dtdW(t) = dW(t)dt = 0$$

$$dW(t)dW(t) = dt.$$

2.5 Exemplo: a lei de Gompertz

Se atendermos ao nosso objeto de estudo, um exemplo de modelo determinístico, que pode traduzir a lei de Gompertz para a mortalidade, pode ser dado por

$$\frac{dX(t)}{dt} = bX(t) \ln \left(\frac{a}{X(t)} \right), \quad (2.5)$$

com $X(t)$ a taxa de mortalidade (que varia com o tempo) dos indivíduos de uma certa idade e sexo (que por agora assumimos como fixos), onde a representa a taxa de mortalidade assintótica e b é uma taxa de aproximação ao regime assintótico.

Por comodidade de cálculo, usamos $Y(t) = \ln(X(t))$ e $A = \ln(a)$ e obtemos a equação, equivalente à equação (2.5)

$$\frac{dY(t)}{dt} = -b(A - Y(t)). \quad (2.6)$$

Para obter o modelo de Gompertz estocástico (MGE), introduzimos, em (2.6), uma fonte de ruído, $\epsilon(t)$, onde $\epsilon(t) = \frac{dW(t)}{dt}$ é o ruído branco padrão. O processo de Wiener $W(t)$ reflete o efeito acumulado das perturbações “ambientais” sobre o fenómeno da mortalidade, até um certo instante t considerado e o coeficiente σ mede a intensidade da variabilidade ambiental resultante das perturbações aleatórias que afetam a variável Y em torno da sua tendência dinâmica. Obtemos assim a EDE autónoma

$$\frac{dY(t)}{dt} = -b(A - Y(t)) + \sigma\epsilon(t), \quad (2.7)$$

com o valor inicial, $Y(0) = y_0$, suposto conhecido.

Em vez de escrevermos a equação (2.7) com recurso à derivada, no sentido das funções generalizadas, de $W(t)$, também se pode escrever usando a notação mais usual

$$dY(t) = -b(Y(t) - A)dt + \sigma dW(t). \quad (2.8)$$

Considerando a forma genérica de EDE apresentada em (2.1) (neste caso para um p.e. $Y(t)$), $f(t, y) = -b(y - A)$ e $g(t, y) = \sigma$.

Seja $Z(t) = \exp\{bt\}(Y(t) - A)$. A solução da equação (2.8) (que vamos usar na subsecção 3.3.1 e cuja resolução apresentamos, neste ponto, a título ilustrativo), obtém-se aplicando a fórmula de Itô, (2.4), a $h(t, y) = \exp\{bt\}(y - A)$ e notando que $Z(t) = h(t, Y(t))$. Vem

$$\begin{aligned} dZ(t) &= b \exp\{bt\}(Y(t) - A)dt + \exp\{bt\}dY(t) + \frac{1}{2}0(dY(t))^2 \\ &= b \exp\{bt\}(Y(t) - A)dt + \exp\{bt\}(-b(Y(t) - A)dt + \sigma \exp\{bt\}dW(t)) \\ &= \sigma \exp\{bt\}dW(t). \end{aligned}$$

Integrando, no intervalo $[0, t]$, vem

$$\int_0^t dZ(s) = \int_0^t \sigma \exp\{bs\}dW(s),$$

donde

$$Z(t) = Z(0) + \sigma \int_0^t \exp\{bs\}dW(s).$$

Invertendo a transformação $Z(t) = \exp\{bt\}(Y(t) - A)$, vem então

$$\exp\{bt\}(Y(t) - A) = y_0 - A + \sigma \int_0^t \exp\{bs\}dW(s),$$

pelo que

$$Y(t) = A + (y_0 - A) \exp\{-bt\} + \sigma \exp\{-bt\} \int_0^t \exp\{bs\}dW(s).$$

Dado que a função integranda é determinística, $\int_0^t \exp\{bs\}dW(s)$ tem distribuição normal com média 0 e variância $\int_0^t (\exp\{bs\})^2 ds$, isto é,

$$\mathcal{N}\left(0, \int_0^t (\exp\{bs\})^2 ds\right) = \mathcal{N}\left(0, \frac{\sigma^2}{2b}(1 - \exp\{-2bt\})\right),$$

pelo que

$$Y(t) \sim \mathcal{N}\left(A + (y_0 - A) \exp\{-bt\}, \frac{\sigma^2}{2b}(1 - \exp\{-2bt\})\right).$$

De $Y(t) = \ln X(t)$ resulta que, a solução para $X(t)$ é dada pela expressão,

$$X(t) = \exp\left\{A + (\ln x_0 - A) \exp\{-bt\} + \sigma \exp\{-bt\} \int_0^t \exp\{bs\}dW(s)\right\}.$$

Obviamente que, como $Y(t) = \ln X(t)$ tem distribuição normal, $X(t)$ tem distribuição log-normal.

3

Modelos univariados de equações diferenciais estocásticas para taxas de mortalidade

3.1	Introdução	20
3.2	Movimento browniano geométrico	20
3.3	Modelo de Gompertz estocástico	30
3.4	Comparação dos resultados entre modelos	39
3.5	Previsões longitudinais em modelos de EDEs	56

3.1 Introdução

Na análise transversal da mortalidade humana, consideramos que há que ter em conta as flutuações aleatórias das condições ambientais, pelo que utilizamos EDEs para modelar as TBMs da população portuguesa.

A partir da análise meramente preliminar das séries temporais das TBMs, no sentido de se encontrarem os primeiros resultados para viabilizar o plano de tese, constatámos que modelos relativamente simples - com dois ou três parâmetros - permitem obter já resultados promissores (que, inclusive, captam a variabilidade das séries nas idades mais avançadas - geralmente mais difíceis de modelar - e permitem fazer previsões); no caso, ajustámos vários modelos unidimensionais, por idade e por sexo. De seguida, apresentamos a aplicação às TBMs do movimento browniano geométrico (MBG) e do modelo de Gompertz estocástico (MGE).

Por ser uma abordagem inovadora, para inferir acerca da adequabilidade dos modelos de EDEs a este tipo de dados, os resultados serão comparados com outros modelos, que em termos metodológicos são também potencialmente adequados, nomeadamente modelos de regressão não linear (RNL) e modelos para séries temporais do tipo ARIMA.

3.2 Movimento browniano geométrico

O movimento browniano geométrico (MBG) é um processo habitualmente utilizado para modelar preços de ações e outras variáveis económicas. Este é a solução da EDE conhecida como *modelo de Black-Scholes*, também designada equação de difusão de Black-Scholes (com coeficientes de tendência e difusão proporcionais à variável de estado, sendo os coeficientes de proporcionalidade aqui designados, respetivamente, μ e σ), dada por

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t), \quad (3.1)$$

com $\sigma > 0$.

Neste caso, $X = X(t)$ pode ser o preço de ativos, mas esta fórmula tem múltiplas aplicações, não só a produtos derivados dos mercados financeiros como também no crescimento de populações [14]. Usando a condição inicial $X(0) = x_0 > 0$, a sua solução, $X(t)$, é, como veremos, o p.e.

$$X(t) = x_0 \exp \left\{ \left(\mu - \frac{1}{2} \sigma^2 \right) t + \sigma W(t) \right\}, \quad (3.2)$$

conhecido por MBG.

Consideremos que as TBMs da população portuguesa seguem também um MBG e tomemos como ponto de partida para a modelação a equação (3.1) ou a sua solução (3.2). A este propósito, note-se que, de facto, quando observamos as séries parece haver uma tendência exponencial decrescente das TBMs ao longo do tempo.

3.2.1 O MBG

Seja $X_k(t)$ a TBM dos indivíduos de uma certa idade i ($i = 1, \dots, 100$) e sexo j ($j = 1$ para o feminino; $j = 2$ para o masculino), no instante t , com $k = i + 100(j - 1)$ para abranger todas as idades do arco da vida e de ambos os sexos. Para simplificar a notação, usamos ao longo de toda a secção simplesmente $X(t)$ em vez de $X_k(t)$, aplicando o modelo a cada idade e por sexo. Suponhamos conhecida a condição inicial $X(t) = x_0$. Fazemos $Y(t) = h(t, X(t)) = \ln(X(t)/x_0)$, com $X(t)$ como em (3.2). $h(t, x) = \ln(x/x_0)$ é uma função estritamente crescente de classe C^2 em x e podemos aplicar a fórmula de Itô (2.4) da secção 2.5, obtendo a EDE

$$dY(t) = Rdt + \sigma dW(t) \quad (3.3)$$

com $Y(0) = 0$, onde $R = \mu - \sigma^2/2$. Note-se que, como convencionámos usar $X(t)$ em vez de $X_k(t)$, o mesmo se aplica aos parâmetros do modelo, que podíamos escrever como R_k e σ_k , representando R_k a taxa média de crescimento de $Y_k(t)$ e σ_k o efeito das flutuações ambientais na dinâmica da mortalidade.

A solução da equação (3.3), para cada idade e sexo, no instante t , é dada por

$$Y(t) = Rt + \sigma W(t), \quad (3.4)$$

com distribuição marginal normal, com média Rt e variância $\sigma^2 t$, isto é,

$$Y(t) \sim \mathcal{N}(Rt, \sigma^2 t), \quad (3.5)$$

donde $X(t)$ tem distribuição log-normal com valor esperado $\mathbb{E}[X(t)] = X_0 \exp\{\mu t\}$. Podemos escrever (3.4) na escala original, tal que

$$X(t) = x_0 \exp\{Rt + \sigma W(t)\}.$$

Note-se que a equação (3.3) é uma EDE autónoma e que a sua solução (3.4) é uma difusão de Itô e um processo de difusão homogêneo com coeficientes de tendência R e de difusão σ^2 .

3.2.2 Estimação

De (3.5) resulta que a função densidade de probabilidade (f.d.p.), $f(t, y)$, de $Y(t)$ é dada por

$$f(t, y) = \frac{1}{\sqrt{2\pi V t}} \exp\left\{-\frac{1}{2} \frac{(y - Rt)^2}{V t}\right\},$$

com $V = \sigma^2$.

Sejam $t_n = t_0 + n$ ($n = 0, 1, 2, \dots, N$) os anos em que as TBMs foram observadas, para cada idade e por sexo (neste caso, todas as séries têm a mesma dimensão). Consideramos que $Y(t_0) = 0$ e

$$Y(t_n) = Y(t_{n-1}) + R(t_n - t_{n-1}) + \sigma(W(t_n) - W(t_{n-1})), \quad (3.6)$$

pelo que, condicionado a $Y(t_{n-1})$, $Y(t_n)$ tem distribuição normal com média $Y(t_{n-1}) + R(t_n - t_{n-1})$ e variância $V(t_n - t_{n-1})$ (uma vez que $Y(t_{n-1})$ é independente de $W(t_n) - W(t_{n-1})$).

Logo, a f.d.p. de transição de $Y(t)$ entre t_{n-1} e t_n é dada por

$$f(Y(t_n)|Y(t_{n-1})) = \frac{1}{\sqrt{2\pi V(t_n - t_{n-1})}} \exp \left\{ -\frac{1}{2} \frac{(Y(t_n) - Y(t_{n-1}) - R(t_n - t_{n-1}))^2}{V(t_n - t_{n-1})} \right\}. \quad (3.7)$$

Note-se que R e V são, respetivamente, a média e a variância das séries dos logaritmos dos retornos das TBMs, $\ln(X(t_n)/X(t_{n-1})) = Y(t_n) - Y(t_{n-1})$. O vetor de parâmetros $\mathbf{p} = (R, V)$ pode ser estimado por máxima verosimilhança (MV). Por $Y(t)$ ser um processo de Markov, a função de log-verosimilhança, \mathcal{L} , dadas as observações $Y(t_1), \dots, Y(t_N)$, pode escrever-se como

$$\begin{aligned} \mathcal{L}(\mathbf{p}|Y(t_1), \dots, Y(t_N)) &= \sum_{n=1}^N \ln(f(Y(t_n)|Y(t_{n-1}))) \\ &= -\frac{N}{2} \ln(2\pi V) - \frac{1}{2} \sum_{n=1}^N \ln(t_n - t_{n-1}) \\ &\quad - \frac{1}{2} \sum_{n=1}^N \frac{(Y(t_n) - Y(t_{n-1}) - R(t_n - t_{n-1}))^2}{V(t_n - t_{n-1})}. \end{aligned} \quad (3.8)$$

Podemos obter as expressões explícitas dos estimadores de MV dos parâmetros (ver [70]), resolvendo o sistema de equações

$$\begin{cases} \frac{\partial \mathcal{L}(y; \mathbf{p})}{\partial R} \Big|_{\hat{R}, \hat{V}} = 0 \\ \frac{\partial \mathcal{L}(y; \mathbf{p})}{\partial V} \Big|_{\hat{R}, \hat{V}} = 0. \end{cases} \quad (3.9)$$

Obtemos, para $t_n - t_{n-1}$ constantes,

$$\hat{R} = \frac{Y(t_N)}{t_N}$$

e

$$\hat{V} = \frac{1}{N} \sum_{n=1}^N \frac{(Y(t_n) - Y(t_{n-1}) - \hat{R}(t_n - t_{n-1}))^2}{t_n - t_{n-1}}.$$

Como, no caso da aplicação às TBMs da população portuguesa, todas as séries são anuais, acima fica $t_n - t_{n-1} = 1$, o que simplifica bastante a computação (ver código no apêndice B). Esta simplificação é válida para todos os modelos aplicados a este conjunto de dados e expostos nas secções e capítulos seguintes.

Para obter os intervalos de confiança, \mathcal{IC} , para os parâmetros, podemos considerar as propriedades assintóticas da estimação por MV. A matriz de informação de Fisher, \mathbf{F} , é dada por

$$\mathbf{F} = \begin{bmatrix} -E \left[\frac{\partial^2 \mathcal{L}}{\partial R^2} \right] & -E \left[\frac{\partial^2 \mathcal{L}}{\partial R \partial V} \right] \\ -E \left[\frac{\partial^2 \mathcal{L}}{\partial V \partial R} \right] & -E \left[\frac{\partial^2 \mathcal{L}}{\partial V^2} \right] \end{bmatrix} = \begin{bmatrix} \frac{t_N}{V} & 0 \\ 0 & \frac{N}{2V^2} \end{bmatrix}.$$

Por sua vez, a variância de cada um dos elementos de $\hat{\mathbf{p}}$ é dada pelos valores da diagonal da inversa da matriz \mathbf{F} . Para cada parâmetro p podemos assim obter uma aproximação dos limites de um intervalo de confiança com um nível de confiança $1 - \alpha$, $\mathcal{IC}_{(1-\alpha) \times 100\%}$, através de $\hat{p} \pm z_{1-\alpha/2} \sqrt{\widehat{Var}[\hat{p}]}$, onde $\widehat{Var}[\hat{p}]$ representa a variância de p com os parâmetros substituídos pelos seus estimadores de MV. Mais concretamente, os respetivos \mathcal{IC} assintóticos, para R e V , são dados por

$$\mathcal{IC}_{(1-\alpha) \times 100\%}(R) = \hat{R} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{V}}{t_N}}$$

e

$$\mathcal{IC}_{(1-\alpha)\times 100\%}(V) = \widehat{V} \pm z_{1-\alpha/2} \sqrt{\frac{2\widehat{V}^2}{N}},$$

onde z_q é o quantil de ordem q da distribuição normal padrão.

Neste caso, podemos também calcular os intervalos de confiança exatos, $\mathcal{IC}_{(1-\alpha)\times 100\%}^e$, usando as distribuições exatas, como em [10]. Com efeito,

$$(\widehat{R} - R) \sqrt{\frac{N-1}{N} \frac{t_N}{\widehat{V}}} \sim t_{N-1}$$

e

$$\frac{N\widehat{V}}{V} \sim \chi_{N-1}^2,$$

onde t_{N-1} representa a distribuição t de Student e χ_{N-1}^2 a distribuição qui-quadrado, em ambos os casos com $N-1$ graus de liberdade. Logo, os intervalos de confiança exatos são dados por

$$\mathcal{IC}_{(1-\alpha)\times 100\%}^e(R) = \widehat{R} \pm t_{1-\alpha/2; N-1} \sqrt{\frac{N}{(N-1)} \frac{\widehat{V}}{t_N}}$$

e

$$\mathcal{IC}_{(1-\alpha)\times 100\%}^e(V) = \left[\frac{N\widehat{V}}{\chi_{1-\alpha/2; N-1}^2}; \frac{N\widehat{V}}{\chi_{\alpha/2; N-1}^2} \right],$$

onde $t_{q; N-1}$ representa o quantil de ordem q da distribuição t de Student e $\chi_{q; N-1}^2$ os quantis de ordem q da distribuição qui-quadrado, em ambos os casos, com $N-1$ graus de liberdade.

Se tivermos observações até um certo instante t_N , com $Y(t_N) = y_{t_N}$, e quisermos obter previsões para um certo instante $t > t_N$, considerando que $Y(t)$ é um processo de Markov, temos

$$E[Y(t)|Y(t_1), \dots, Y(t_N)] = E[Y(t)|Y(t_N)].$$

De (3.7), vem

$$Y(t)|Y(t_N) \sim \mathcal{N}(Y(t_N) + R(t - t_N), V(t - t_N)).$$

Podemos usar para previsões a longo prazo (LP), em cada idade, para $t > t_N$,

$$\widehat{Y}(t) = \widehat{E}[Y(t)|Y(t_N) = y_{t_N}] = y_{t_N} + \widehat{R}(t - t_N), \quad (3.10)$$

onde \widehat{E} representa o valor aproximado da esperança matemática, pois, como não conhecemos o valor exato de R , substituímo-lo pelo do seu estimador de MV, \widehat{R} .

As previsões passo-a-passo (PP) são estimadas da mesma forma que em (3.10), mas atualizando t e a última observação, bem como as estimativas dos parâmetros, cada vez que se progride um passo no tempo (no nosso caso, um ano). No apêndice B apresentamos o código R correspondente a todo o processo de modelação.

Finalmente, podemos recorrer à técnica de simulação para obter a distribuição aproximada dos erros de previsão $\widehat{Y}(t) - Y(t)$ e intervalos de confiança de previsão. De (3.7) conhecemos a média e a variância de $Y(t_n)|Y(t_{n-1}) = y_{t_{n-1}}$. Usamos, para cada idade e por sexo, as estimativas de MV de p e simulamos um número suficientemente elevado de réplicas (trajetórias) $Y^*(t)$, digamos r (neste

caso, usamos $r = 1000$). Deste modo, obtemos até um certo ano t_N as estimativas de MV, para cada uma das r réplicas simuladas, um novo vetor de parâmetros, p^\bullet , as previsões $\hat{Y}^\bullet(t)$ (para $t > t_N$), os erros de previsão $\hat{Y}^\bullet(t) - Y^\bullet(t)$, bem como a média e a variância empíricas destes no conjunto das r réplicas, para estimar a média e variância do erro de previsão.

Designemos por M_t e V_t as respetivas médias e variâncias empíricas. Seja $\hat{Y}(t)$ a previsão inicial no instante t . Podemos obter uma aproximação dos limites do $\mathcal{IC}_{(1-\alpha)\times 100\%}$, para uma certa idade e sexo considerados, através de

$$\hat{Y}(t) - M_t \pm z_{1-\alpha/2} \sqrt{V_t}. \quad (3.11)$$

3.2.3 Resultados

Ajustamos o MBG aos dados da mortalidade da população portuguesa, para cada uma das idades selecionadas do arco da vida (0 a 99 anos) e por sexo. Para o efeito, usamos a variável $Y(t) = \ln(X(t)/x_0)$, com $X(t)$ cada uma das séries temporais das TBMs.

Nas figuras 3.1 e 3.2, representam-se as estimativas dos parâmetros do modelo, respetivamente \hat{R} e \hat{V} , estimados para todas as idades e por sexo, bem como os intervalos de confiança, \mathcal{IC} , que lhes estão associados. Se considerarmos o “comportamento” dos parâmetros estimados com a idade, constatamos que, relativamente a R , existe uma ligeira tendência crescente, mais notória até às primeiras idades adultas, crescendo muito lentamente depois da idade 20.

Se considerarmos a evolução do “comportamento” dos parâmetros estimados por sexo, fixando cada idade, constatamos que, se os valores de R são bastante semelhantes, já os valores de V apresentam um padrão diferente nas primeiras idades, com mais oscilações (mais notório entre as idades 18 a 30), e também nas últimas (depois da idade 95, sobretudo no caso do sexo masculino).

Quanto aos \mathcal{IC} , estimamos os intervalos assintóticos, mas como, neste caso, é também possível estimar os intervalos exatos, apresentamos os resultados obtidos pelas duas vias para um grau de confiança de 95%, respetivamente, $\mathcal{IC}_{95\%}$ e $\mathcal{IC}_{95\%}^e$.

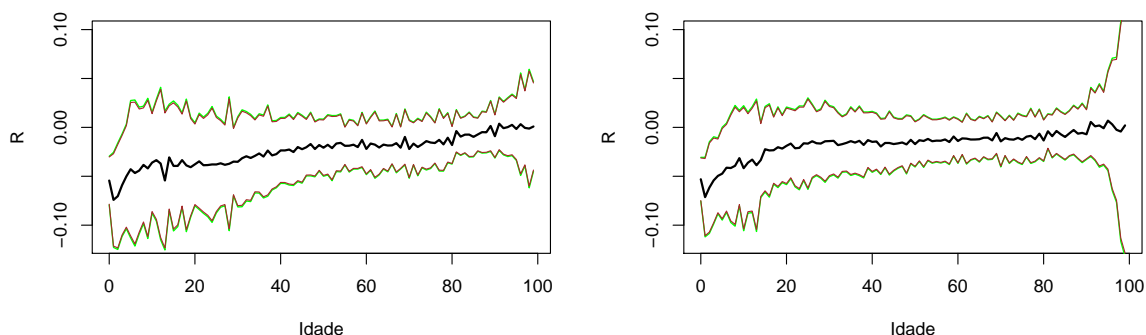


Figura 3.1: Estimativas de R do MBG com $\mathcal{IC}_{95\%}^e$ (a verde) e $\mathcal{IC}_{95\%}$ (a castanho), por idade e por sexo (feminino, à esquerda e masculino, à direita)

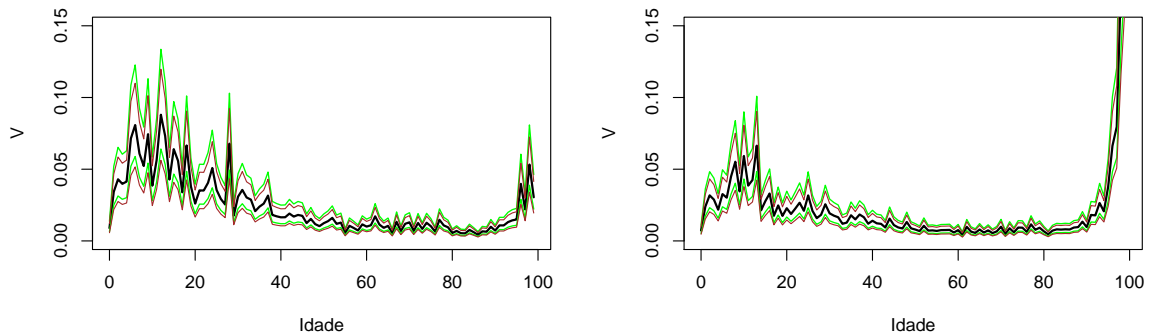


Figura 3.2: Estimativas de V do MBG com $IC_{95\%}^e$ (a verde) e $IC_{95\%}$ (a castanho), por idade e por sexo (feminino, à esquerda e masculino, à direita)

Para ambos os parâmetros, os IC assintóticos e exatos estimados pouco diferem e não há vantagens significativas na utilização dos intervalos exatos (nas figuras anteriores, as duas representações quase que se sobrepõem). As amplitudes dos IC de R e V são aproximadamente proporcionais a \sqrt{V} e a V , o que determina a maior amplitude dos IC de R comparativamente aos de V . No caso do parâmetro R , destaca-se ainda a enorme amplitude dos limites dos IC depois da idade 95 do sexo masculino. Nas caixas 3.1 e 3.2, respetivamente para R e V , mostramos as estimativas dos parâmetros bem como os IC associados para as primeiras dez idades do sexo masculino.

Caixa 3.1 MBG: estimativas do parâmetro R e respetivos $IC_{95\%}$ (IC , assintóticos; ICe , exatos) para as idades 0 a 9 do sexo masculino

```
> IC.R.M <- cbind(lim.inf.assin.R,lim.inf.exa.R,R,lim.sup.exa.R,lim.sup.assin.R)
> IC.R.M[1:10]
      lim inf IC lim inf ICe      R      lim sup ICe      lim sup IC
[1,] -0.07477691 -0.07543025 -0.05303565 -0.0306410552 -0.031294396
[2,] -0.11079548 -0.11198154 -0.07132691 -0.0306722732 -0.031858333
[3,] -0.10707953 -0.10844656 -0.06158875 -0.0147309436 -0.016097975
[4,] -0.09752380 -0.09882547 -0.05420811 -0.0095907419 -0.010892410
[5,] -0.08725503 -0.08838646 -0.04960430 -0.0108221420 -0.011953574
[6,] -0.09341776 -0.09480346 -0.04730588  0.0001917078 -0.001193988
[7,] -0.08554126 -0.08687276 -0.04123296  0.0044068362  0.003075339
[8,] -0.09564886 -0.09727390 -0.04157253  0.0141288373  0.012503804
[9,] -0.09922982 -0.10103133 -0.03928069  0.0224699437  0.020668429
[10,] -0.07888325 -0.08030825 -0.03146344  0.0173813742  0.015956374
```

Os resultados dos ajustamentos e previsões foram revertidos para a escala original, $X(t)$, das TBMs, em vez de $Y(t)$. Na figura 3.3, ilustramos uma concretização do ajustamento (fazendo $\sigma = 0$ em (3.4) e substituindo os parâmetros pelos seus estimadores de MV) e previsões, neste caso, para a idade 8 do sexo masculino.

Caixa 3.2 MBG: estimativas do parâmetro V e respetivos $IC_{95\%}$ (IC , assintóticos; ICe , exatos) para as idades 0 a 9 do sexo masculino

```
> IC.V.M <- cbind(lim.inf.assin.V,lim.inf.exa.V,V,lim.sup.exa.V,lim.sup.assin.V)
> IC.V.M[1:10]
```

	lim inf IC	lim inf ICe	V	lim sup ICe	lim sup IC
[1,]	0.004639824	0.005292021	0.00725954	0.01102663	0.009879255
[2,]	0.015290974	0.017440350	0.02392449	0.03633928	0.032558010
[3,]	0.020313230	0.023168559	0.03178239	0.04827476	0.043251550
[4,]	0.018417171	0.021005980	0.02881579	0.04376874	0.039214403
[5,]	0.013914867	0.015870811	0.02177141	0.03306893	0.029627961
[6,]	0.020871713	0.023805546	0.03265620	0.04960200	0.044440690
[7,]	0.019270920	0.021979737	0.03015158	0.04579769	0.041032233
[8,]	0.028704288	0.032739107	0.04491117	0.06821626	0.061118049
[9,]	0.035277513	0.040236297	0.05519574	0.08383765	0.075113959
[10,]	0.022072521	0.025175145	0.03453500	0.05245574	0.046997487

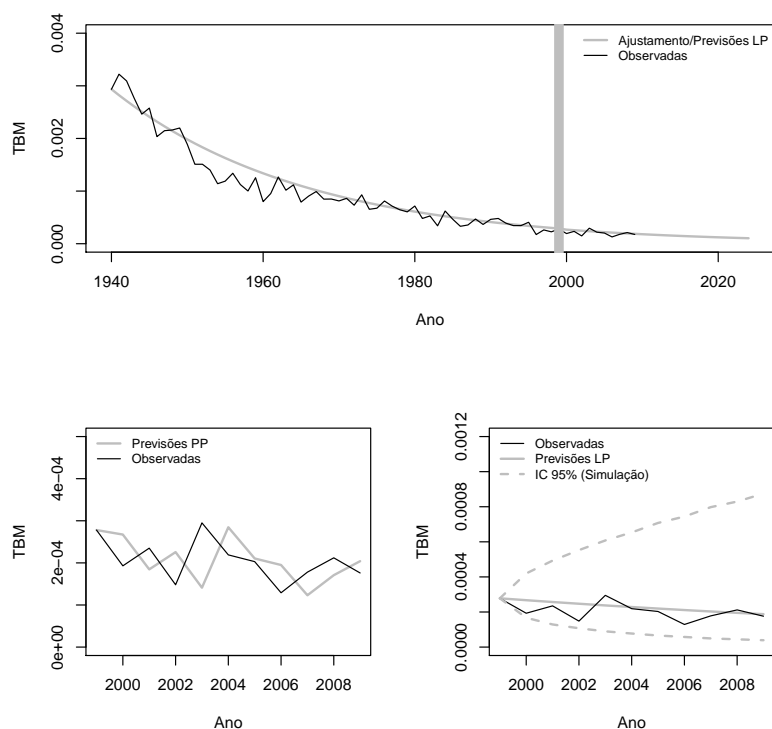


Figura 3.3: Ajustamento do MBG com previsões a LP (25 anos: de 2000 a 2024) para a idade 8 do sexo masculino (em cima); previsões, no período de 2000 a 2009, PP e a LP com $IC_{95\%}$ assintóticos (respetivamente, à esquerda e à direita, em baixo)

No caso das previsões a LP representamos também, na figura anterior, os IC associados, obtidos por simulação (usando a expressão (3.11)). Recordemos que usámos para o ajustamento os dados de 1940 a 1999, reservando os de 2000 a 2009 para a previsão. Note-se que optámos

por representar também esses valores na figura conjunta do ajustamento e previsões (precisamente à direita da barra vertical cinzenta que marca o início do período de previsões), pois traduzem uma informação adicional à estimativa do erro, que resulta da comparação entre a sua tendência e a das previsões.

Na figura 3.4, mostramos as 1000 réplicas das simulações realizadas para obter os IC das previsões para a idade 8 do sexo masculino (ilustrada no exemplo da figura 3.3).

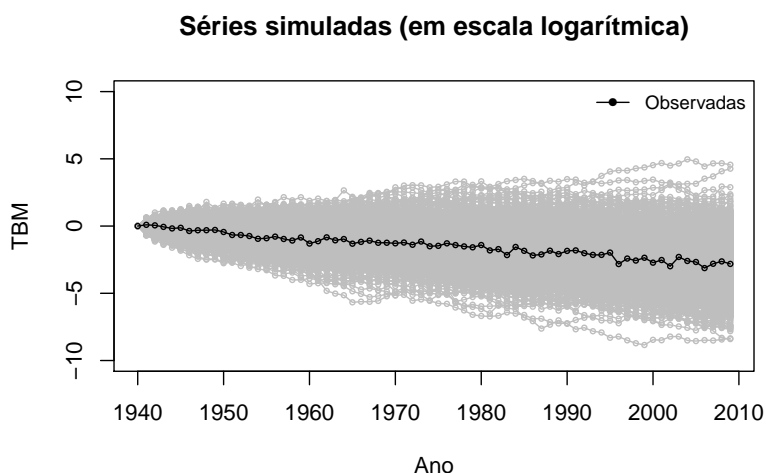


Figura 3.4: 1000 réplicas simuladas (de 1940 a 2009) da TBM pelo MBG para a idade 8 do sexo masculino

De um modo geral, os resultados da aplicação do modelo são bastante satisfatórios. O MBG ajusta-se bem aos dados e fornece “boas” previsões. Na caixa 3.3, mostramos as previsões (obtidas usando (3.10)) a LP, para o período de 2000 a 2009, para a idade 8 de ambos os sexos, antecedidas pelos valores das TBMs observadas, de modo a melhor enquadrar a ordem de grandeza. Para medir a “bondade” do ajustamento, usamos como critério quantitativo o erro quadrático médio (EQM).

Caixa 3.3 MBG: previsões a LP, para o período 2000 a 2009, das TBMs da idade 8 de ambos os sexos

```
> tx.prev.MBG <- cbind(Taxas.F[60:70,9],Prev.F[,9],Taxas.M[60:70,9],Prev.M[,9])
      Observ.F prevLP.F Observ.M prevLP.M
[1,] 0.000165 0.000227 0.000193 0.000267
[2,] 0.000207 0.000158 0.000235 0.000184
[3,] 0.000156 0.000199 0.000148 0.000226
[4,] 0.000156 0.000149 0.000295 0.000141
[5,] 0.000077 0.000150 0.000219 0.000284
[6,] 0.000134 0.000073 0.000203 0.000210
[7,] 0.000154 0.000128 0.000129 0.000195
[8,] 0.000037 0.000148 0.000178 0.000123
[9,] 0.000092 0.000035 0.000212 0.000171
[10,] 0.000149 0.000088 0.000176 0.000204
```

Analisando globalmente os resultados, quer o ajustamento, quer as previsões são melhores (segundo os critérios anteriores) para as séries de dados relativos ao sexo feminino. Nas figuras 3.5 e 3.6 ilustramos os respetivos EQMs.

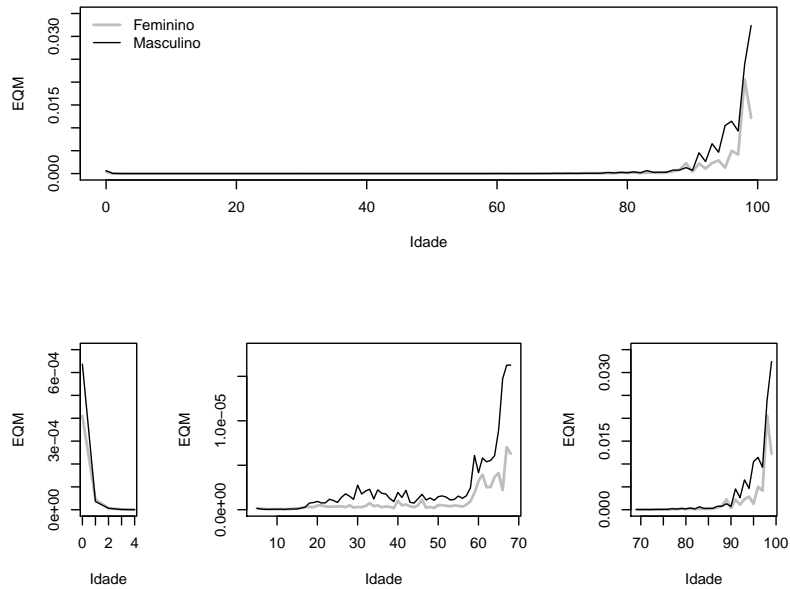


Figura 3.5: EQM do ajustamento do MBG às TBMs, por idade e por sexo (em cima, representação para todas as idades; em baixo, ampliação por grupos de idades)

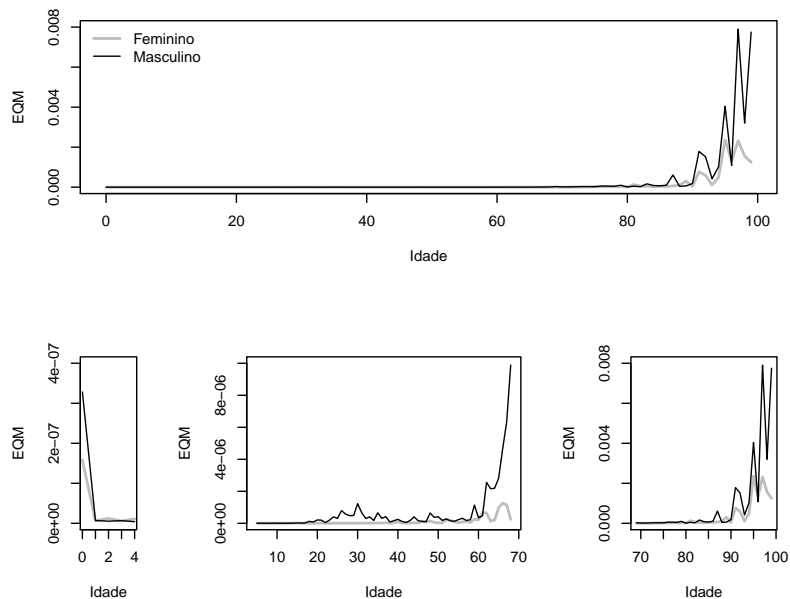


Figura 3.6: EQM das previsões a LP das TBMs, para o período 2000 a 2009, do MBG, por idade e por sexo (em cima, representação para todas as idades; em baixo, ampliação por grupos de idades)

A diferença no desempenho dos modelos entre sexos é mais notória entre as idades 18 e 45 (corresponde a um conjunto de idades em que, ao longo do tempo, o padrão de mortalidade do sexo masculino sofre uma inflexão relativamente à proeminente tendência decrescente geral). Também a partir da idade 91, em ambos os sexos, contudo mais expressivo no masculino, o modelo não consegue captar a variabilidade das séries e obter um bom ajustamento, daí o expressivo aumento do EQM ilustrado na figura 3.5. No entanto, e apesar do EQM das previsões ser também elevado nas idades mais avançadas relativamente às restantes (ver figura 3.6), o modelo consegue ainda assim, nessas idades, fornecer algumas previsões a considerar, já que tendem fortemente para as médias das séries (veja-se a ilustração de dois casos particulares que correspondem a esta descrição na figura 3.7).

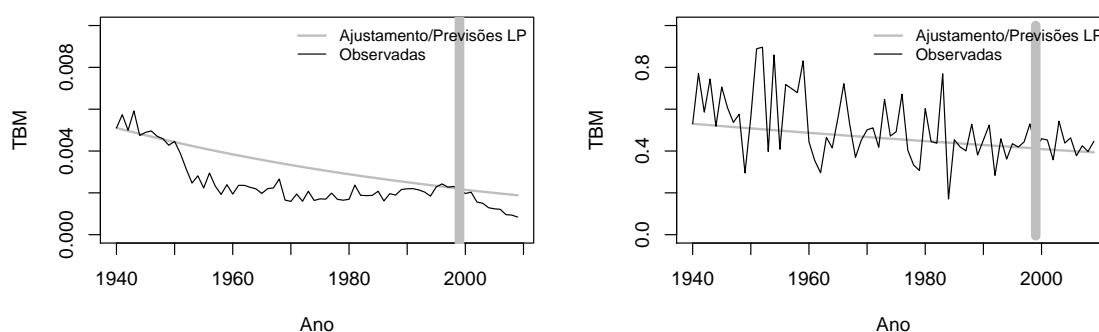


Figura 3.7: Ajustamento do MBG com previsões a LP (de 2000 a 2009) para as idades 29 (à esquerda) e 98 (à direita) do sexo masculino

Na caixa 3.4, mostramos os valores dos EQMs para o ajustamento, previsões a LP e PP das TBMs, no período de 2000 a 2009, das idades 0 a 9 de ambos os sexos.

Caixa 3.4 MBG: EQM do ajustamento às TBMs e das previsões PP e a LP, para o período 2000 a 2009, das TBMs das idades 0 a 9 de ambos os sexos

```
> EQM.MBG
      EQM.ajust.F  EQM.ajust.M  EQM.prevLP.F  EQM.prevLP.M  EQM.prevPP.F  EQM.prevPP.M
[1,] 4.093210e-04 6.366660e-04 1.584287e-07 3.282315e-07 2.618232e-07 2.738289e-07
[2,] 4.446729e-05 3.563518e-05 6.340558e-09 7.015487e-09 1.795989e-08 1.443912e-08
[3,] 6.792582e-06 6.213509e-06 1.239100e-08 5.307098e-09 1.215987e-08 1.140113e-08
[4,] 9.337624e-07 1.125694e-06 5.920429e-09 6.618440e-09 1.314745e-08 2.355860e-09
[5,] 3.138975e-07 2.521653e-07 1.105865e-08 4.102888e-09 4.835788e-09 3.271105e-09
[6,] 1.623475e-07 1.534057e-07 8.391448e-09 4.006402e-09 5.014393e-09 5.891796e-09
[7,] 1.205340e-07 7.478043e-08 1.782075e-09 1.023516e-08 3.510684e-09 7.638037e-09
[8,] 8.450619e-08 4.341852e-08 4.104093e-09 2.394581e-09 5.108838e-09 3.254216e-09
[9,] 5.871227e-08 4.294925e-08 4.633425e-09 2.480245e-09 3.372764e-09 4.723781e-09
[10,] 5.355684e-08 6.032932e-08 1.897361e-09 1.043273e-08 4.630891e-09 4.180100e-09
```

Parecendo-nos adequado o MBG para modelar, como ponto de partida, este tipo de dados e tendo obtido resultados que consideramos satisfatórios, mostramos, na secção seguinte, a aplicação de um outro modelo de EDEs, o modelo de Gompertz estocástico (MGE). No final do capítulo, comparamos o desempenho dos dois modelos entre si e também com modelos alternativos.

3.3 Modelo de Gompertz estocástico

O *modelo determinístico monomolecular* (do início do século XIX), inicialmente proposto para descrever uma reação química irreversível de primeira ordem, também conhecido por *modelo de Mitscherlich* na área da nutrição das plantas e fertilização dos solos, pode ser enunciado por

$$\frac{dY(t)}{dt} = b(A - Y(t)) \quad (3.12)$$

e tem como solução, admitindo $Y(t_0) = y_0$ conhecido,

$$Y(t) = A - (A - y_0) \exp\{-b(t - t_0)\}.$$

No caso das plantas, considera-se que a evolução da taxa de crescimento ao longo do tempo, $dY(t)/dt$, onde vamos supor que $Y(t)$ é uma medida de biomassa, proporcional à diferença entre uma biomassa máxima ou assintótica e a biomassa entretanto formada, com b a constante de proporcionalidade. Se quisermos, o modelo em (3.12) pode ainda ser escrito de uma forma mais geral pondo $\frac{dY^c(t)}{dt} = b(a^c - Y^c(t))$ (nos exemplos que se seguem, usamos sempre $c = 1$).

Este modelo está na origem de outros, que podem obter-se através de transformações, usando uma variável $Y(t) = h(X(t))$, em que h é uma função monótona de classe C^1 . Um exemplo, é o *modelo de Gompertz determinístico*, que se pode escrever na forma

$$\frac{d \ln(X(t))}{dt} = b(\ln(a) - \ln(X(t))), \quad (3.13)$$

e que se reduz a (3.12), pondo $Y(t) = h(X(t)) = \ln(X(t))$ e $A = h(a) = \ln(a)$.

O modelo de Gompertz estocástico (MGE), que abordamos com mais pormenor de seguida, obtém-se adicionando ao modelo de Gompertz determinístico, em (3.12), um ruído branco que descreve as flutuações aleatórias do ambiente sobre a mortalidade.

3.3.1 O MGE

Consideremos, como na subsecção 3.2, a simplificação de notação $X(t) = X_k(t)$, para a TBM dos indivíduos de uma certa idade i e sexo j , com $k = i + 100(j - 1)$, no instante t . Sejam $Y(t)$ e A definidos como em (3.12), isto é, $Y(t) = \ln(X(t))$ e $A = \ln(a)$, o MGE pode escrever-se na forma

$$dY(t) = b(A - Y(t))dt + \sigma dW(t), \quad (3.14)$$

com $Y(t_0) = y_{t_0}$, a condição inicial suposta conhecida, $W(t)$ um processo de Wiener padrão e parâmetros $A = \ln(a)$, onde a representa a taxa média de mortalidade assintótica, b , que regula

a velocidade de aproximação ao regime assintótico e σ , que mede a intensidade das flutuações aleatórias do ambiente. Note-se que foram experimentadas exaustivamente várias transformações h , de acordo com as recomendações dadas na literatura (ver, por exemplo, [82]), no sentido de diminuir a variância das séries e tentar obter séries com padrão mais linear ou curvilíneo suavizado, de modo a facilitar a modelação, mas de facto a transformação logarítmica, utilizada com frequência em taxas de crescimento de organismos na biologia, revelou-se a mais vantajosa para o conjunto dos dados.

A solução da equação (3.14), para cada idade e sexo, no instante t , é dada por

$$Y(t) = A + (y_{t_0} - A) \exp \{-b(t - t_0)\} + \sigma \exp \{-bt\} \int_{t_0}^t \exp \{bs\} dW(s) \quad (3.15)$$

com distribuição normal, com média

$$A + (y_{t_0} - A) \exp \{-b(t - t_0)\}$$

e variância

$$\sigma^2(1 - \exp \{-2b(t - t_0)\})/(2b),$$

isto é,

$$Y(t) \sim \mathcal{N}(A + (y_{t_0} - A) \exp \{-b(t - t_0)\}, \sigma^2(1 - \exp \{-2b(t - t_0)\})/(2b)). \quad (3.16)$$

A equação (3.14) é uma EDE autónoma. Por sua vez, a solução, (3.15), é uma difusão de Itô e um processo de difusão homogêneo com coeficiente de tendência $a(y) = b(A - y)$ e coeficiente de difusão $b(y) = \sigma^2$.

3.3.2 Estimação

Podemos supor $t_0 = 0$. Sejam $t_n = n$ ($n = 0, 1, 2, \dots, N$) os anos em que as TBMs foram observadas, para cada idade e por sexo. A f.d.p. de transição de $Y(t)$ entre t_{n-1} e t_n é dada por

$$f(y_n | Y(t_{n-1})) = \frac{1}{\sqrt{2\pi s^2_{Y,n|n-1}}} \exp \left\{ -\frac{1}{2} \frac{(y_n - \mu_{Y,n|n-1})^2}{s^2_{Y,n|n-1}} \right\}$$

em que

$$\mu_{Y,n|n-1} = E[Y(t_n) | Y(t_{n-1})] = A + (Y(t_{n-1}) - A) \exp \{-b(t_n - t_{n-1})\}$$

e

$$s^2_{Y,n|n-1} = \text{Var}[Y(t_n) | Y(t_{n-1})] = \sigma^2(1 - \exp \{-2b(t_n - t_{n-1})\})/(2b).$$

O vetor de parâmetros, $\mathbf{p} = (A, b, \sigma)$, pode também ser estimado por MV. Por $Y(t)$ ser um processo de Markov, a função de log-verosimilhança, \mathcal{L} , é dada por

$$\begin{aligned} \mathcal{L}(\mathbf{p} | Y(t_1), \dots, Y(t_N)) &= \sum_{n=1}^N \ln(f(y_n | Y(t_{n-1}))) \\ &= -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(s^2_{Y,n|n-1}) - \frac{1}{2} \sum_{n=1}^N \frac{(Y(t_n) - \mu_{Y,n|n-1})^2}{s^2_{Y,n|n-1}}, \end{aligned} \quad (3.17)$$

a partir da qual é possível definir as expressões dos estimadores de MV dos parâmetros, \hat{p} . Resolvemos, então, o sistema de equações

$$\begin{cases} \frac{\partial \mathcal{L}(y; p)}{\partial A} \Big|_{\hat{A}, \hat{b}, \hat{\sigma}} = 0 \\ \frac{\partial \mathcal{L}(y; p)}{\partial b} \Big|_{\hat{A}, \hat{b}, \hat{\sigma}} = 0 \\ \frac{\partial \mathcal{L}(y; p)}{\partial \sigma} \Big|_{\hat{A}, \hat{b}, \hat{\sigma}} = 0. \end{cases}$$

Fixando \hat{b} (pois não é possível obter explicitamente as expressões para os três parâmetros), vem

$$\hat{A} = \frac{1}{N} \sum_{n=1}^N \frac{Y(t_n) - Y(t_{n-1}) \exp\{-\hat{b}(t_n - t_{n-1})\}}{1 - \exp\{-\hat{b}(t_n - t_{n-1})\}}$$

e

$$\hat{\sigma} = \left(\frac{2\hat{b}}{N} \sum_{n=1}^N \frac{(Y(t_n) - \hat{A} - (Y(t_{n-1}) - \hat{A}) \exp\{-\hat{b}(t_n - t_{n-1})\})^2}{1 - \exp\{-2\hat{b}(t_n - t_{n-1})\}} \right)^{\frac{1}{2}}.$$

Para efeitos de cálculo, atenderemos a que $t_n - t_{n-1} = 1$, que recordamos se deve ao facto de os dados serem séries temporais anuais. Resulta do exposto que definimos \hat{A} como função de \hat{b} , tal que $\hat{A} := \zeta_1(\hat{b})$, e $\hat{\sigma}$ como função de \hat{A} e de \hat{b} , tal que $\hat{\sigma} := \zeta_2(\hat{A}, \hat{b})$. Obtemos assim uma nova função, \mathcal{L}^* , como os mesmos valores ótimos que a função de log-verosimilhança definida em (3.17), mas que depende apenas do parâmetro b , e que podemos escrever na forma

$$\mathcal{L}^*(b|Y(t_1), \dots, Y(t_N)) := -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(s^{*2}_{Y,n|n-1}) - \frac{1}{2} \sum_{n=1}^N \frac{(Y(t_n) - \mu^*_{Y,n|n-1})^2}{s^{*2}_{Y,n|n-1}}, \quad (3.18)$$

com

$$\mu^*_{Y,n|n-1} := \zeta_1(b) + (Y(t_{n-1}) - \zeta_1(b)) \exp\{-b\}$$

e

$$s^{*2}_{Y,n|n-1} := (\zeta_2(\zeta_1(b), b))^2 (1 - \exp\{-2b\}) / (2b).$$

O estimador de MV de b , para cada idade e sexo, é obtido pela minimização do simétrico de (3.18), utilizando a função `optimize` do R (conforme código, no apêndice C). Este método (descrito em [30] e aplicado também em modelo análogo em [14]), utilizando \mathcal{L}^* em vez de \mathcal{L} para encontrar os estimadores de MV do vetor de parâmetros p , é especialmente útil quando é difícil encontrar uma expressão explícita para os estimadores, com a vantagem de ser computacionalmente eficiente (sem recurso a métodos numéricos mais complicados de implementar). Uma vez obtido \hat{b} , os estimadores de MV \hat{A} e $\hat{\sigma}$ obtém-se, respetivamente, a partir de $\hat{A} = \zeta_1(\hat{b})$ e $\hat{\sigma} = \zeta_2(\hat{A}, \hat{b})$.

Para obter uma aproximação dos intervalos de confiança para os parâmetros, assumimos que estamos em regime assintótico, considerando as propriedades de estimação por MV, e fazemos, também, uma aproximação da matriz de informação de Fisher pela simétrica da inversa da matriz hessiana, de cuja diagonal obtemos uma aproximação às variâncias dos parâmetros estimados.

Considerando um parâmetro p e o seu estimador de MV, \hat{p} , uma aproximação dos limites do $\mathcal{IC}_{(1-\alpha) \times 100\%}$ pode ser obtida analogamente ao utilizado no caso do MBG através de

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\widehat{Var}[\hat{p}]},$$

em que $\widehat{Var}[\hat{\rho}]$ representa uma estimativa da variância do parâmetro obtida a partir da inversa da matriz hessiana pelo método acima descrito.

Se tivermos observações até um certo tempo t_N , e quisermos obter previsões para um certo tempo $t > t_N$, considerando que $Y(t)$ é um processo de Markov, temos que

$$E[Y(t)|Y(t_1), \dots, Y(t_N)] = E[Y(t)|Y(t_N)].$$

Como

$$Y(t)|Y(t_N) \sim \mathcal{N}(A + (Y(t_N) - A) \exp\{-b(t - t_N)\}, \sigma^2(1 - \exp\{-2b(t - t_N)\})/(2b)),$$

podemos usar para previsões a longo prazo, em cada idade,

$$\widehat{Y}(t) = \widehat{E}[Y(t)|Y(t_N) = y_{t_N}] = \widehat{A} + (y_{t_N} - \widehat{A}) \exp\{-\widehat{b}(t - t_N)\}, \quad (3.19)$$

onde \widehat{E} representa o valor aproximado da esperança matemática, substituindo os valores exatos de A e de b pelos dos seus estimadores de MV, respetivamente, \widehat{A} e \widehat{b} .

As previsões PP são estimadas da mesma forma que em (3.19), mas atualizando t , e a última observação, bem como as estimativas dos parâmetros de cada vez (no nosso caso anualmente) que se avança um passo no tempo (ver no apêndice C o código R correspondente).

Obtidas as previsões, se quisermos encontrar um intervalo de confiança para os seus erros, dados por $\widehat{Y}(t) - Y(t)$, podemos usar, em alternativa à técnica de simulação apresentada na secção anterior, o *método Delta* para estimar a sua variância (ver [20, 70]). Este método utiliza-se para estimar a média e a variância de funções dos parâmetros, a partir de estimativas da média e da variância dos parâmetros. Em (3.14), fizemos $A = h(a) = \ln(a)$, donde, para reverter ao parâmetro inicial a , também aqui podemos usar o método de Delta (usando uma função g , tal que $g(A) = h^{-1}(A) = \exp\{A\}$, de modo a obter, designadamente, os limites do $\mathcal{IC}_{(1-\alpha) \times 100\%}$ a partir da expressão $g(\widehat{A}) \pm z_{1-\alpha/2} g'(\widehat{A}) \sqrt{\widehat{Var}[g(\widehat{A})]}$, onde g' representa a derivada de g .

Note-se que, para a aplicação deste método, a função g tem que ser diferenciável, pois este baseia-se na expansão em série de Taylor dessa mesma função (neste caso, usamos apenas o termo linear), através de uma generalização do *teorema do limite central*. Relativamente aos erros de previsão do MGE, aplicado a cada idade e por sexo, fazemos (como em [25])

$$\begin{aligned} \widehat{Y}(t) - Y(t) = g_t(\widehat{A}, \widehat{b}, S) &= \widehat{A} + (Y(t_N) - \widehat{A}) \exp\{-\widehat{b}(t - t_N)\} \\ &- A - (Y(t_N) - A) \exp\{-b(t - t_N)\} - \sigma \exp\{-bt\} S, \end{aligned} \quad (3.20)$$

em que $S = \int_{t_N}^t \exp\{bs\} dW(s)$ tem distribuição normal com média nula e variância $\frac{\exp\{2bt\} - \exp\{2bt_N\}}{2b}$. A partir do método Delta, vem que

$$E[\widehat{Y}(t) - Y(t)] \approx g_t(A, b, 0) = 0$$

e

$$Var[\widehat{Y}(t) - Y(t)] \approx E \left[\left((\widehat{A} - A) \frac{\partial g_t(A, b, 0)}{\partial A} + (\widehat{b} - b) \frac{\partial g_t(A, b, 0)}{\partial b} + S \frac{\partial g_t(A, b, 0)}{\partial S} \right)^2 \right],$$

sendo que a variância pode ser decomposta em dois termos: um termo V_P , que corresponde à variabilidade dos erros de estimação dos parâmetros, e um termo V_E , associado à variabilidade devida às flutuações aleatórias do ambiente (através do integral estocástico). Podemos assim obter uma aproximação da variância na forma

$$Var[\hat{Y}(t) - Y(t)] \approx V_P + V_E,$$

com

$$V_P = (1 - E_N^b)^2 Var[\hat{A}] + (Y(t_N) - A)(t - t_N)^2 E_N^{2b} Var[\hat{b}] - 2(Y(t_N) - A)(t - t_N) E_N^b Cov[\hat{A}, \hat{b}],$$

onde

$$E_N = \exp\{-(t - t_N)\},$$

e

$$V_E = \frac{\sigma^2}{2b} (1 - E_N^{2b}).$$

As variâncias e a covariância V_p podem obter-se aproximadamente usando a matriz inversa da simétrica da matriz de informação de Fisher empírica.

3.3.3 Resultados

Tal como na secção anterior, relativamente ao MBG, foi possível ajustar o MGE aos dados da mortalidade da população portuguesa, para cada uma das idades seleccionadas do arco da vida (0 a 99 anos) e por sexo. Para o efeito, usámos, neste caso, a variável $Y(t) = \ln(X(t))$, com $X(t)$ a representar cada uma das séries temporais das TBMs.

Na figura 3.8, podemos observar as estimativas dos parâmetros do modelo, a , b e σ , por idade e por sexo.

Recordamos que estimámos o valor de $A = \ln(a)$, mas optamos por mostrar o parâmetro na escala original, a , que representa a TBM assintótica média (média geométrica). Na figura referida, e à esquerda, apresentamos os valores estimados para todas as idades, e à direita, excluimos as últimas 10 idades, de modo a mostrar com mais detalhe o comportamento dos parâmetros estimados nas idades adultas e tornar possível uma melhor compreensão da forma descrita em cada gráfico (principalmente no relativo ao parâmetro b).

Na realidade, os resultados das estimativas dos parâmetros do modelo vêm de encontro ao expectável, considerando o conhecimento do fenómeno em estudo. Assim, a , que representa a taxa de mortalidade assintótica, aumenta com a idade, tendo valores muitíssimo mais elevados nas últimas idades do arco de vida, em que a probabilidade de morte é maior. b pode informalmente traduzir-se como uma medida da velocidade de aproximação ao regime assintótico. Neste caso, apresenta uma ligeira tendência crescente, logo nas primeiras idades, mantendo-se depois, num patamar que oscila, em termos médios, em torno do valor 0.06 para o sexo feminino e 0.09 para o masculino. A partir da idade 80, os valores de b estimados aumentam até cerca de dez vezes mais, comparativamente aos níveis médios.

Quanto a σ , está associado ao termo do integral estocástico do modelo e mede a intensidade das flutuações aleatórias do ambiente sobre a mortalidade. Os valores estimados apresentam uma tendência crescente nas primeiras idades, que dizem respeito à infância e juventude. Esse padrão traduz a vulnerabilidade dos indivíduos nos primeiros anos de vida. A partir da idade 18, verifica-se um decréscimo lento nesses valores, estabilizando apenas entre as idades 60 e 80, a partir das quais o padrão descrito pelo parâmetro apresenta nova tendência crescente, que traduz a suscetibilidade das últimas idades do arco de vida, em que qualquer acontecimento aleatório pode provocar o óbito.

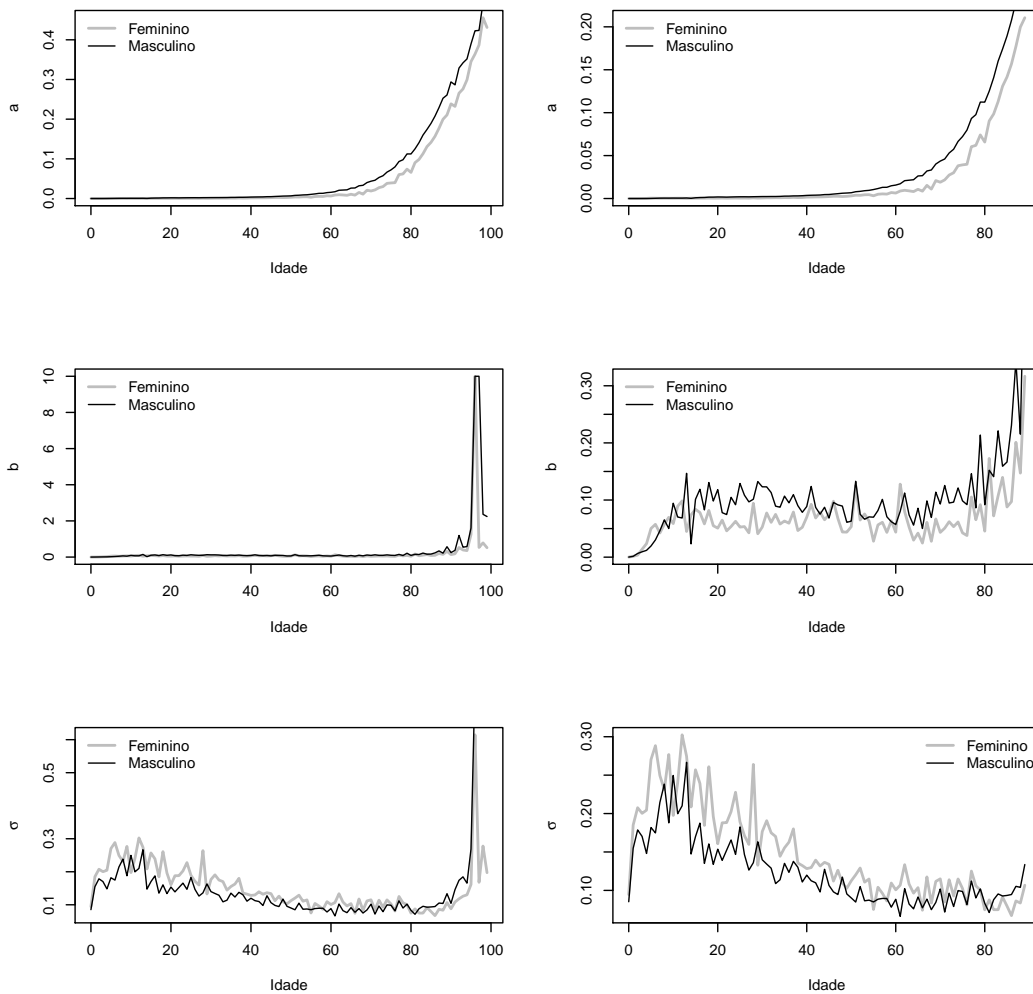


Figura 3.8: Estimativas dos parâmetros do MGE a , b e σ (respectivamente, de cima para baixo), por idade e por sexo; à esquerda para todas as idades e à direita excluindo as últimas 10 idades, o que permite ampliar a escala dos gráficos

Os gráficos anteriores sugerem ainda uma maior variabilidade das estimativas dos parâmetros entre idades consecutivas para b e σ comparativamente a a . Quando observamos o padrão dessas estimativas em função da idade, embora este seja semelhante em ambos os sexos, no caso de a e b , os valores estimados são superiores para o sexo masculino, verificando-se o contrário no caso de

σ . Na caixa 3.5, apresentamos os valores estimados, por MV, de $A = \ln(a)$, b e σ , a título ilustrativo, para as idades 30 a 39 de ambos os sexos.

À semelhança do que mostrámos antes para o MBG, podemos também, no caso do MGE, obter intervalos de confiança, \mathcal{IC} , para os parâmetros, considerando as propriedades assintóticas da estimação por MV e a aproximação à matriz de informação de Fisher empírica pela simétrica da matriz hessiana. Assim, na caixa 3.5 mostramos também as margens de erro dos $\mathcal{IC}_{95\%}$ assintóticos associados aos parâmetros estimados para as idades referidas.

Caixa 3.5 Estimativas dos parâmetros do MGE, e margens de erro dos $\mathcal{IC}_{95\%}$ assintóticos, para as idades 30 a 39 de ambos os sexos

```

> par.MGE.F <- cbind(A,A.tol,b,b.tol,sigma,sigma.tol)[31:40]
      A.F      +/-      b.F      +/-      sigma.F      +/-
[1,] -7.330300 1.2151180 0.05261091 0.06665715 0.1765242 0.03237092
[2,] -7.061650 0.7836088 0.07735702 0.08702334 0.1907148 0.03534781
[3,] -7.081565 0.9302608 0.06094323 0.07094953 0.1749462 0.03214633
[4,] -6.930213 0.7179070 0.07487668 0.07455329 0.1701812 0.03132316
[5,] -6.981973 0.8667725 0.05790942 0.06665168 0.1443982 0.02647807
[6,] -6.828785 0.7952847 0.06331128 0.07255061 0.1553509 0.02856811
[7,] -6.866995 0.9630023 0.05940186 0.07703418 0.1627824 0.03000745
[8,] -6.614026 0.6783814 0.07902816 0.09058847 0.1800799 0.03344864
[9,] -6.835085 1.1670100 0.04657088 0.06698154 0.1356130 0.02487415
[10,] -6.707579 1.0076630 0.05223754 0.07583681 0.1323194 0.02437844

> par.MGE.M <- cbind(A,A.tol,b,b.tol,sigma,sigma.tol)[131:140]
      A.M      +/-      b.M      +/-      sigma.M      +/-
[1,] -6.070924 0.3051248 0.12316052 0.10134970 0.1398543 0.02613352
[2,] -6.076342 0.2947742 0.12303593 0.10439780 0.1342401 0.02513628
[3,] -6.050484 0.3203390 0.11311080 0.09281066 0.1289315 0.02396517
[4,] -6.043381 0.3523670 0.08924373 0.07537648 0.1088865 0.02004617
[5,] -5.997587 0.3779968 0.08742278 0.08558403 0.1138314 0.02107642
[6,] -5.885910 0.3454955 0.10686380 0.09798301 0.1352105 0.02521885
[7,] -5.873156 0.3665756 0.09511612 0.08978781 0.1234700 0.02291550
[8,] -5.787396 0.3433829 0.10946598 0.10396580 0.1377752 0.02579887
[9,] -5.783250 0.4249930 0.09019019 0.09446539 0.1301408 0.02422680
[10,] -5.722045 0.4088995 0.07850427 0.08958123 0.1109272 0.02059138

```

Na figura 3.9 ilustramos os valores estimados, na escala original dos dados, do ajustamento (fazendo $\sigma = 0$, que reflete a ausência de variabilidade ambiental e substituindo os parâmetros pelos seus estimadores de MV) e previsões a 25 anos, no período de 2000 a 2024 (em cima). Mostramos ainda, em pormenor, as previsões PP e a LP (a 10 anos: de 2000 a 2009) para as TBMs, com \mathcal{IC} associados, obtidos por duas vias: através do método Delta (que ilustrámos a partir da expressão (3.20) e seguintes) e com recurso a simulação (que já havíamos utilizado no caso do MBG e que aplicámos de modo semelhante). Neste caso, os resultados dizem respeito à idade 39 do sexo feminino, mas foi possível estimar os $\mathcal{IC}_{95\%}$ associados às previsões das TBMs de todas as idades.

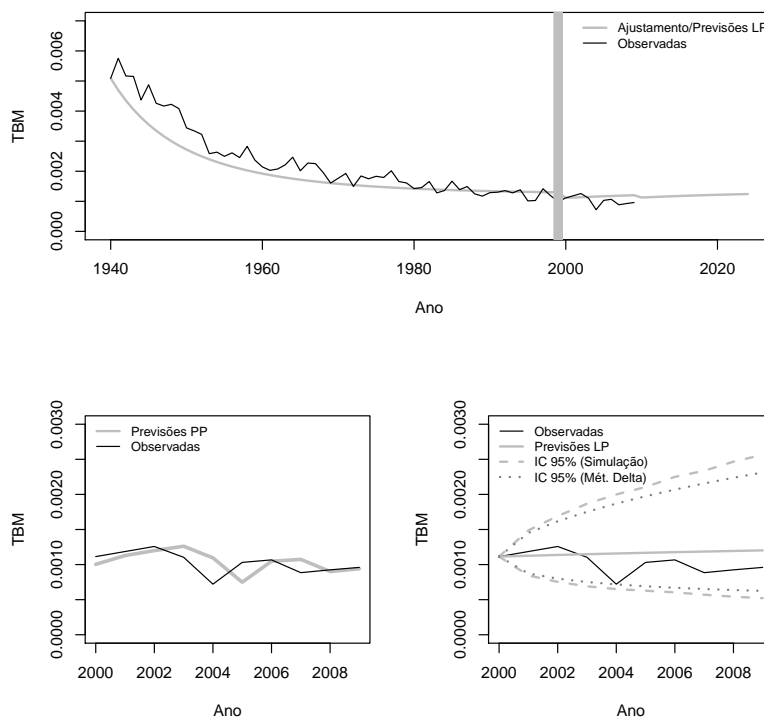


Figura 3.9: Ajustamento do MGE com previsões a LP (25 anos: de 2000 a 2024) para a idade 39 do sexo feminino (em cima); previsões, no período de 2000 a 2009, PP e a LP com $IC_{95\%}$ assintóticos (respetivamente, à esquerda e à direita, em baixo)

Analisando globalmente os resultados da aplicação do MGE, consideramos que, de um modo geral, são bastante satisfatórios. O modelo ajusta-se bem aos dados e fornece boas previsões (segundo os critérios usuais). Na caixa 3.6, mostramos as previsões a LP (para os anos 2000 a 2009), para a idade 39, por sexo, antecedidas pelos valores das TBMs, de modo a melhor enquadrar a ordem de grandeza dos valores estimados.

Caixa 3.6 MGE: previsões a LP, para o período 2000 a 2009, das TBMs da idade 39 e por sexo

```
> tx.prev.MGE <- cbind(Taxas.F[60:70,40],Prev.F[,40],Taxas.M[60:70,40],Prev.M[,40])
      Observ.F prevLP.F Observ.M prevLP.M
[1,] 0.001114 0.000997 0.003187 0.003099
[2,] 0.001186 0.001007 0.003007 0.003112
[3,] 0.001258 0.001017 0.003245 0.003124
[4,] 0.001104 0.001027 0.002608 0.003135
[5,] 0.000721 0.001036 0.002316 0.003145
[6,] 0.001031 0.001045 0.002654 0.003154
[7,] 0.001067 0.001053 0.002535 0.003163
[8,] 0.000885 0.001061 0.002080 0.003171
[9,] 0.000925 0.001069 0.002214 0.003179
[10,] 0.000960 0.001076 0.002094 0.003186
```

Considerando a qualidade do ajustamento, quer o ajustamento propriamente dito, quer as previsões são em geral melhores nas séries do sexo feminino (tal como no MBG). Esta diferença entre sexos é mais expressiva depois da idade 80, sendo que é também neste grupo de idades - e depois da idade 95 - que ocorre pontualmente (e apenas nas previsões) um maior EQM associado ao sexo feminino (ver, respetivamente, as figuras 3.10 e 3.11). Na caixa 3.7, mostramos os valores dos EQMs para o ajustamento e para as previsões a LP e PP para as idades 30 a 39, por sexo.

Caixa 3.7 MGE: EQM do ajustamento às TBMs e das previsões PP e a LP (2000 a 2009) das TBMs das idades 30 a 39 de ambos os sexos

```
> EQM.MGE
      EQM.ajust.F  EQM.ajust.M  EQM.prevLP.F  EQM.prevLP.M  EQM.prevPP.F  EQM.prevPP.M
[1,] 3.512614e-07 3.680007e-07 1.982331e-08 1.369462e-06 7.948045e-09 1.010044e-07
[2,] 2.711189e-07 3.311887e-07 6.220008e-08 9.216094e-07 8.820688e-09 9.565390e-08
[3,] 3.510966e-07 3.430556e-07 6.584368e-08 6.936619e-07 8.792051e-09 4.865990e-08
[4,] 2.689496e-07 2.850296e-07 9.235446e-08 7.085137e-07 7.955384e-09 8.569175e-08
[5,] 2.118849e-07 3.953059e-07 7.734999e-08 4.550797e-07 1.749258e-08 1.055565e-07
[6,] 3.227023e-07 5.612827e-07 9.842296e-08 9.998682e-07 2.062012e-08 1.417330e-07
[7,] 3.391223e-07 5.337844e-07 2.599870e-08 6.938674e-07 3.504697e-08 5.151480e-08
[8,] 4.537314e-07 7.396971e-07 1.685915e-07 8.361819e-07 2.545801e-08 4.950123e-08
[9,] 3.509596e-07 7.680497e-07 2.520924e-08 4.147590e-07 1.322905e-08 7.169715e-08
[10,] 2.085337e-07 6.668248e-07 2.741231e-08 4.956685e-07 3.011183e-08 9.330526e-08
```

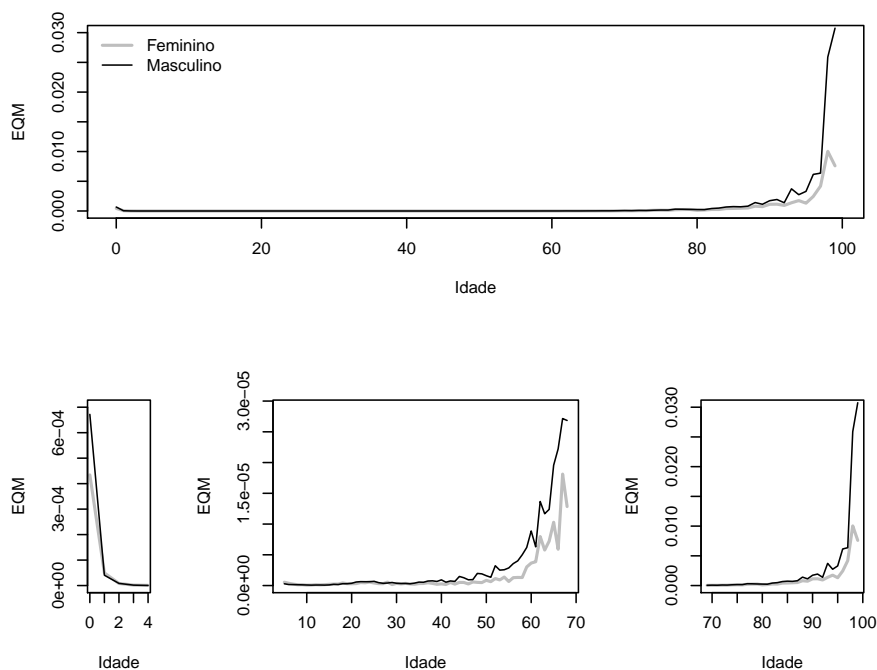


Figura 3.10: EQM do ajustamento às TBMs do MGE, por idade e por sexo (em cima, representação para todas as idades; em baixo, ampliação por grupos de idades)

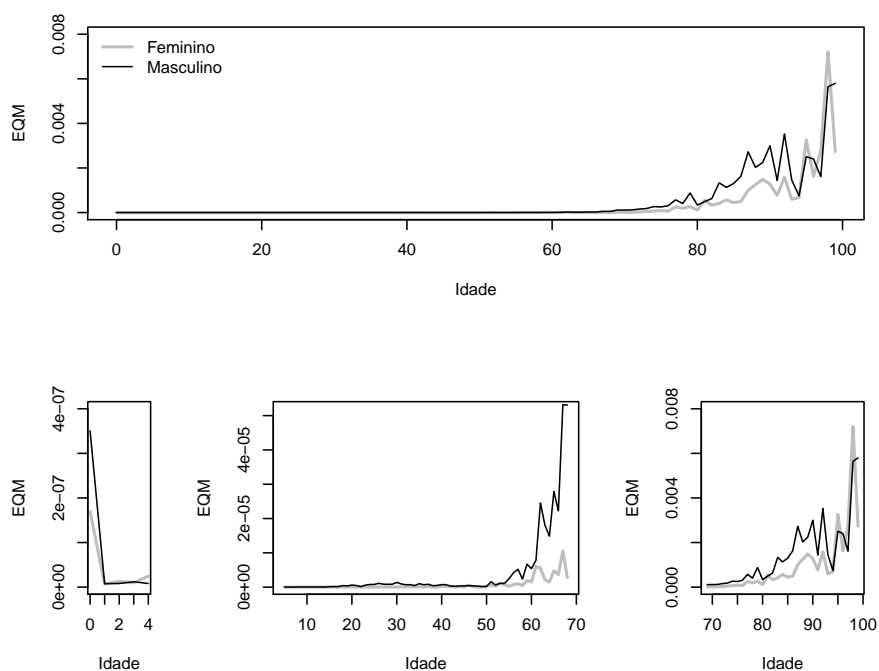


Figura 3.11: EQM das previsões a LP das TBMs, para o período 2000 a 2009, do MGE, por idade e por sexo (em cima, representação para todas as idades; em baixo, ampliação por grupos de idades)

Pelo exposto, e à semelhança dos resultados obtidos na secção anterior relativamente ao MBG, também o MGE nos parece adequado para modelar este tipo de dados, considerando igualmente os resultados promissores, enquanto contribuição para as etapas seguintes da investigação.

Na secção seguinte, porque a aplicação da metodologia é inovadora neste tipo de dados e porque queremos validar os resultados antes de aumentar a complexidade dos modelos de EDEs, vamos comparar entre si o desempenho dos dois modelos de EDEs aplicados unidimensionalmente. Adicionalmente, vamos também comparar o desempenho de um desses modelos - neste caso o modelo de Gompertz - e um modelo de regressão não linear (RNL) e também com modelos do tipo ARIMA (modelos clássicos de análise de séries temporais, que embora amplamente difundidos não têm vindo a ser muito aplicados a este tipo de dados).

3.4 Comparação dos resultados entre modelos

Recordemos que o objetivo deste trabalho é captar tendências evolutivas nas séries temporais das TBMs, por idade e por sexo, e fazer previsões da mortalidade para Portugal. Nesta secção, e no sentido de validarmos os resultados, para além de compararmos entre si os resultados dos dois modelos de EDEs aplicados nas duas secções anteriores, comparamo-los também com os obtidos por duas metodologias alternativas de modelação, que também se adequam a este tipo de dados. São eles um modelo geral (com parâmetros diferentes para cada idade e sexo) de RNL, que

representa uma curva da resposta média da parte determinística do MGE (utilizado como em [25]), e modelos do tipo ARIMA (uma escolha habitual, dentro dos métodos estocásticos, para modelar séries temporais).

3.4.1 MBG vs MGE

No decurso da investigação, equacionámos (e experimentámos) variações do MBG e do MGE (outras transformações de variáveis, redução do horizonte temporal ao encontro de um período estável de cada série, ou, ao invés, separação e estimação dos parâmetros consoante as possíveis fases do horizonte temporal em análise). Não reconhecendo vantagens significativas, em termos de resultados, desses modelos alternativos, optámos por apresentar os resultados obtidos através da aplicação destes dois modelos de EDEs tal como foram descritos nas secções anteriores.

Recordamos que foi possível ajustar ambos os modelos e fazer previsões para todos os grupos etários anuais (que temos vindo a designar por idades) dos 0 aos 99 anos dos dois sexos, sendo que consideramos que ambos apresentam previsões realísticas com valores na mesma ordem grandeza e com EQMs próximos, que não nos permitem afirmar, numa primeira análise, que, quer em termos de previsão quer de ajustamento, um modelo seja, em geral, melhor do que o outro.

A figura 3.12 ilustra a aplicação dos dois modelos de EDEs à idade 19, para ambos os sexos (os resultados são apresentados na escala original dos dados).

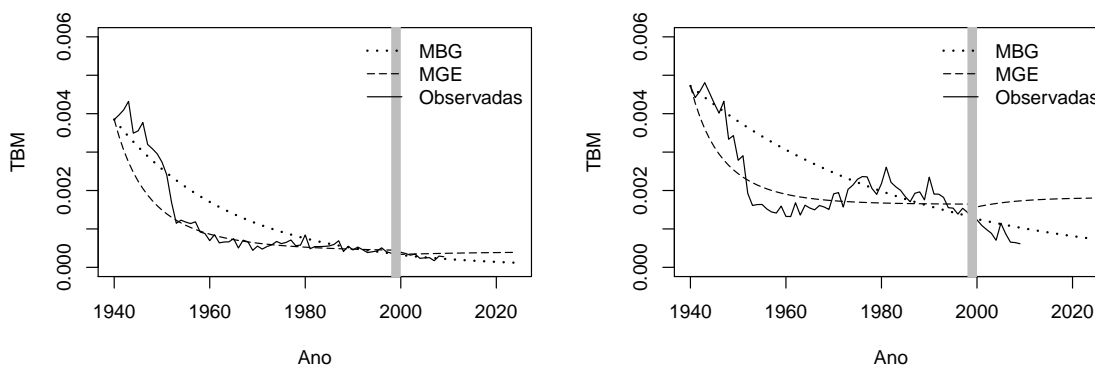


Figura 3.12: Comparação entre o ajustamento do MBG e do MGE com previsões a LP (previsões a 25 anos, de 2000 a 2024) para a idade 19 do sexo feminino (à esquerda) e do masculino (à direita)

Selecionámos esta idade (idade 19), porque constitui o típico exemplo do “comportamento” dos valores estimados, quer em termos de ajustamento, quer no que se refere à tendência das previsões, no que distingue o MBG do MGE. Assim, para a generalidade das idades, e para ambos os sexos, o ajustamento pode ilustrar-se por um gráfico semelhante ao do lado esquerdo da figura 3.12. Note-se que, a curva estimada pelo MBG só acompanha a variabilidade da série no início e final do período de ajustamento, enquanto que a do MGE, embora não acompanhe a curva dos valores observados nos primeiros anos, capta mais cedo a variabilidade da série que o MBG. No gráfico do lado direito,

é visível a exceção a estes comportamentos. Sensivelmente entre as idades 17 a 37 do sexo masculino ocorre um efeito tipo “bossa” (neste caso, entre 1960 e 1999, mas este período pode variar consoante a idade) que reflete um aumento da mortalidade neste grupo etário e que provoca a principal diferença no padrão da mortalidade entre sexos.

Em termos de previsões, na generalidade das idades (exceção para as idades depois da idade 85, em que as previsões não têm tendência expressiva, pois tendem para a média da série), o MBG subestima com tendência decrescente (mais ou menos forte, consoante a idade), enquanto que o MGE sobrestima com tendência crescente (como, de resto, também é observável na figura 3.12).

Embora o desempenho de nenhum dos modelos sobressaia explicitamente relativamente ao outro, se analisarmos para ambos a diferença entre os respetivos EQMs, para cada idade e por sexo, o MBG apresenta vantagens face ao MGE. De facto, quer para o ajustamento (exceção para algumas idades, a maior parte das quais entre os 25 e os 49 anos e também depois dos 85, do sexo masculino), quer para as previsões PP ou a LP (neste caso as exceções são ainda mais pontuais e sobretudo nas últimas duas ou três idades) constatamos que há uma tendência geral para que o erro associado ao MBG seja inferior ao associado ao MGE.

Nos gráficos das figuras 3.13, 3.14 e 3.15, representam-se precisamente as diferenças para todas as idades, por sexo, entre o EQM associado ao MBG e ao MGE, isto é $EQM_{MBG} - EQM_{MGE}$, para os ajustamentos, previsões PP e a LP. Note-se que, devido à ordem de grandeza das estimativas dos erros, que são frequentemente muito próximos e reduzidos para várias idades, as diferenças vêm multiplicadas por 10000.

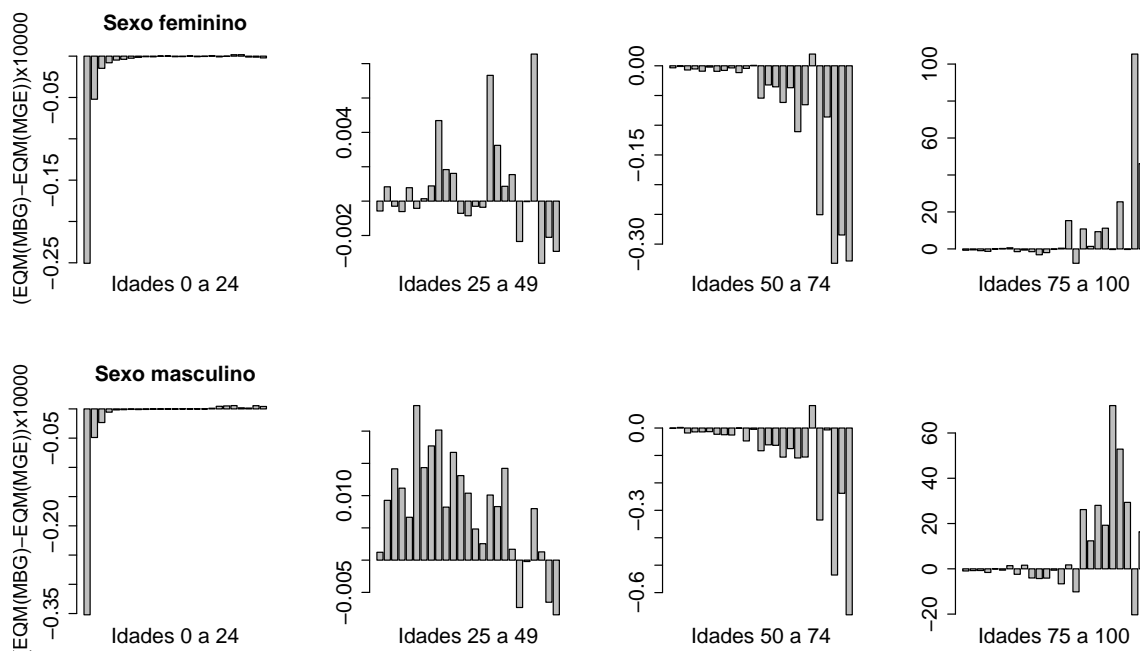


Figura 3.13: Diferença ($\times 10000$) entre os EQMs associados ao ajustamento às TBMs do MBG e do MGE, por idade; sexo feminino, em cima, e masculino, em baixo

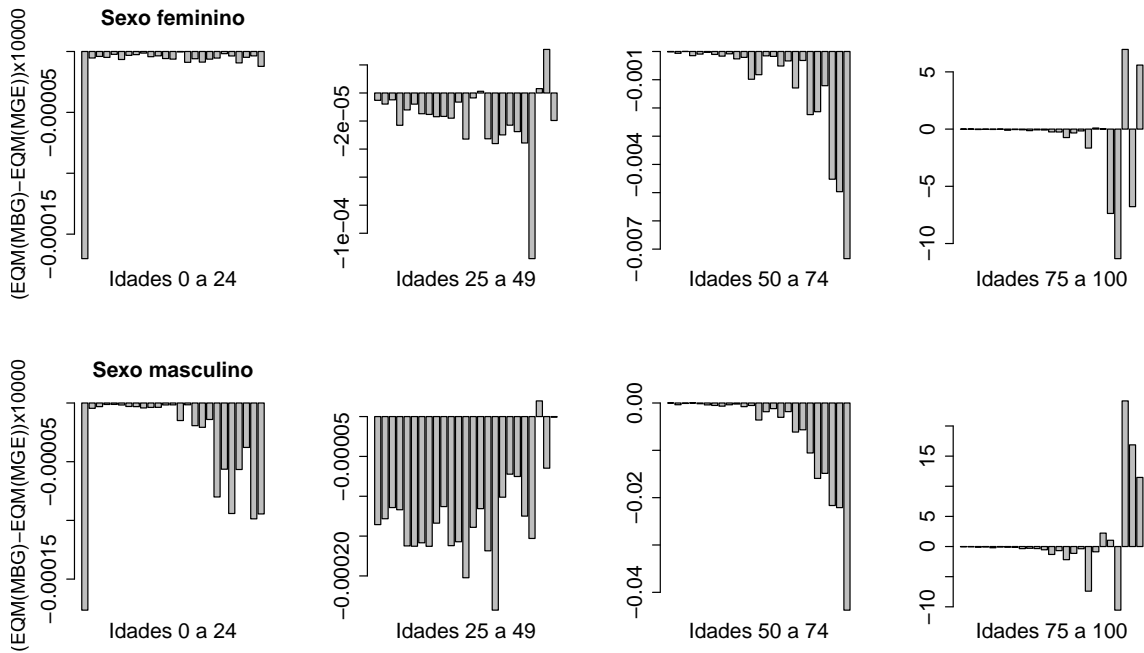


Figura 3.14: Diferença ($\times 10000$) entre os EQMs associados às previsões PP das TBMs (de 2000 a 2009) do MBG e do MGE, por idade; sexo feminino, em cima, e masculino, em baixo

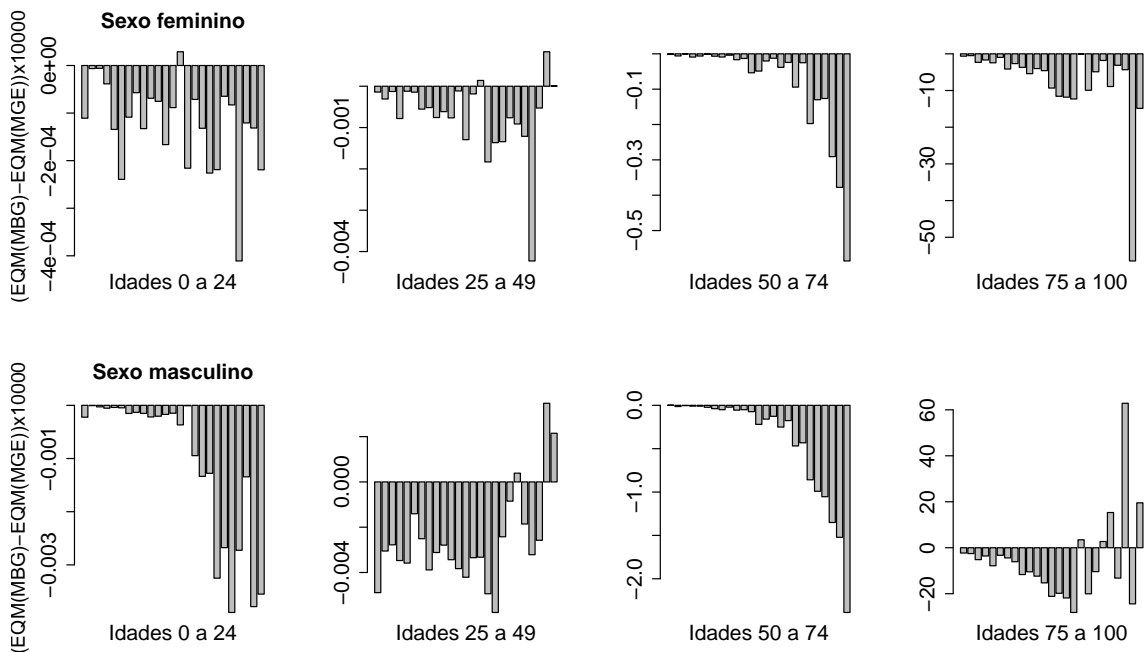


Figura 3.15: Diferença ($\times 10000$) entre os EQMs associados às previsões a LP das TBMs (de 2000 a 2009) do MBG e do MGE, por idade; sexo feminino, em cima, e masculino, em baixo

A partir da representação gráfica dos ajustamentos, constatamos que é no período inicial, digamos desde 1940 e até 1950-60, que ocorrem as maiores diferenças entre os dois modelos. Se, por um lado, os erros do MGE podem aumentar devido ao desvio das estimativas do modelo, face aos valores observados, por outro, o MBG também não acompanha a variabilidade dos dados nesse período (as estimativas do modelo apenas lhes são mais próximas).

Pelo exposto, porque o MGE é claramente mais estável a partir dos anos 60 do século passado e apresenta um melhor ajustamento, não podemos afirmar que o MBG é indiscutivelmente o melhor modelo para aplicar a este tipo de dados. Para algumas idades, poderá haver vantagem em aplicar um ou outro, mas como nos capítulos seguintes pretendemos aplicar um único modelo base para ambos os sexos e/ou para todas as idades, não nos vamos restringir a escolher apenas um modelo nesta fase.

Face à possível contribuição do período inicial em análise (digamos de 1940 a 1960) para uma menos boa prestação do MGE, questionámo-nos acerca das vantagens de usar séries mais curtas, por exemplo, a partir de 1960, em vez da série completa, mas também aí não foi possível tomar uma decisão consensual, para todas as idades e ambos os sexos, sobre qual o melhor modelo. Na maior parte dos casos, aumenta a variabilidade ao reduzir-se o número de observações e o EQM associado a ajustamentos e previsões pode mesmo ser superior quando se aplicam estes modelos às séries mais curtas (veja-se a figura 3.16), pelo que não há vantagens, para estes modelos de EDEs, em não usar as séries completas.

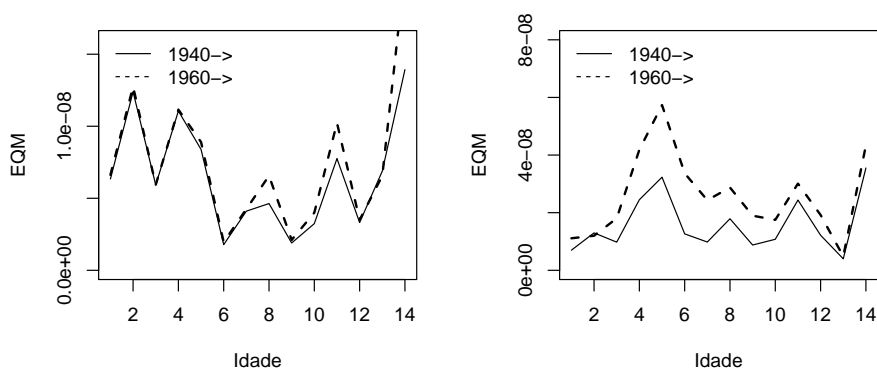


Figura 3.16: Comparação entre o EQM das previsões a LP das TBMs (de 2000 a 2009) do MBG (à esquerda) e do MGE (à direita) para as idades 1 a 15 do sexo feminino

Para finalizar este capítulo, apresentamos, nas subsecções seguintes, os comparativos entre os resultados obtidos pelos modelos de EDEs e os obtidos via modelos de RNL e do tipo ARIMA.

3.4.2 Modelos de EDEs vs RNL

Consideremos que as taxas variam com o tempo e que queremos modelar uma v.a., Y , observando $(t_1, y_1), (t_2, y_2), \dots, (t_N, y_N)$. Podemos modelar Y , enquanto descritor da mortalidade, para cada

idade e por sexo, através de um modelo de regressão não linear (RNL), que tem por base o mesmo modelo determinístico que o MGE definido em (3.14) e tem também erros associados a cada “observação”. Um tal modelo, pode escrever-se na forma

$$Y_n = f(t_n, \mathbf{p}) + \epsilon_n, \quad (3.21)$$

em que f é uma função de regressão à qual se soma o erro ϵ_n . A função determinística f terá a mesma forma que a média do modelo (3.14), isto é, seria a forma (3.15) com $\sigma = 0$. f resulta da combinação não linear dos parâmetros do modelo ($\mathbf{p} = (A, b, y_0)$, representa o vetor dos parâmetros) e depende da variável explicativa tempo, tal que

$$f(t_n, \mathbf{p}) = A + (y_0 - A) \exp\{-b(t_n - t_0)\},$$

com $n = 1, \dots, N$. Os erros ϵ_n constituem uma sequência de v.a. independentes e identicamente distribuídas, com distribuição normal com média zero e variância, σ_ϵ^2 , isto é, $\epsilon_n \sim \mathcal{N}(0, \sigma_\epsilon^2)$ i.i.d..

Para estimar os parâmetros do modelo, A, b e y_0 , usamos o método dos mínimos quadrados (os aspetos teóricos relacionados com este método, bem como os extensivos à regressão não linear, estão exaustivamente descritos na bibliografia de referência - veja-se, por exemplo, [4]).

Como queremos comparar este modelo de RNL com o referido MGE, o ajustamento e as previsões do modelo de RNL são dados por

$$\hat{Y}_n = \hat{A} + (\hat{y}_0 - \hat{A}) \exp\{-\hat{b}(t_n - t_0)\}.$$

Foi possível ajustar o modelo a todas as idade e por sexo (para o efeito, usámos a função \exp do \mathbb{R} , como em [29]). Os parâmetros obtidos nos ajustamentos serviram de base às previsões a LP para o período de 2000 a 2009. Já nas previsões PP, em que era preciso estimar novamente os parâmetros em cada passo (por se ter mais uma observação), ocorreram dificuldades, relacionadas com os valores de arranque e a ordem de grandeza dos parâmetros a estimar, tendo sido possível estimar os parâmetros apenas para algumas idades (como é o caso, por exemplo, das idades cujos resultados ilustramos na figura 3.17). Por esse motivo, não mostraremos gráficos comparativos dos EQMs associados às previsões PP obtidas via RNL.

Ajustado o modelo, para cada idade e por sexo, foi feito o diagnóstico do mesmo, nomeadamente, testando os resíduos, de modo a verificar os pressupostos iniciais. Muitas das técnicas utilizadas para o efeito no caso da regressão linear (sobretudo no que toca à representação gráfica) são também úteis para modelos de regressão não linear como o aplicado (para uma descrição mais pormenorizada destas técnicas, bem como dos testes usuais, veja-se, pelo seu carácter prático, [77] e também [24]). A análise dos resíduos revela alguns problemas (que podem ou não ocorrer em simultâneo) em várias idades: a não normalidade, não independência e ausência de média nula e variância constante.

A figura 3.17 ilustra os resultados da aplicação do modelo de RNL e do MGE, para três idades distintas: idade 0 do sexo feminino e idades 50 e 84 do sexo masculino. Fazemos a representação simultânea, para os ajustamentos, previsões a LP (neste caso a 10 anos, de 2000 a 2009) e previsões

PP, no mesmo período. Como podemos observar na figura, para as idades selecionadas, não há uma tendência comum aos dois modelos. No caso das previsões a LP, enquanto que na idade 0 do sexo feminino os valores estimados estão mais próximos das TBMs “observadas”, na idade 84 do sexo masculino inverte-se esta tendência (o modelo de RNL é melhor), sendo que os valores estimados através do MGE têm inclusivé uma tendência crescente, contrária à das taxas observadas. Para as idades consideradas, apenas nas previsões PP, os valores estimados pelo MGE são sempre mais próximos dos observados (na idade 50 do sexo masculino, inclusivé, quase que se sobrepõem).

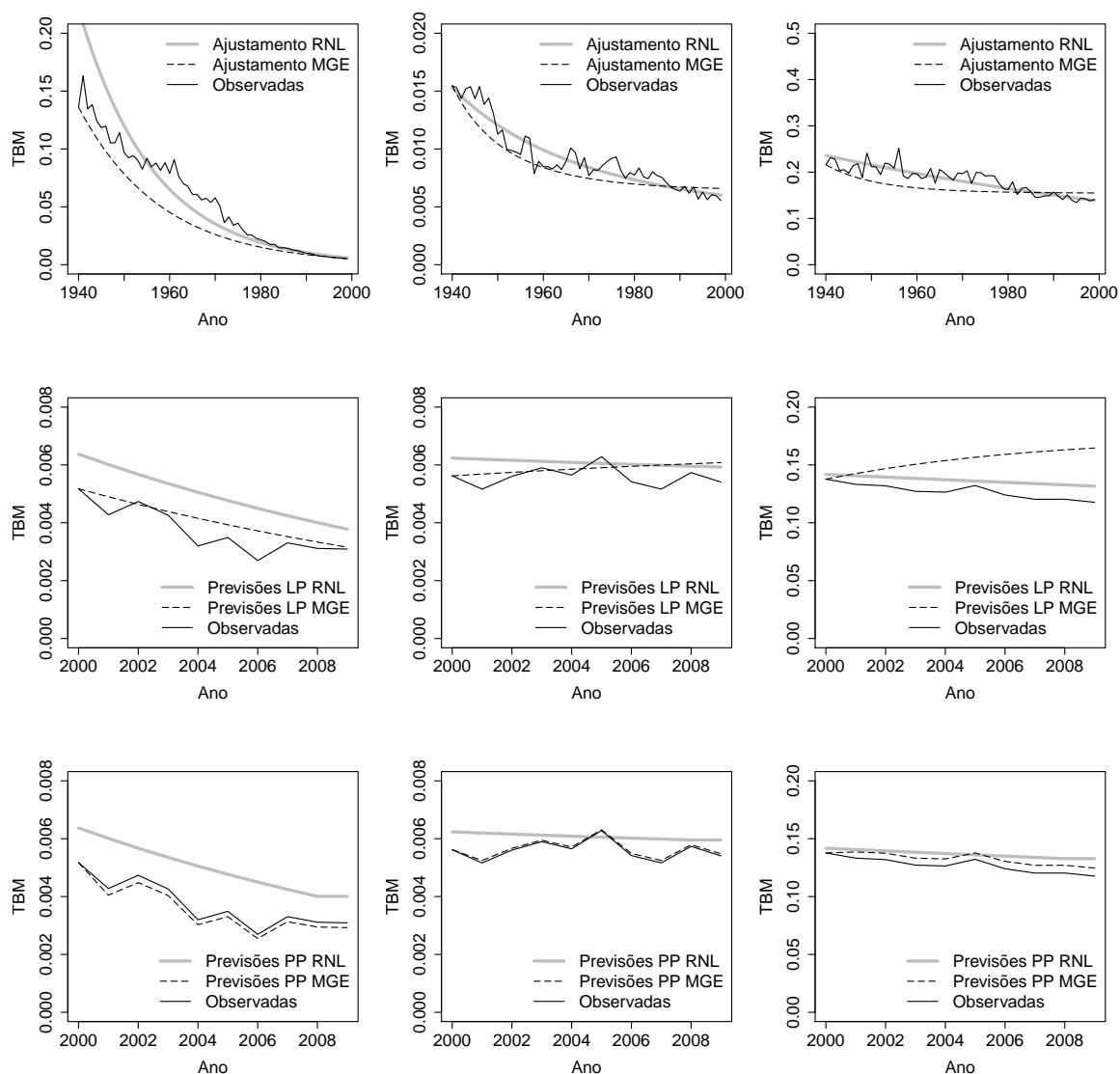


Figura 3.17: Comparação (de cima para baixo) dos ajustamentos, previsões a LP e previsões PP (ambas para o período de 2000 a 2009), entre o modelo de RNL e o MGE para as idades 0 do sexo feminino e 50 e 84 do sexo masculino (da esquerda para a direita)

Na tabela 3.1, comparam-se as estimativas dos parâmetros obtidas para o MGE e para o modelo de RNL, para as idades cujos resultados ilustrámos na figura 3.17. Note-se que a disparidade entre os valores estimados do parâmetro A e dos restantes parâmetros vem do facto de se representar o

logaritmo da taxa de mortalidade assintótica (daí que, se fizéssemos $\exp\{A\}$, íriamos obter valores muito próximos de zero, o que vem de acordo com as TBMs daquelas idades).

Ainda na tabela 3.1, quando comparamos as estimativas dos parâmetros obtidos pelos dois modelos, notamos que os valores de A e b apresentam grandes diferenças. Estas devem-se, em parte, à forma como o cálculo é efetuado, pois, inicialmente, foram dados como valores de arranque para o modelo de RNL as estimativas dos parâmetros do MGE, para cada idade e por sexo. No entanto, devido à ordem de grandeza dos dados, para haver convergência, foi necessário alterar esses valores bem como a escala de cálculo (o que se traduz nos resultados apresentados).

Tabela 3.1: Comparação entre as estimativas dos parâmetros dadas pelo MGE e pelo modelo de RNL para as idades 0 do sexo feminino (F0) e 50 e 84 do sexo masculino (M50 e M84, respetivamente)

Idade	Modelo	A	b	y_0
F0	RNL	-4014.543	1.524e-05	-1.5084
	MGE	-355.808	1.5575e-04	–
M50	RNL	-26.791	0.000712	-4.969
	MGE	-5.6717	0.091103	–
M85	RNL	-153.084	6.326e-5	-1.32515
	MGE	-1.7665	9.641e-02	–

Ao usarmos o método dos mínimos quadrados no modelo de RNL, não constitui também surpresa que os erros associados ao ajustamento sejam sempre inferiores aos resultantes da aplicação do MGE, pois aquele método estima os parâmetros precisamente de modo a minimizar a soma dos quadrados dos erros. Na figura 3.18, relativa ao ajustamento, podemos observar a diferença entre o EQM associado ao MGE e ao modelo de RNL, isto é, $EQM_{MGE} - EQM_{RNL}$. Com exceção do grupo de idades entre os 0 e os 24 anos, esta diferença aumenta com a idade e é maior para o sexo masculino.

Já no que toca às previsões, os gráficos da figura 3.19 para a diferença referida no parágrafo anterior (mas agora para as previsões a LP, no período que temos vindo a usar, de 2000 a 2009), sugerem também um melhor desempenho do modelo de RNL comparativamente ao MGE na globalidade das idades e em ambos os sexos. No entanto, inverte-se esta relação sobretudo nas últimas idades (depois dos 85 anos e em ambos os sexos), nas idades 15 a 25 do sexo masculino e também em algumas das restantes idades, isoladamente. Em qualquer caso, não esqueçamos que as diferenças são pequenas, surgindo ampliadas nos gráficos por efeito da multiplicação por 10000.

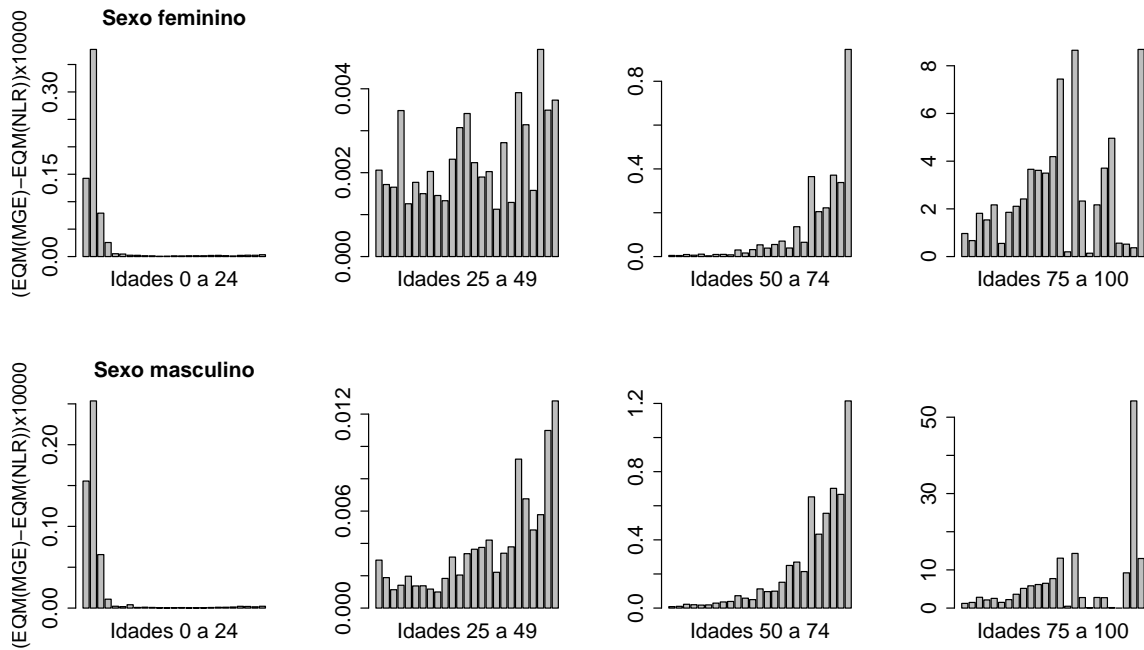


Figura 3.18: Diferença ($\times 10000$) entre os EQMs associados ao ajustamento às TBMs do MGE e do modelo de RNL, por idade; sexo feminino, em cima, e masculino, em baixo

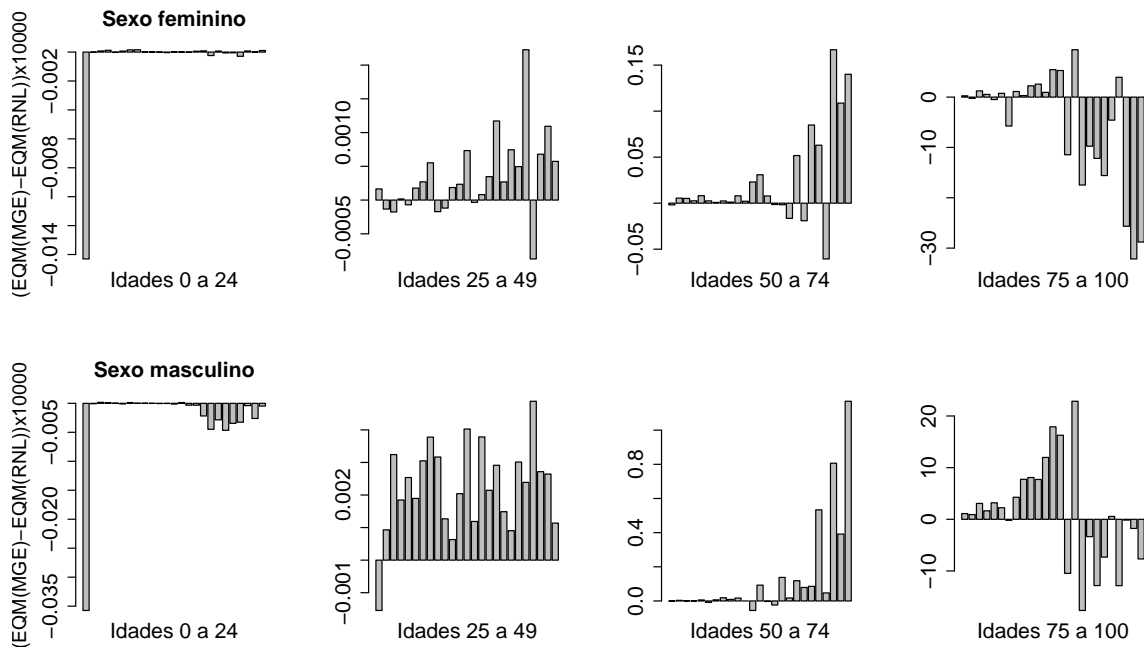


Figura 3.19: Diferença ($\times 10000$) entre os EQMs associados às previsões a LP das TBMs (de 2000 a 2009) do MGE e do modelo de RNL, por idade; sexo feminino, em cima, e masculino, em baixo

3.4.3 Modelos de EDEs vs ARIMA

A análise de séries temporais e as suas aplicações tem vindo a ganhar importância em várias áreas de investigação, nomeadamente na economia, agronomia, meteorologia, engenharia, medicina e também nas ciências sociais e políticas. Apesar da mortalidade ser uma variável demográfica muito estudada, através de modelos determinísticos e, mais recentemente, estocásticos, existem ainda poucas referências (muito devido à abordagem usual ser longitudinal e não necessariamente transversal ao longo do tempo) a trabalhos que utilizem a análise de séries temporais para fazer previsões da mortalidade [2, 22, 45, 85].

Neste ponto, consideramos que a análise das séries temporais constitui um bom ponto de partida para ajudar a explicar o fenómeno em estudo (não obstante a utilização dos modelos de EDEs, em cuja aplicação incide a temática principal deste trabalho). Assim, nesta secção, mostramos os resultados da aplicação, às TBMs, de modelos autorregressivos integrados e de médias móveis (ARIMA), pressupondo que as observações presentes dependem das passadas e/ou de impulsos aleatórios. Comparamos então os resultados obtidos por esta metodologia com os obtidos através do MGE e do MBG.

Os modelos ARIMA dizem-se de ordem (p, d, q) , com p e q a ordem dos termos autorregressivos e de médias móveis, respetivamente, e d a ordem da diferenciação. Estes modelos podem escrever-se na forma

$$(1 - B)^d Z(t) = \mu + \phi_1 Z(t-1) + \dots + \phi_p Z(t-p) + \epsilon(t) - \theta_1 \epsilon(t-1) - \dots - \theta_q \epsilon(t-q), \quad (3.22)$$

com $t, d \in \mathbb{Z}$, B o operador atraso que pode definir-se por $B^d = Z(t) - Z(t-d)$, μ a constante do modelo, (ϕ_1, \dots, ϕ_p) e $(\theta_1, \dots, \theta_p)$ os parâmetros associados aos polinómios autorregressivo e de médias móveis, respetivamente, tal que $\phi_p, \theta_q \in \mathbb{R} \setminus \{0\}$ e ϵ_t é um processo ruído branco (WN), de média nula e variância constante, isto é, $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ ($\sigma_\epsilon^2 > 0$).

Como a primeira condição para a aplicação da metodologia é a estacionaridade, verificou-se que a transformação que resulta melhor, para o efeito e para a maior parte das idades de ambos os sexos, é a transformação logarítmica. Assim, em (3.22) usámos $Z(t) = \ln(X(t))$, com $X(t)$ a TBM correspondente a uma certa idade e sexo. Também para a maioria das séries, aos logaritmos das TBMs aplicámos uma diferenciação simples (ou seja, $d = 1$), o que corresponde a $B = \ln(X(t)) - \ln(X(t-1))$.

A identificação do tipo de modelo, autorregressivo (AR) ou de médias móveis (MA), e da respetiva ordem, é normalmente conseguida através da combinação de métodos empíricos e analíticos, que incidem sobretudo no estudo da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP). Os aspetos metodológicos relativos a essa identificação, bem como a estimação dos parâmetros e avaliação de diagnóstico (avaliação da significância dos parâmetros, da estabilidade do modelo ajustado e da qualidade do ajustamento) estão exaustivamente expostos na bibliografia de referência. Em particular, as estimativas dos parâmetros, para cada idade e por sexo, foram obtidas por MV, utilizando a função `arima` do R (veja-se, por exemplo, a propósito da teoria e da aplicação dos modelos ARIMA, [7, 16, 35, 43, 64, 79, 86]).

No caso das TBMs da população portuguesa, foi possível aplicar modelos ARIMA a cada uma das idades selecionadas, por sexo, que são, na maioria dos casos, estatisticamente significativos e válidos.

Considerando os critérios usuais de estabilidade associados ao ajustamento deste tipo de modelos e o estudo dos resíduos, ocorrem em algumas idades (frequentemente, depois da idade 85, mas não só), problemas relacionados com a heterocedasticidade dos resíduos ou a sua não independência, ainda que as hipóteses de homocedasticidade testadas tenham sido rejeitadas no limiar dos níveis de significância usuais.

À semelhança da subsecção anterior (relativamente aos erros associados ao modelo de RNL comparados com os do modelo de EDEs), também neste caso, não apresentamos o comparativo que diz respeito à diferença entre os erros associados às previsões PP dadas pelos modelos de EDEs e as obtidas via modelos do tipo ARIMA. O processo revelou-se computacionalmente exaustivo (no caso desta última metodologia), pois em cada idade e por sexo as ordens p e q variavam, bem como a opção por incluir ou não uma constante no modelo. Por este motivo, porque não era um objetivo principal da investigação e porque consideramos ser mais vantajoso automatizar o processo através do desenvolvimento de um código mais eficiente (a fim de evitar a dispersão e permitir obter resultados mais robustos), optámos por não apresentar, neste ponto, os resultados das previsões PP.

De qualquer modo, consideramos que os modelos ARIMA ajustados explicam razoavelmente a variabilidade das séries, acompanhando a sua tendência geral, com as previsões a LP, para o período de 2000 a 2009, a apresentarem uma ligeira tendência decrescente na maior parte das idades de ambos os sexos. Os modelos têm coeficientes diferentes para cada idade, como podemos ver na tabela 3.2, para as idades 6 e 25 do sexo feminino e idade 70 do sexo masculino.

Tabela 3.2: Estimativas dos parâmetros do modelo AR(1) aplicado aos logaritmos das TBMs das idades 6 e 25 do sexo feminino, respetivamente, F6 e F25 e à idade 70 do sexo masculino, M70 (para esta idade usou-se $\mu = 0$, uma vez que o valor estimado não era significativo)

Idade	μ	ϕ_1
F6	-0.0875	-0.6038
F25	-0.0579	-0.4474
M70	0	-0.3682

Nas tabelas F.1 e F.2 do apêndice F, apresentamos os modelos ajustados para todas as idades, considerando, porém, os grupos etários quinquenais usuais em demografia (exceção para as TBMs relativas à idade 0, cujo óbito se dá no primeiro ano de vida e das idades 1 a 4).

Na figura 3.20, ilustramos os resultados dos ajustamentos e das previsões obtidos através da aplicação de um modelo com componente autorregressiva de primeira ordem, AR(1), para as três idades cujos parâmetros estimados mostrámos na tabela 3.2. Nos gráficos representamos, simultaneamente, os resultados equivalentes obtidos via modelos de EDEs, isto é através do MBG e do MGE, de modo a ilustrar diferenças no desempenho das duas metodologias.

Neste ponto recorde-se que temos vindo a fazer o ajustamento, nos modelos de EDEs, numa só curva (como se, em cada instante, não fossem observados valores nos instantes anteriores), o que se adequou ao comparativo da subsecção anterior com os modelos de RNL. Note-se, no entanto, que para podermos comparar o ajustamento feito através dos modelos do tipo ARIMA com os modelos de EDEs, concretamente com o MBG e o MGE, foi necessário fazer o ajustamento PP nestes últimos, uma vez que um modelo ARIMA faz o ajustamento, num dado instante, com recurso à observação do instante anterior.

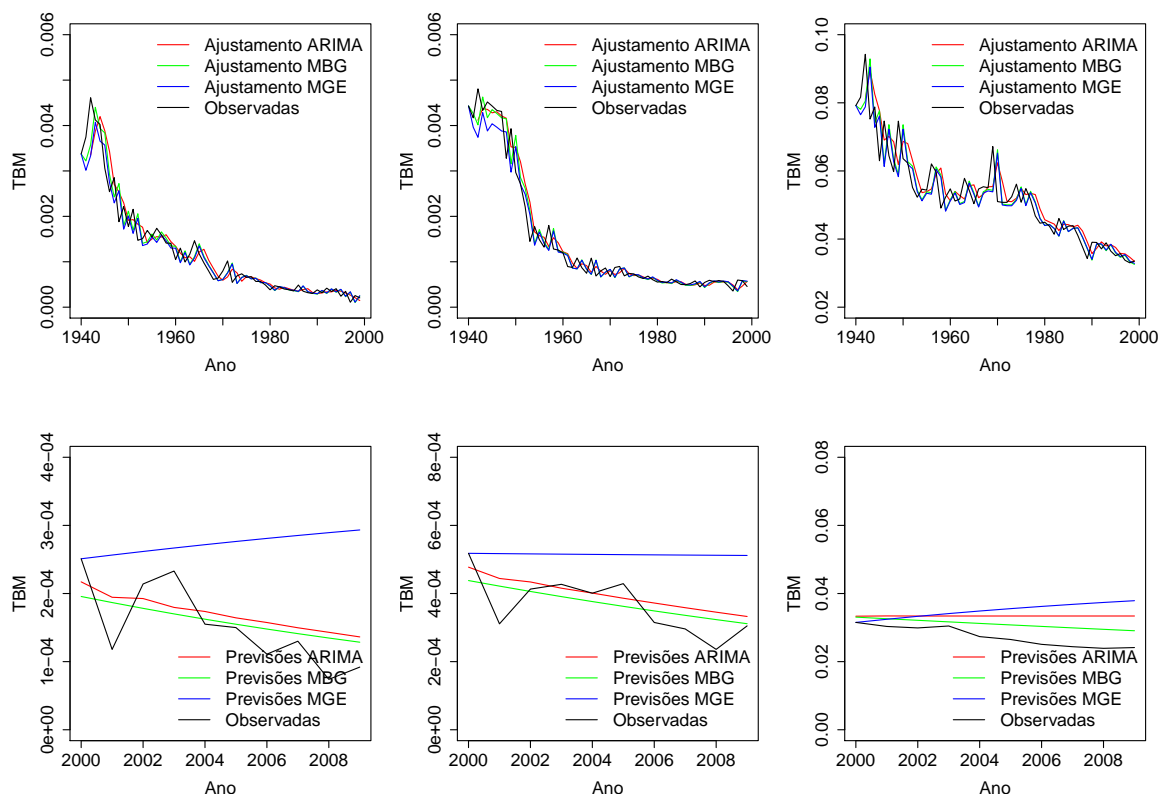


Figura 3.20: Comparação (ajustamento, em cima, e previsões a LP de 2000 a 2009, em baixo) entre o modelo ARIMA e os modelos MBG e MGE para as idades (da esquerda para a direita) 6 e 25 do sexo feminino e 70 do sexo masculino

Os gráficos da figura 3.20 sugerem, para as idades consideradas, ajustamentos semelhantes obtidos pelos três modelos e uma maior proximidade entre os valores estimados a partir dos modelos ARIMA e do MBG, no caso das previsões a 10 anos para o período de 2000 a 2009. Sobretudo no caso das previsões, é também evidente, para estas idades, um melhor desempenho dos modelos ARIMA e do MBG face ao MGE.

Consideremos agora os resultados globais da aplicação dos modelos do tipo ARIMA às séries anuais da mortalidade da população portuguesa. No apêndice F, já referido anteriormente, apresentamos uma lista dos modelos tipo ARIMA ajustados às séries das TBMs de todos os grupos etários quinquenais. Como as TBMs para os grupos etários anuais (que temos vindo a designar por

idade) pouco diferem, em termos de ordem de grandeza, tendência e variabilidade, das do grupo quinquenal a que pertencem, tão pouco diferem, globalmente, os modelos ajustados.

Assim, quanto aos modelos ARIMA, e na maior parte das idades, predominam os modelos ajustados apenas com componente autorregressiva de primeira ou segunda ordem - ainda que em algumas idades seja necessário usar uma ordem superior -, sendo que a componente de médias móveis é mais comum nas séries relativas ao sexo feminino. Refira-se ainda que, em termos de ajustamento, não há outras diferenças significativas entre os dois sexos, tendo os modelos estimados a capacidade de captar a variabilidade e mesmo acontecimentos excepcionais das séries, sobretudo, até à idade 80. Nas idades seguintes, estes modelos perdem grande parte da capacidade de captar a variabilidade anual das séries e fazem como que um ajustamento em torno dos seus valores médios, o que faz com que o EQM seja claramente superior ao estimado para as restantes idades. Esta situação verifica-se em ambos os sexos, com os erros maiores nas séries relativas ao sexo masculino (no fundo, à semelhança do observado com as outras metodologias de modelação aplicadas e cujos resultados ilustrámos nas subsecções anteriores). Relativamente às previsões, o EQM também aumenta nas idades mais avançadas e também é superior no sexo masculino, comparativamente ao feminino.

Nas figuras 3.21 e 3.22, representamos a diferença entre o EQM associado ao MBG e aos modelos do tipo ARIMA, isto é, $EQM_{MBG} - EQM_{ARIMA}$, respetivamente, para os ajustamentos PP e para as previsões a LP para o período de 2000 a 2009. Como nas secções anteriores, alterámos a escala dos resultados, multiplicando os valores estimados para a diferença por 10000 para assim fazer sobressair discrepâncias.

Os resultados das diferenças, no caso dos ajustamentos, sugerem, com exceção para algumas idades entre as consideradas, um melhor desempenho global por parte dos modelos do tipo ARIMA (sobretudo entre as idades 25 a 49 e no sexo feminino). Já no que toca às previsões, não há uma vantagem evidente, em termos de desempenho, de um dos dois tipos de modelos, analisando globalmente todas as idades de ambos os sexos. Quando muito, se atendermos aos primeiros três grupos de idades selecionadas, isto é, até à idade 74, os gráficos sugerem uma vantagem, em ambos os sexos (mais discreta depois da idade 50), para o MBG, que de novo se evidencia em termos de um melhor desempenho.

Estes resultados, em particular no que toca às diferenças entre EQMs dos ajustamentos, não são de todo surpreendentes uma vez que os modelos ARIMA assentam numa metodologia que se adequa a este tipo de dados. Ainda assim, recordamos que se tratam de erros com valores muito reduzidos e para a maior parte das idades, com a mesma ordem de grandeza nos três tipos de modelos. No entanto, e como veremos no capítulo 4, os modelos ARIMA não têm as vantagens quer em termos de modelação quer em termos de computação que os modelos de EDEs apresentam.

Vamos ainda comparar as diferenças entre os erros associados aos ajustamentos e previsões estimados a partir da aplicação do MGE e os modelos do tipo ARIMA, isto é, $EQM_{MGE} - EQM_{ARIMA}$. Representamos esses resultados, respetivamente, nas figuras 3.23 e 3.24.

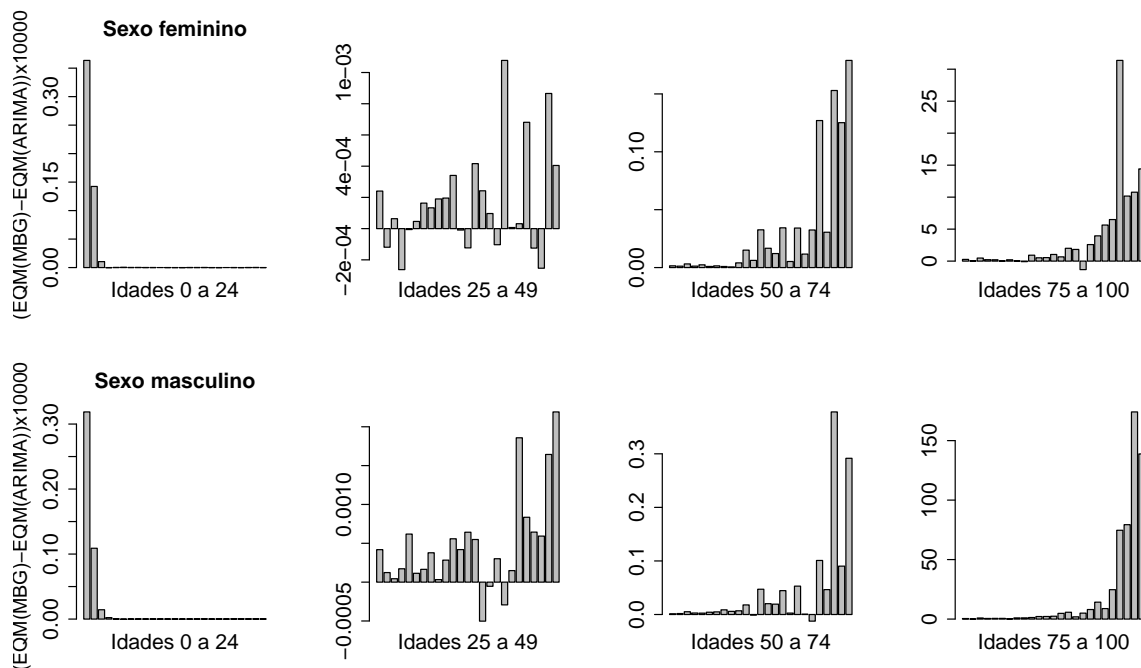


Figura 3.21: Diferença ($\times 10000$) entre os EQMs associados ao ajustamento às TBMs do MBG e do modelo ARIMA, por idade; sexo feminino, em cima, e masculino, em baixo

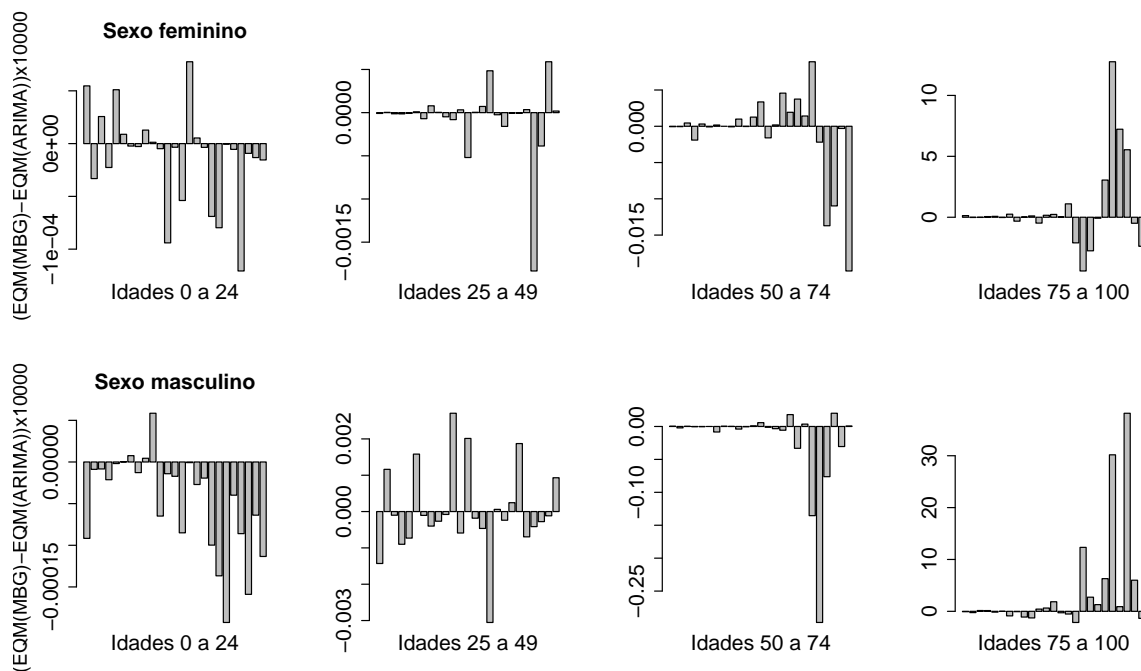


Figura 3.22: Diferença ($\times 10000$) entre os EQMs associados às previsões a LP das TBMs (de 2000 a 2009) do MBG e do modelo ARIMA, por idade; sexo feminino, em cima, e masculino, em baixo

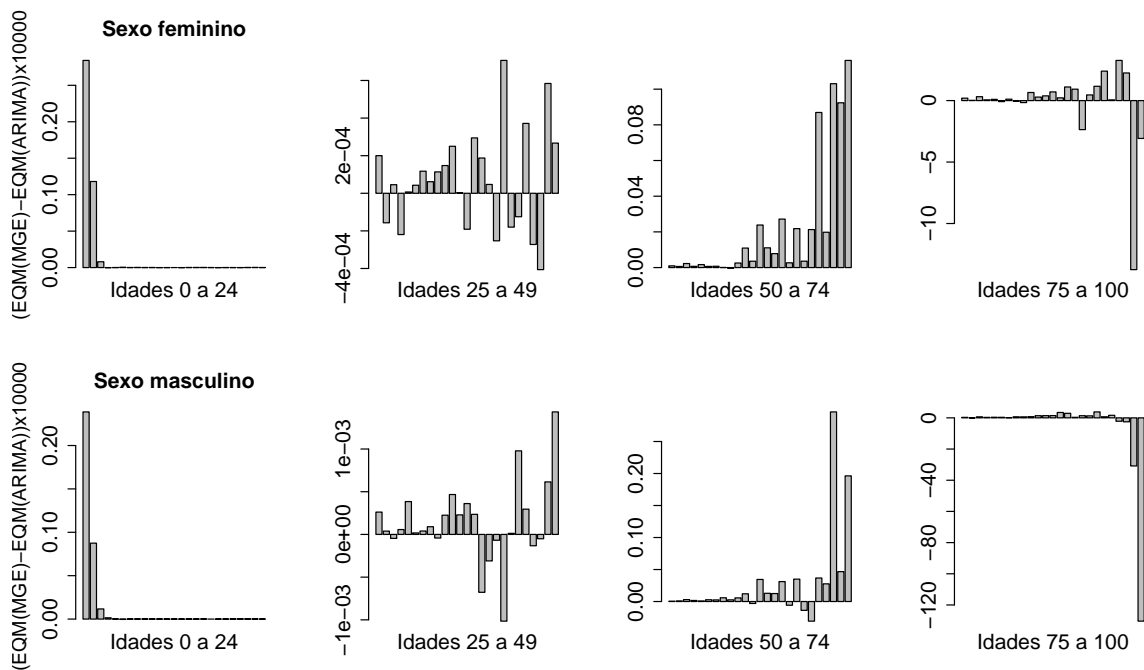


Figura 3.23: Diferença ($\times 10000$) entre os EQMs associados ao ajustamento às TBMs do MGE e do modelo ARIMA, por idade; sexo feminino, em cima, e masculino, em baixo

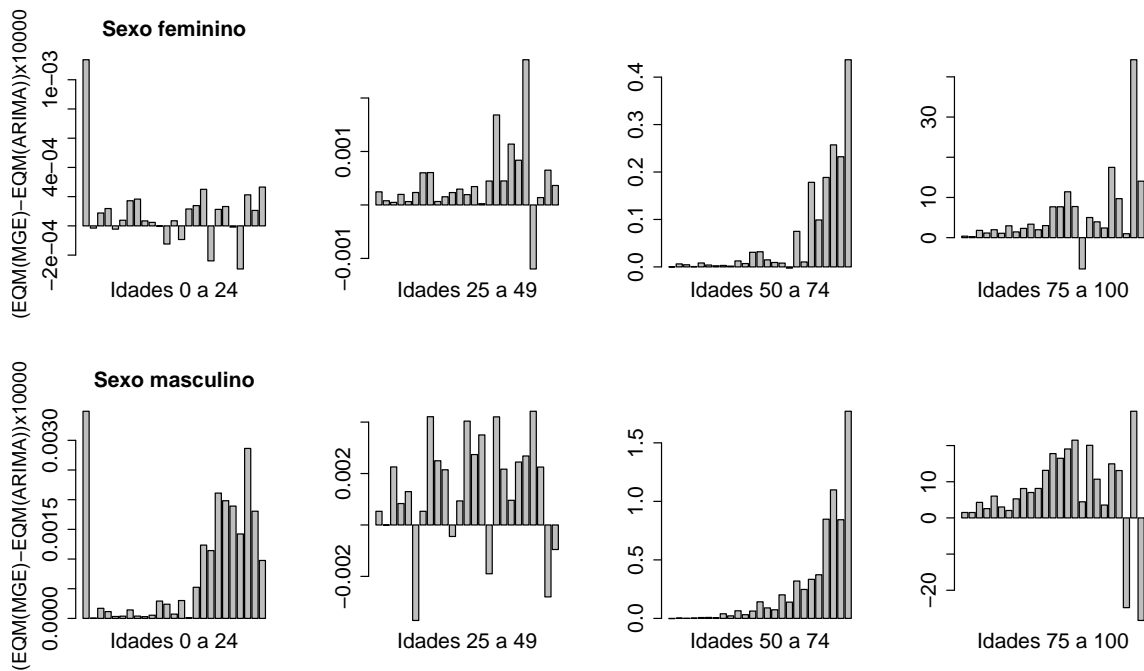


Figura 3.24: Diferença ($\times 10000$) entre os EQMs associados às previsões a LP das TBMs (de 2000 a 2009) do MGE e do modelo ARIMA, por idade; sexo feminino, em cima, e masculino, em baixo

De novo, os resultados para os ajustamentos vão ao encontro dos obtidos no comparativo anterior: os gráficos sugerem que os modelos do tipo ARIMA têm, para a generalidade das idades de ambos os sexos (exceção também para as duas últimas idades do sexo masculino), um melhor desempenho que o modelo de EDE em análise - neste caso -, o MGE. No caso das previsões a LP, embora os gráficos sugiram também uma melhor prestação dos modelos ARIMA em detrimento do MGE, ocorre uma certa inversão dessa tendência, em algumas das idades entre os 0 e os 24 anos do sexo feminino e entre os 25 e os 49 anos do sexo masculino.

Finalmente, a título de exemplo, apresentamos na figura 3.25 previsões a LP para 25 anos (de 2000 a 2024) obtidas pelos modelos de EDEs, MBG e MGE, pelo modelo de RNL e ainda por modelos do tipo ARIMA para as idades 4 e 15 do sexo masculino e 29 e 95 do feminino, com estes últimos a serem do tipo autorregressivo para as idades 4, 15 e 95 (respetivamente de ordem 2, 1 e 2, todos com constante) e com componente de médias móveis, de primeira ordem, para a idade 29. Em geral, nas primeiras e também nas últimas idades, ocorre uma maior proximidade entre os resultados obtidos pelos quatro modelos; nas idades jovens ou adultas o comportamento é bastante variável.

Os gráficos que constituem a figura 3.25 sugerem também que (nas idades seleccionadas, é exceção a idade 95 do sexo feminino), os modelos do tipo ARIMA estimam com uma ligeira tendência decrescente, aproximando-se das estimativas obtidas pelo MBG. Os resultados ilustrados na figura sugerem mesmo que estes modelos têm globalmente, para as idades consideradas, o melhor desempenho. De resto, entre os modelos de EDEs, o MGE apresenta, neste caso, resultados menos favoráveis (em parte devido à tendência crescente das previsões). Quanto aos modelos, digamos alternativos, o de RNL apresenta resultados menos favoráveis (não por uma questão de tendência, mas pelo afastamento dos valores observados disponíveis no período inicial das previsões).

Por considerarmos globalmente promissores os resultados obtidos na modelação das taxas de mortalidade através dos modelos de EDEs, MBG e MGE, baseámo-nos nestes modelos para a construção dos novos modelos que propomos nos capítulos seguintes. Deste modo, o MBG servirá de base à construção de um modelo bidimensional, com uma estrutura de correlação entre sexos para cada idade (que propomos no capítulo 4), e de um modelo multidimensional, com correlações entre idades diferentes do mesmo sexo e de sexos diferentes, que preencherá todo o capítulo 5. Quanto ao MGE, será a base para a construção de um outro modelo bidimensional (que propomos também no capítulo 4, na secção 4.3), que considera também a correlação entre sexos fixando a idade. No capítulo 4, faremos assim o paralelismo entre os dois modelos bidimensionais, à semelhança da análise apresentada para os casos unidimensionais.

Por os modelos do tipo ARIMA apresentarem aquilo que consideramos serem bons ajustamentos e resultados credíveis também nas previsões, propomo-nos ainda aplicar uma versão bidimensional dos mesmos no final do capítulo 4. Trata-se de um modelo da mesma família, mas agora um modelo de vetores autorregressivos e/ou de médias móveis, VARMA (no caso, bivariado), para fazer analogamente o estudo comparativo dos resultados com os obtidos via modelos de EDEs.

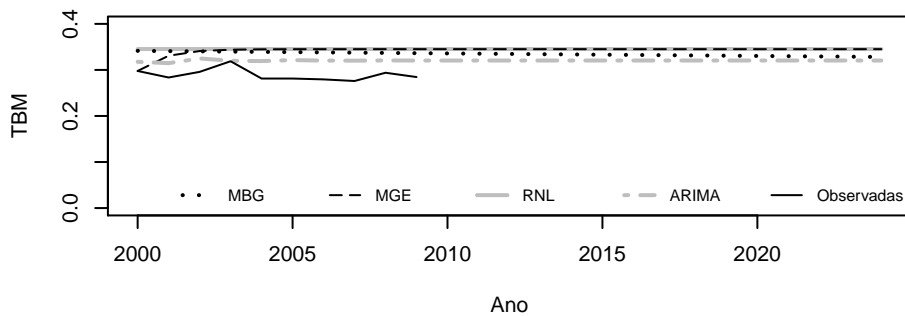
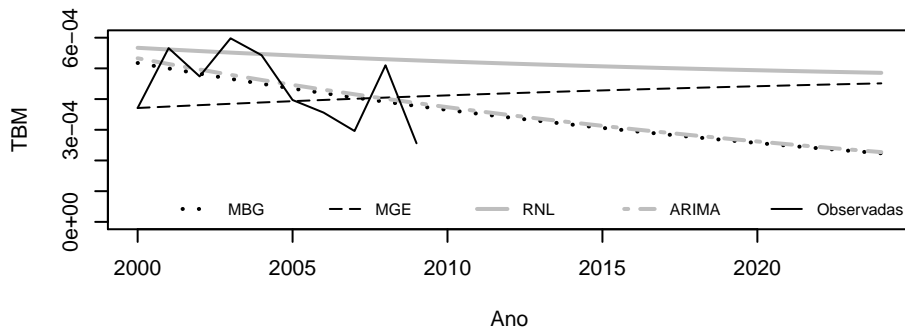
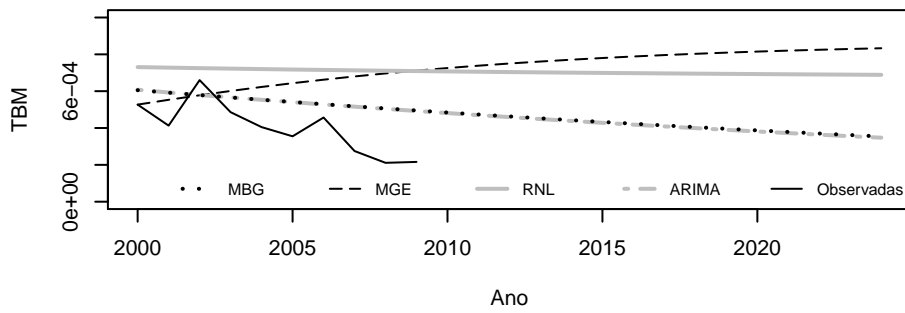
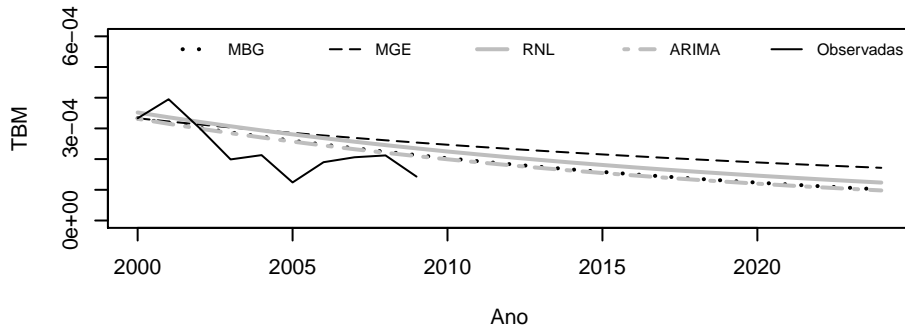


Figura 3.25: Comparação entre as previsões a LP (previsões a 25 anos, de 2000 a 2024) do modelo de RNL, do modelo ARIMA, do MBG e do MGE, para as idades (de cima para baixo) 4 e 15 do sexo masculino e 29 e 95 do sexo feminino

3.5 Previsões longitudinais em modelos de EDEs

No final deste capítulo, e apesar do trabalho exposto estar assente na análise transversal da mortalidade ao longo do tempo, não podemos deixar de mostrar outra das potencialidades da metodologia, de carácter prático e que consideramos útil, que tem a ver com o facto de facilmente podermos representar os resultados obtidos numa perspetiva de coorte. Para o efeito, como dispomos das previsões para todas as idades consecutivas para o mesmo horizonte temporal, fixamos uma certa idade, i , no início desse horizonte temporal, digamos t , e a respetiva previsão pontual obtida pelo modelo de EDEs. A previsão que se segue é obtida da previsão para a idade $i + 1$ para o ano $t + 1$ e assim sucessivamente (ver código no final do apêndice B).

Na figura 3.26, mostramos, a título de exemplo, as previsões da mortalidade para a coorte de indivíduos que nasceu no ano de 1999 no horizonte temporal de 2000 a 2009. Para este efeito, usamos os valores estimados através do MBG unidimensional (para grupos etários anuais, que temos vindo a designar por idade) que, muito embora seja um modelo simples, consideramos ilustrar convenientemente as previsões por coorte.

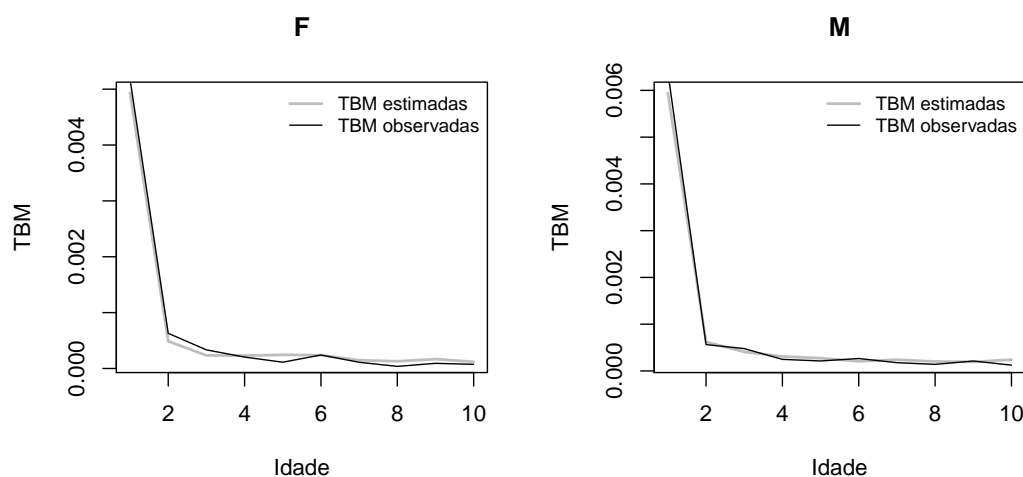


Figura 3.26: Representação das previsões a 10 anos do MBG para a coorte dos indivíduos que nasceu em 1999, por sexo (feminino, à esquerda e masculino à direita)

Para os modelos que propomos nos capítulos seguintes, o tipo de cálculo é semelhante ao agora apresentado. Se trabalharmos com grupos de idades diferentes dos anuais (quinquenais por exemplo) teríamos que avançar, de cada vez, em lugar de 1 ano de calendário, como agora fizemos, o número de anos correspondente à amplitude do respetivo grupo etário (no caso dos grupos etários quinquenais avançaríamos, de cada vez, 5 anos no tempo).

Da mesma forma que indicámos anteriormente (com recurso a simulação ou outro tipo de técnica que se mostre mais conveniente para o tipo de modelo) podemos também, neste caso, estimar uma aproximação dos intervalos de confiança para as previsões.

4

Modelos bivariados de equações diferenciais estocásticas para taxas de mortalidade

4.1	Introdução	58
4.2	Movimento browniano geométrico bidimensional com processos de Wiener correlacionados	59
4.3	Modelo de Gompertz estocástico bidimensional com processos de Wiener correlacionados	69
4.4	Comparação entre o movimento browniano geométrico bidimensional e um modelo de vetores autorregressivos e/ou de médias móveis	79

4.1 Introdução

Ao aplicar os modelos de EDEs a cada série temporal das TBMs da população portuguesa, por idade e por sexo (conforme ilustrado no capítulo 3), notámos que existem padrões semelhantes nos dados para a mesma idade de sexos diferentes, de idades diferentes do mesmo sexo e inclusivé de sexos diferentes. Estas semelhanças ocorrem quer em termos de tendência, quer em termos de variação anual. Este facto sugere que, ao longo do tempo, existem fatores ambientais que podem afetar de forma relacionada as TBMs dos indivíduos de ambos os sexos e/ou com idades diferentes.

A figura 4.1 ilustra a evolução do padrão da mortalidade em Portugal, concretamente através da representação da superfície do conjunto dos dados relativos a todas as idades do arco de vida, por sexo. Na figura, faz-se um cruzamento entre a abordagem transversal e a abordagem longitudinal ao longo do tempo (no período 1940 a 2009), conferindo a profundidade que as representações da figura 1.1 (que apresentámos no capítulo introdutório, para ilustrar a diferença entre aquelas duas abordagens) não permitia.

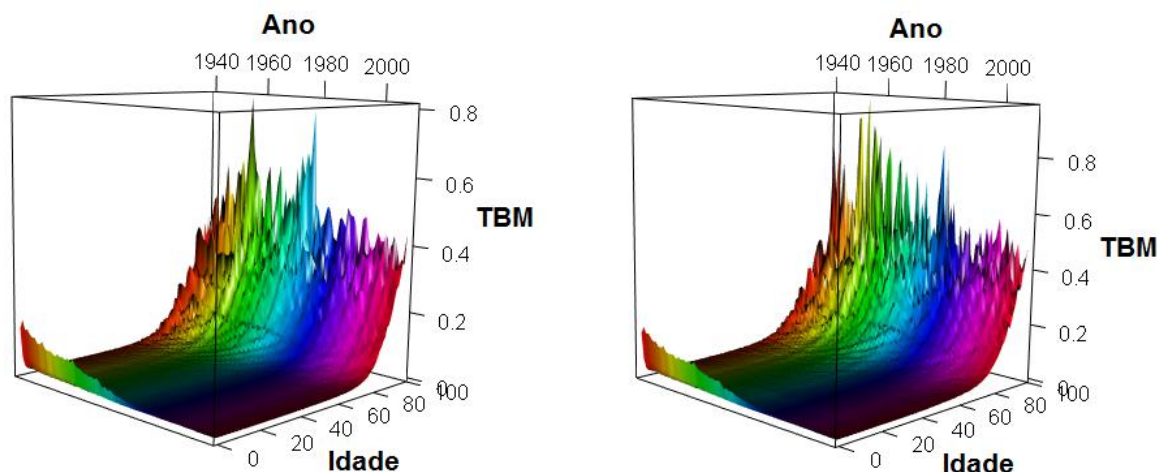


Figura 4.1: Gráfico de superfície das TBMs da população portuguesa, no período de 1940 a 2009, de todas as idades, por sexo (feminino à esquerda e masculino à direita)

Nos gráficos, é evidente o período de maior estabilidade na mortalidade a partir das primeiras 2 ou 3 idades e até, sensivelmente, à idade 60, comum ao longo do tempo em ambos os sexos. Ainda que neste ângulo seja pouco explícito, é ainda visível, no perfil lateral à esquerda (vemos como que uma dobra nas primeiras idades), a redução da mortalidade ao longo do tempo, comum a todas as idades e em ambos os sexos. Este comportamento da mortalidade com traços semelhantes para as mesmas idades de sexos diferentes sugere que os processos de Wiener, que estão associados aos respetivos coeficientes que explicam a variabilidade ambiental nos modelos de EDEs de cada sexo, devem estar correlacionados. Neste caso, não destacamos, mas é também notória a presença do mesmo padrão entre idades dentro de cada sexo e também entre idades diferentes de sexos diferentes, evidências em que assentam os pressupostos para a construção do modelo multivariado que apresentamos no capítulo 5.

Neste capítulo, propomos dois novos modelos bivariados de EDEs. Em ambos consideramos a existência de correlação entre sexos (fixando a idade), usando processos de Wiener correlacionados por sexo: na secção 4.2, apresentamos o modelo que designamos por movimento browniano geométrico bidimensional (MBGB) e, na secção 4.3, o modelo de Gompertz estocástico bidimensional (MGEB).

No capítulo 3, comparámos os resultados obtidos através dos modelos, no caso, unidimensionais de EDEs, entre si, e ainda com os obtidos via modelos de séries temporais e também de RNL. Neste capítulo, para, por um lado, validar os modelos e, por outro, aferir qual o modelo que melhor se ajusta aos dados ou com as melhores previsões, comparamos, analogamente, os modelos bidimensionais de EDEs entre si e ainda os resultados obtidos pelo MBGB com os obtidos por um modelo do tipo de vetores autorregressivos e de médias móveis (VARMA).

4.2 Movimento browniano geométrico bidimensional com processos de Wiener correlacionados

Na sequência dos resultados obtidos a partir do MBG, que apresentámos na secção 3.2, usamos como modelo base de evolução temporal das TBMs, para uma certa idade, o MBG e admitimos que indivíduos de ambos os sexos (fixando a idade) estão a ser afetados pelo mesmo tipo de aleatoriedade ambiental no que diz respeito à mortalidade.

Quando observamos os dados (veja-se, a título ilustrativo, as séries das TBMs representadas na figura 4.2, relativas às idades 2, 13, 17, 37, 72 e 83 de ambos os sexos), reparamos que existem picos de mortalidade descendentes e ascendentes (que correspondem, respetivamente, a anos de mortalidade reduzida e a anos de mortalidade elevada), que são comuns aos dois sexos, pelo que, em geral, há fortes indícios para indivíduos de sexos diferentes terem idêntico comportamento face ao fenómeno em estudo.

Esta correlação, observada praticamente em todas as idades, entre os efeitos dos fatores ambientais nas TBMs dos dois sexos, sugere que os $W(t)$ associados às EDEs que as modelam devem estar correlacionados, daí usarmos um modelo bidimensional.

No final desta secção, testamos o efeito da correlação entre sexos, comparando o modelo bidimensional, com correlação entre os $W(t)$ unidimensionais associados às TBMs de uma certa idade, do sexo feminino e masculino, com o modelo sem o efeito da correlação entre os dois sexos.

4.2.1 O MBGB

Consideremos a existência de uma variabilidade ambiental que influencia simultaneamente as TBMs dos dois sexos. Seja $X_j(t)$ a TBM de uma certa idade ($j = 1$ para o sexo feminino e $j = 2$ para o masculino), no ano t . Assumimos que $X_j(t)$ segue um MBG unidimensional. Logo, os logaritmos dos retornos, $Z_j(t) = \ln(X_j(t)/X_j(t-1))$, satisfazem

$$Z_j(t) = R_j + \sigma_j(W_k(t) - W_k(t-1))$$

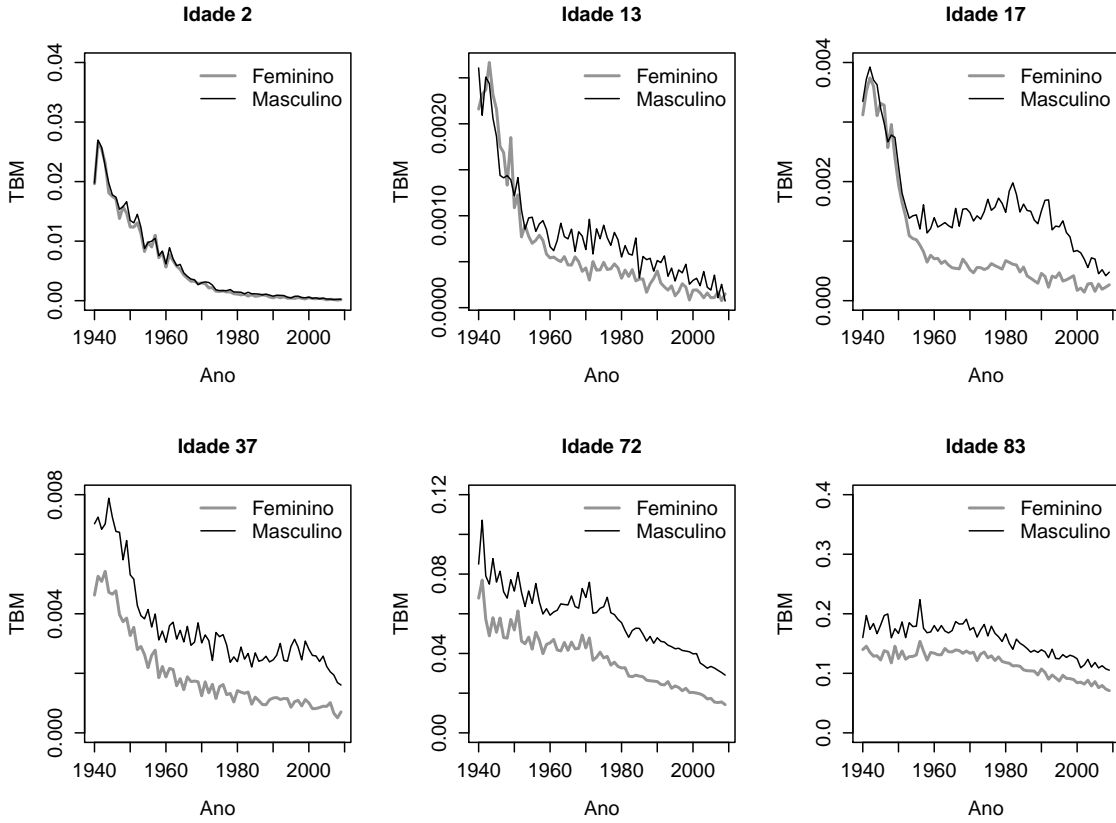


Figura 4.2: TBMs da população portuguesa, no período de 1940 a 2009, das idades 2, 13, 17 (em cima) e 37, 72 e 83 (em baixo) de ambos os sexos

com R_j , σ_j e $W_j(t)$ como R , σ e $W(t)$ em (3.4), e $Z(t) = [Z_1(t) Z_2(t)]'$ satisfaz

$$Z(t) = R + \Lambda^{1/2}(W(t) - W(t-1)), \quad (4.1)$$

com

$$R = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \quad \text{e} \quad W(t) = \begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix}.$$

Ao modelo descrito em (4.1) chamamos movimento browniano geométrico bidimensional (MBGB). Note-se que o vetor R representa as médias, R_j , dos $Z_j(t)$ e que, por sua vez, a matriz Λ é uma matriz diagonal com as variâncias dos $Z_j(t)$, σ_j^2 . $W(t)$ é o vetor dos processos de Wiener padrão, associados a cada um dos $Z_j(t)$, admitindo-se que $W_1(t)$ e $W_2(t)$ estejam correlacionados.

Seja ρ o coeficiente de correlação entre $W_1(t) - W_1(t-1)$ e $W_2(t) - W_2(t-1)$. Supomos que $|\rho| \neq 1$ pois, caso contrário, as TBMs, para cada idade, seriam funções determinísticas das TBMs do outro sexo. Usando dois processos de Wiener padrão independentes, $W_1^*(t)$ e $W_2^*(t)$, ambos normais com média nula e variância t , isto é, $W_j^*(t) \sim \mathcal{N}(0, t)$ i.i.d., podemos ter processos de Wiener correlacionados fazendo

$$W_1(t) = \alpha W_1^*(t) + \beta W_2^*(t) \quad (4.2)$$

e

$$W_2(t) = \beta W_1^*(t) + \alpha W_2^*(t) \quad (4.3)$$

com

$$\alpha = \left(\frac{1 + (1 - \rho^2)^{1/2}}{2} \right)^{1/2} \quad (4.4)$$

e

$$\beta = \text{sinal}(\rho) \left(\frac{1 - (1 - \rho^2)^{1/2}}{2} \right)^{1/2} = \text{sinal}(\rho)(1 - \alpha^2)^{1/2}. \quad (4.5)$$

Das expressões anteriores resulta que

$$E[W_j(t)^2] = t \quad (4.6)$$

e

$$E[W_1(t)W_2(t)] = \rho t. \quad (4.7)$$

Note-se ainda que, das operações e simplificações acima, resulta também que $\alpha^2 + \beta^2 = 1$ e que $\rho = 2\alpha\beta$ (o que facilmente se pode verificar usando diretamente as expressões (4.4) e (4.5)). Usando as propriedades de um processo de Wiener padrão, sabendo que $E[W_j^*(t)W_j^*(t-1)] = t-1$ e que $E[W^*(s)W^*(t)] = E[W^*(s)]E[W^*(t)] = 0 \times 0 = 0$, obtemos

$$E[(W_j(t) - W_j(t-1))^2] = 1 \quad (4.8)$$

e

$$E[(W_1(t) - W_1(t-1))(W_2(t) - W_2(t-1))] = \rho. \quad (4.9)$$

Assim, a expressão (4.6), que corresponde à variância de um $W_j(t)$, obtém-se, no caso de $W_1(t)$, por exemplo, usando (4.2) e através do desenvolvimento do quadrado do binómio no valor esperado $E[(\alpha W_1^*(t) + \beta W_2^*(t))^2]$. Como conhecemos a distribuição dos $W^*(t)$ independentes, sabemos que $E[W_j^*(t)] = 0$, para $j = 1, 2$ e podemos usar a propriedade $E[(W_j^*(t))^2] = \text{Var}[W_j^*(t)]$. Como $\text{Var}[W_j^*(t)] = t$, facilmente se chega à igualdade $E[W_j(t)^2] = (\alpha^2 + \beta^2)t$. O resultado final obtém-se substituindo, no membro do lado direito, α e β pelas expressões que lhes atribuímos, respetivamente, em (4.4) e (4.5) (neste caso, da igualdade à esquerda).

A expressão (4.7), que corresponde à covariância de $W_1(t)$ e $W_2(t)$, também se obtém diretamente do desenvolvimento de $E[W_1(t)W_2(t)]$, substituindo $W_1(t)$ e $W_2(t)$ tal como os definimos em (4.2) e (4.3). Novamente, usando o conhecimento da distribuição dos $W^*(t)$ independentes e as propriedades referidas a propósito da obtenção de (4.6), facilmente se chega à igualdade $E[W_1(t)W_2(t)] = 2\alpha\beta t$. Mais uma vez, o resultado final obtém-se substituindo agora, no membro do lado direito, α e β pelas expressões que vêm de (4.4) e (4.5) (neste caso, usamos a igualdade à direita).

Para além das propriedades anteriores, podemos ainda mostrar que $W_1(t)$ e $W_2(t)$ são também, enquanto combinações lineares de processos de Wiener (ver (4.2) e (4.3)), processos de Wiener. $W_1(t)$ verifica as propriedades que definem um processo de Wiener (para $W_2(t)$ a justificação é análoga), pois:

- $W_1(0) = \alpha W_1^*(0) + \beta W_2^*(0) = \alpha \times 0 + \beta \times 0 = 0$.

- Os incrementos $W_1(t) - W_1(s)$ ($s < t$) têm distribuição normal com média 0 e variância $t - s$. Com efeito, pela sua definição e pelas propriedades da soma de v.a. independentes e com distribuição normal, vem $W_1(t) - W_1(s) = \alpha W_1^*(t) + \beta W_2^*(t) - (\alpha W_1^*(s) + \beta W_2^*(s)) = \alpha(W_1^*(t) - W_1^*(s)) + \beta(W_2^*(t) - W_2^*(s))$, donde $W_1(t) - W_1(s) \sim \mathcal{N}(0, (\alpha^2 + \beta^2)(t - s)) = \mathcal{N}(0, t - s)$.
- $W_1(t)$ tem incrementos independentes. Também pela sua definição, para $s < t$, $W_1(t) - W_1(s) = \alpha(W_1^*(t) - W_1^*(s)) + \beta(W_2^*(t) - W_2^*(s))$, mas nem $W_1^*(t) - W_1^*(s)$ ou $W_2^*(t) - W_2^*(s)$ dependem da informação anterior de W_1^* ou W_2^* , e, por isso, o mesmo sucede a $W_1(t) - W_1(s)$. Logo, o incremento $W_1(t) - W_1(s)$ não depende da informação anterior de $W_1(t)$.

Sabemos que $Z(t)$ tem uma distribuição normal bivariada (veja-se, por exemplo, [84]), com média \mathbf{R} e variância Σ , isto é, $Z(t) \sim \mathcal{N}(\mathbf{R}, \Sigma)$ i.i.d. com $\Sigma = \Lambda^{1/2} \mathbf{S} \Lambda^{1/2}$ e \mathbf{S} a matriz de correlações de $W(t) - W(t - 1)$, dada por

$$\mathbf{S} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

em que

$$\rho = \text{Corr}[W_1(t) - W_1(t - 1), W_2(t) - W_2(t - 1)]. \quad (4.10)$$

Note-se que $E[W_1(t)W_2(t)] = \rho t$. Vamos supor que $\rho \neq 1$ pois, caso contrário, os logaritmos dos retornos das TBMs de um sexo seriam funções lineares dos logaritmos dos retornos das TBMs do outro sexo.

A matriz Σ é dada por

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}.$$

A distribuição conjunta de $Z_1(t)$ e $Z_2(t)$ é normal bivariada com coeficiente de correlação, $r(t_0, t)$, dado por

$$r(t_0, t) = \frac{\text{Cov}[Z_1(t), Z_2(t)]}{s_{Z_1(t)}s_{Z_2(t)}} = \frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2} = \rho.$$

4.2.2 Estimação

Sejam $t_n = t_0 + n$ ($n = 0, 1, 2, \dots, N$) os anos em que as TBMs foram observadas, para cada idade e por sexo, e seja $Z_j(t_n) = \ln(X_j(t_n)/X_j(t_{n-1}))$. Dada a independência entre os $Z(t_n) = [z_1(t_n) \ z_2(t_n)]'$ para diferentes valores de n , a f.d.p. conjunta de transição de $(Z_1(t), Z_2(t))$ entre t_{n-1} e t_n , que é normal e não depende de n , pode escrever-se como

$$\begin{aligned} f(z_{1,n}, z_{2,n} | Z_1(t_{n-1}) = z_{1,n-1}, Z_2(t_{n-1}) = z_{2,n-1}) &= f(z_{1,n}, z_{2,n}) \\ &= \frac{1}{2\pi s_{Z_{1,n}} s_{Z_{2,n}} \sqrt{1 - r_{n|n-1}^2}} \exp \left\{ -\frac{1}{2} Q(z_{1,n}, z_{2,n}) \right\} \end{aligned}$$

com

$$\begin{aligned} Q(z_{1,n}, z_{2,n}) &= \frac{1}{1 - r_{n|n-1}^2} \left[\left(\frac{z_{1,n} - \mu_{Z_{1,n}|n-1}}{s_{Z_{1,n}|n-1}} \right)^2 - 2r_{n|n-1} \frac{(z_{1,n} - \mu_{Z_{1,n}|n-1})(z_{2,n} - \mu_{Z_{2,n}|n-1})}{s_{Z_{1,n}|n-1} s_{Z_{2,n}|n-1}} + \right. \\ &\quad \left. \left(\frac{z_{2,n} - \mu_{Z_{2,n}|n-1}}{s_{Z_{2,n}|n-1}} \right)^2 \right], \end{aligned}$$

em que

$$\mu_{Z_j, n|n-1} = R_j,$$

$$s^2_{Z_j, n|n-1} = \sigma_j^2$$

e

$$r_{n|n-1} = \rho.$$

Isto é,

$$f(z_{1,n}, z_{2,n}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}Q(z_1, z_2)\right\},$$

com

$$Q(z_1, z_2) = \frac{1}{1-\rho^2} \left[\left(\frac{z_1 - R_1}{\sigma_1}\right)^2 - 2\rho \frac{(z_1 - R_1)(z_2 - R_2)}{\sigma_1\sigma_2} + \left(\frac{z_2 - R_2}{\sigma_2}\right)^2 \right].$$

Sejam $z_{1,n}$ e $z_{2,n}$ ($n = 1, \dots, N$), os valores observados de $Z_1(t_n)$ e de $Z_2(t_n)$, e seja $\mathbf{p} = (R_1, R_2, \sigma_1, \sigma_2, \rho)$ o vetor dos parâmetros. Usando a propriedade de Markov em (4.1), a função de log-verossimilhança, \mathcal{L} , dos logaritmos dos retornos das TBMs, para uma certa idade, é dada por

$$\mathcal{L}(\mathbf{p}|Z_1(t_1), Z_2(t_1) \dots Z_1(t_N), Z_2(t_N)) = \sum_{n=1}^N \ln(f(Z_1(t_n), Z_2(t_n))). \quad (4.11)$$

As estimativas dos parâmetros, para cada idade, foram obtidas por métodos numéricos através da minimização do simétrico da função \mathcal{L} em (4.11), utilizando a função `optim` do R. Para a estimação, usa-se o código no apêndice E, com as simplificações resultantes da escrita do corrente modelo, o MBGB e, recordamos, a simplificação $t_n - t_{n-1} = 1$.

Para obter os intervalos de confiança para os parâmetros, assumimos que estamos em regime assintótico, considerando as propriedades de estimação por MV, e fazemos uma aproximação da *matriz de informação de Fisher* pela simétrica da inversa da *matriz hessiana*, de cuja diagonal obtemos as respetivas variâncias assintóticas (esta técnica foi aplicada, tal como descrito no capítulo anterior, para obter os *IC* assintóticos dos parâmetros estimados pelos modelos unidimensionais de EDEs). Para cada parâmetro, obtemos assim uma aproximação dos limites de um intervalo de confiança a $(1 - \alpha) \times 100\%$ através de

$$\hat{p} \pm z_{1-\alpha/2}(\widehat{Var}[\hat{p}])^{1/2},$$

com $\widehat{Var}[\hat{p}]$ uma estimativa da variância obtida como acima descrito.

Se tivermos observações até um certo instante t_N e quisermos obter previsões para $t > t_N$, considerando que $Z_j(t)$ é um processo de Markov, temos que

$$E[Z_j(t)|Z_1(t_1), Z_2(t_1), \dots, Z_1(t_N), Z_2(t_N)] = E[Z_j(t)|Z_1(t_N), Z_2(t_N)].$$

Como

$$Z_j(t)|Z_1(t_N) = z_{1,N}, Z_2(t_N) = z_{2,N} \sim \mathcal{N}(R_j, \sigma_j^2(t - t_N)),$$

podemos usar para previsões a LP, para cada idade,

$$\widehat{Z}_j(t) = \widehat{E}[Z_j(t)|Z_1(t_N) = z_{1,N}, Z_2(t_N) = z_{2,N}] = \widehat{R}_j, \quad (4.12)$$

onde \hat{E} representa o valor aproximado da esperança matemática, substituindo os valores exatos de R_j pelos dos seus estimadores de MV, isto é, \hat{R}_j . Para obtermos as previsões das TBMs, vem

$$\hat{X}_j(t) = X_j(t-1) \exp\{\hat{R}_j(t-t_N)\}.$$

As previsões PP são estimadas da mesma forma que em (4.12), mas atualizando t e os valores observados, bem como as estimativas dos parâmetros de cada vez (no nosso caso anualmente) que avançamos um passo no tempo.

4.2.3 Resultados

Tal como no MBG unidimensional, cujos resultados apresentámos em 3.2.3, ajustámos o MBGB aos dados da mortalidade da população portuguesa, para as idades 0 aos 99 anos, neste caso conjuntamente para ambos os sexos. Na figura 4.3, representamos os parâmetros estimados para todas as idades. No que se refere às formas descritas nos gráficos dos valores estimados de R , σ e ρ , note-se que, para R e σ , estas são diferentes das obtidas para o modelo unidimensional (ver figuras 3.1 e 3.2). No caso do parâmetro R (que traduz a tendência, em escala logarítmica, de uma

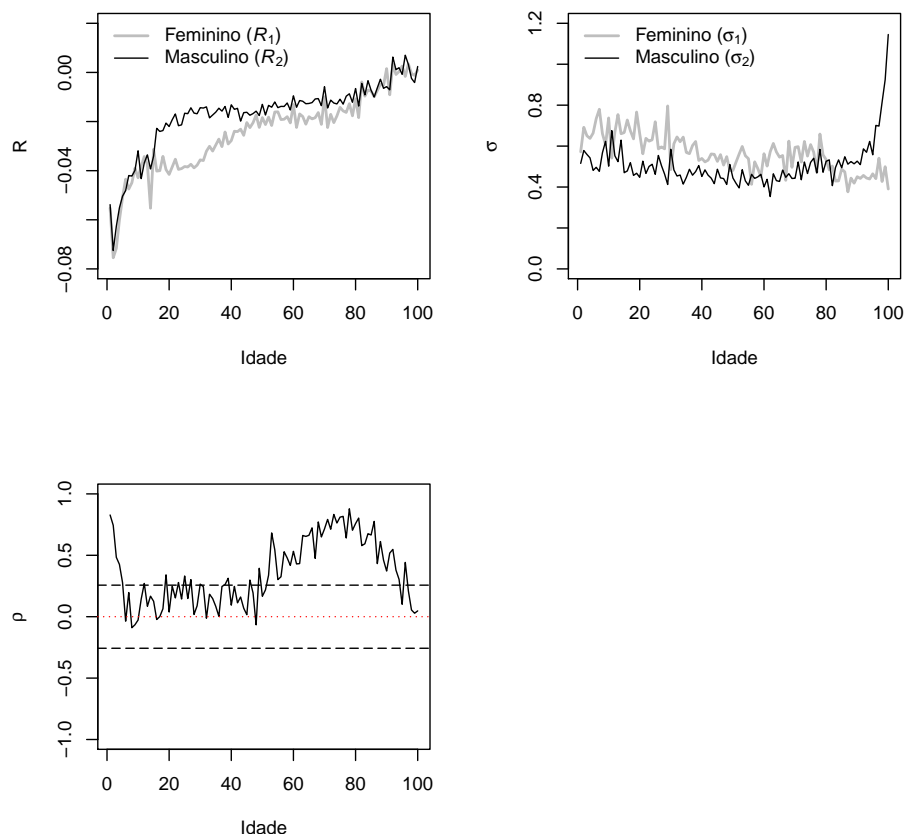


Figura 4.3: Estimativas dos parâmetros do MBGB, por idade e para ambos os sexos. No último gráfico, indicamos a região de aceitação (a tracejado) da hipótese $\rho = 0$ para o nível de significância de 5% (para o efeito usámos a transformação de Fisher como sugerido, por exemplo, em [82])

taxa média de mortalidade) essas diferenças ocorrem sobretudo nas primeiras idades, pois os valores, para ambos os sexos, sofrem um aumento acentuado até sensivelmente à idade 20 e posteriormente ocorre um padrão diferenciado por sexo até à idade 40. Depois da idade 80 os valores estimados de R aumentam ligeiramente em ambos os sexos. Relativamente a σ (recordamos que no modelo unidimensional não estimámos σ , mas sim σ^2 , a que chamámos V e que traduz o efeito das flutuações ambientais sobre a mortalidade), até à idade 60 observa-se uma ligeira tendência decrescente nos valores estimados, para ambos os sexos, que depois se inverte sensivelmente até à idade 80. Até essa mesma idade os valores mais elevados estão associados ao sexo feminino, mas a partir daí, aumentam abruptamente as “intensidades das flutuações” sobre a mortalidade do sexo masculino.

Quanto a ρ , os seus valores são mais elevados nas primeiras idades, sendo que, relativamente às restantes idades adultas (e também às idades jovens), ocorre um aumento progressivo do valor deste parâmetro sensivelmente a partir da idade 50 e até à 80 (em que ocorre um valor de correlação próximo de 0.8). A partir dessa idade, inverte-se a tendência crescente, de tal modo que 0.1 é sensivelmente o valor estimado para a idade 99.

Para testar se as correlações são significativamente diferentes de zero, usámos a transformação de Fisher como sugerido, por exemplo, em [82]. No gráfico, indicamos a região de aceitação na escala original (a tracejado) da hipótese $\rho = 0$ para o nível de significância de 5%. Os resultados apontam para correlações significativas nas primeiras idades, em algumas das idades adultas e para quase todas as idades depois da idade 50.

Ainda relativamente aos parâmetros, estimámos os intervalos de confiança assintóticos de modo análogo ao que mostrámos na secção 3.2 para o modelo unidimensional. Note-se que no caso bidimensional, no entanto, as margens de erro são mais amplas, sendo que as associadas a R_1 e R_2 são mesmo, para todas as idades, superiores aos valores estimados dos parâmetros. Na tabela 4.1, apresentamos os valores agora obtidos para as primeiras três idades e para uma confiança de 95%.

Tabela 4.1: Estimativas dos parâmetros do MBGB com semi-amplitude dos $IC_{95\%}$ assintóticos para as idades 0, 1 e 2

Idade	$R_1(ano^{-1})$	$R_2(ano^{-1})$	$\sigma_1(ano^{-1/2})$	$\sigma_2(ano^{-1/2})$	ρ
0	-0.0552 (± 0.0745)	-0.0538 (± 0.0671)	0.5722 (± 0.2602)	0.5155 (± 0.2344)	0.8282 (± 0.2473)
1	-0.0755 (± 0.0901)	-0.0725 (± 0.0753)	0.6919 (± 0.1945)	0.578 (± 0.1626)	0.7442 (± 0.2172)
2	-0.0716 (± 0.0847)	-0.0626 (± 0.0728)	0.6506 (± 0.1535)	0.5595 (± 0.1320)	0.4835 (± 0.3131)

Na figura 4.4, apresentamos uma concretização do ajustamento do modelo e previsões PP e a LP, neste caso para a idade 1 do sexo feminino. Em geral, consideramos que o MBGB nos parece ser adequado para este tipo de dados, apresentando ajustamentos e previsões que consideramos satisfatórias. No entanto, embora tenha sido possível aplicar o modelo, a cada uma das idades, em

simultâneo para ambos os sexos, os resultados sugerem diferenças, quer no ajustamento, quer nas previsões, quanto ao caso unidimensional, que pretendemos evidenciar de seguida.

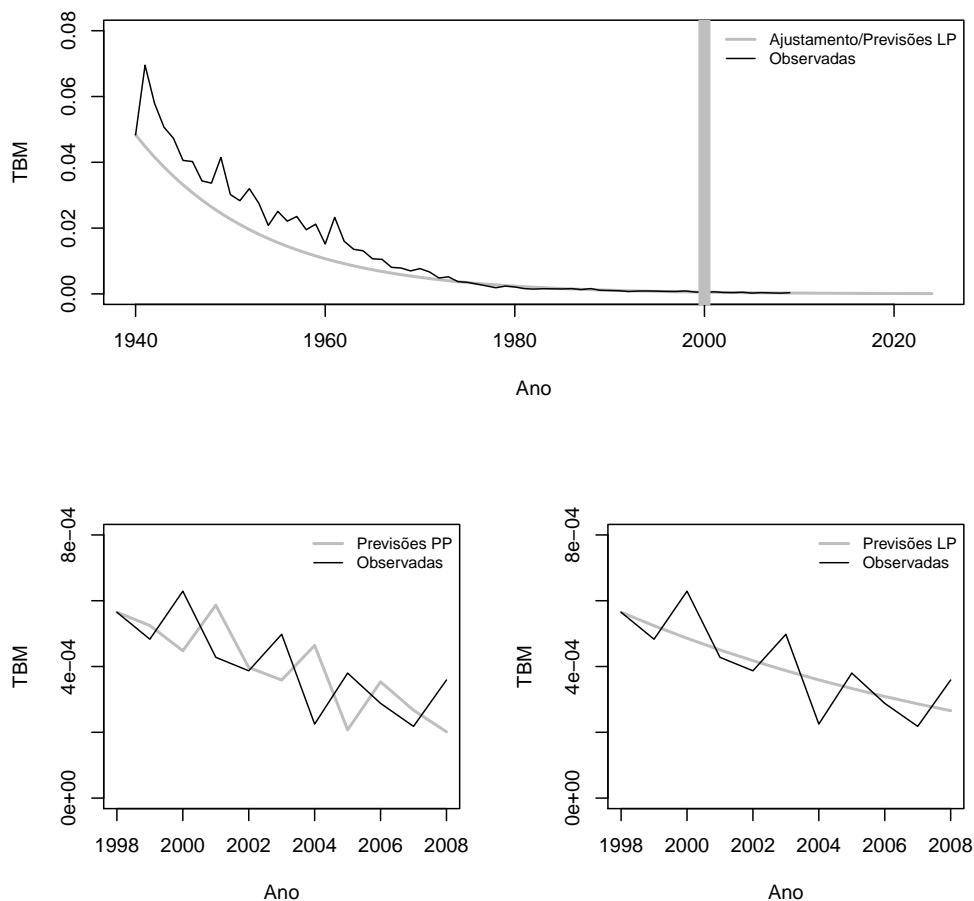


Figura 4.4: Ajustamento do MBGB com previsões a LP (25 anos: de 2000 a 2024) para a idade 1 do sexo feminino (em cima); ampliação das previsões (PP e a LP, no período de 2000 a 2009, respetivamente, à esquerda e à direita, em baixo)

Começamos por comparar o modelo com o efeito das correlações por sexo (MBGB) com o modelo sem a correlação (MBG). Para o efeito, podemos aplicar um teste de razão de verosimilhanças com

$$H_0 : \rho = 0 \quad \text{vs} \quad H_1 : \rho \neq 0$$

e usando a estatística de teste

$$D = -2 \ln \left(\frac{\text{Verosimilhança do modelo com } \rho = 0}{\text{Verosimilhança do modelo com } \rho \neq 0} \right),$$

que, para a hipótese nula, tem uma distribuição aproximadamente qui-quadrado com um grau de liberdade.

Na tabela 4.2, apresentamos os valores estimados da estatística de teste D para todas as idades. Para um nível de significância $\leq 5\%$, nas idades entre os 51 e os 93 anos, rejeitamos a hipótese nula, ou seja, o parâmetro ρ é significativo para o modelo.

Não só nas idades referidas, mas também logo nas primeiras quatro idades e também em outras idades dispersas fora desse intervalo, os resultados sugerem que o modelo agora proposto, com correlações entre os processos de Wiener que descrevem o efeito das flutuações ambientais sobre cada idade de ambos os sexos, apresenta diferenças significativas face ao modelo sem o efeito das correlações. Pelo exposto, somos levados a concluir que há vantagens na modelação conjunta das TBMs de ambos os sexos, através do MBGB.

Tabela 4.2: Resultados do teste de razão de verosimilhanças entre o MBGB e o MBG para todas as idades (* representa valores-p < 0.05; ** representa valores-p < 0.01)

Idade	D	Idade	D	Idade	D	Idade	D
0	68.29165**	25	1.33425	50	2.87724	75	64.89495**
1	47.65775**	26	5.64588*	51	7.19543**	76	31.21364**
2	15.70391**	27	0.01963	52	36.92957**	77	87.37983**
3	11.79149**	28	0.46251	53	20.47175**	78	40.39463**
4	4.73077*	29	4.29113*	54	5.73761*	79	50.51581**
5	0.07621	30	3.64455	55	6.62437*	80	61.15839**
6	2.32805	31	0.00929	56	19.41620**	81	24.51149**
7	0.47799	32	2.01501	57	15.25178**	82	25.57473**
8	0.25485	33	1.31600	58	11.33184**	83	35.94525**
9	0.04126	34	0.42279	59	19.67510**	84	34.55056**
10	1.20288	35	0.00099	60	12.03278**	85	54.26652**
11	4.50663*	36	3.68435	61	12.10104**	86	12.44581**
12	0.42442	37	4.22576*	62	33.92692**	87	27.74834**
13	1.67499	38	6.04685*	63	32.86903**	88	14.04769**
14	0.93537	39	0.52305	64	34.14643**	89	8.77142**
15	0.02915	40	3.71503	65	43.85197**	90	18.42994**
16	0.00007	41	0.76560	66	14.86694**	91	21.78611**
17	0.243831	42	1.53293	67	53.36631**	92	9.18828**
18	7.27383**	43	0.27895	68	32.32503**	93	5.69948**
19	0.08917	44	0.01543	69	42.15327**	94	0.60970
20	3.73939	45	5.52891*	70	58.30814**	95	12.81615**
21	1.39773	46	2.25050	71	41.72734**	96	2.66071
22	4.76288**	47	0.25036	72	69.77974**	97	0.18274
23	1.23322	48	9.91717**	73	51.38017**	98	0.04889
24	6.89081**	49	1.68935	74	63.34674**	99	0.13433

Para além dos resultados apresentados na tabela 4.2, podemos comparar ainda as versões uni e bidimensionais do MBG concretamente através dos EQMs estimados nos ajustamentos e previsões obtidos pelas duas vias. Assim, nas figuras 4.5 e 4.6, representamos a diferença entre o EQM associado ao MBG e o MBGB, isto é, $EQM_{MBG} - EQM_{MBGB}$, respetivamente, para os ajustamentos e para as previsões a LP.

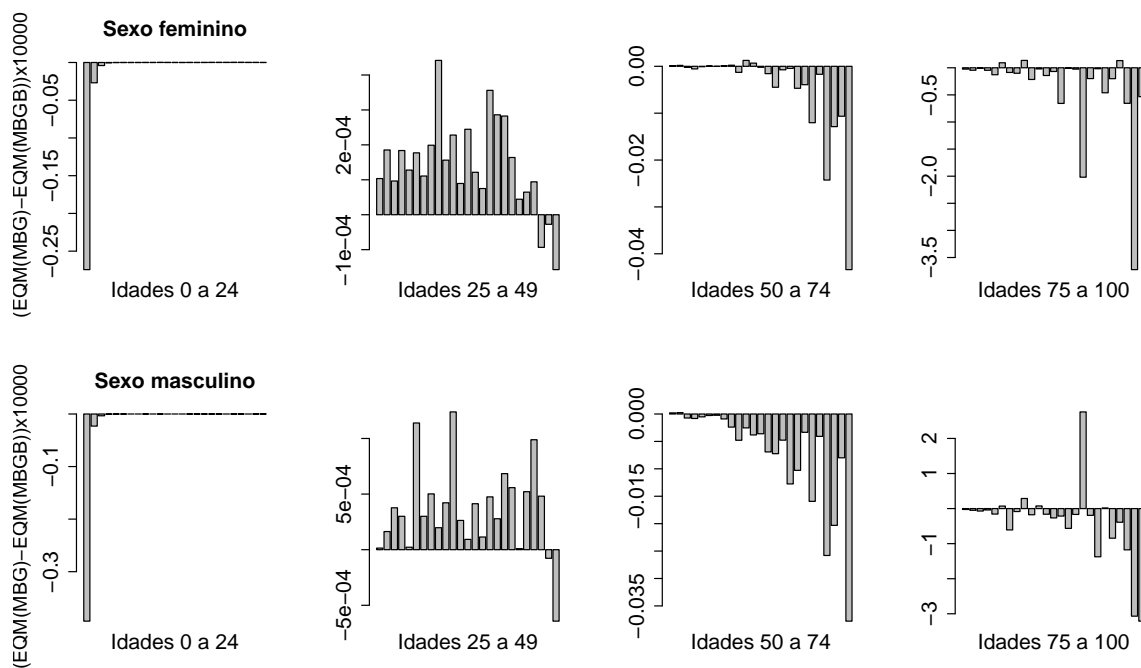


Figura 4.5: Diferença ($\times 10000$) entre os EQMs associados ao ajustamento às TBMs do MBG e do MBGB, por idade; sexo feminino, em cima, e masculino, em baixo

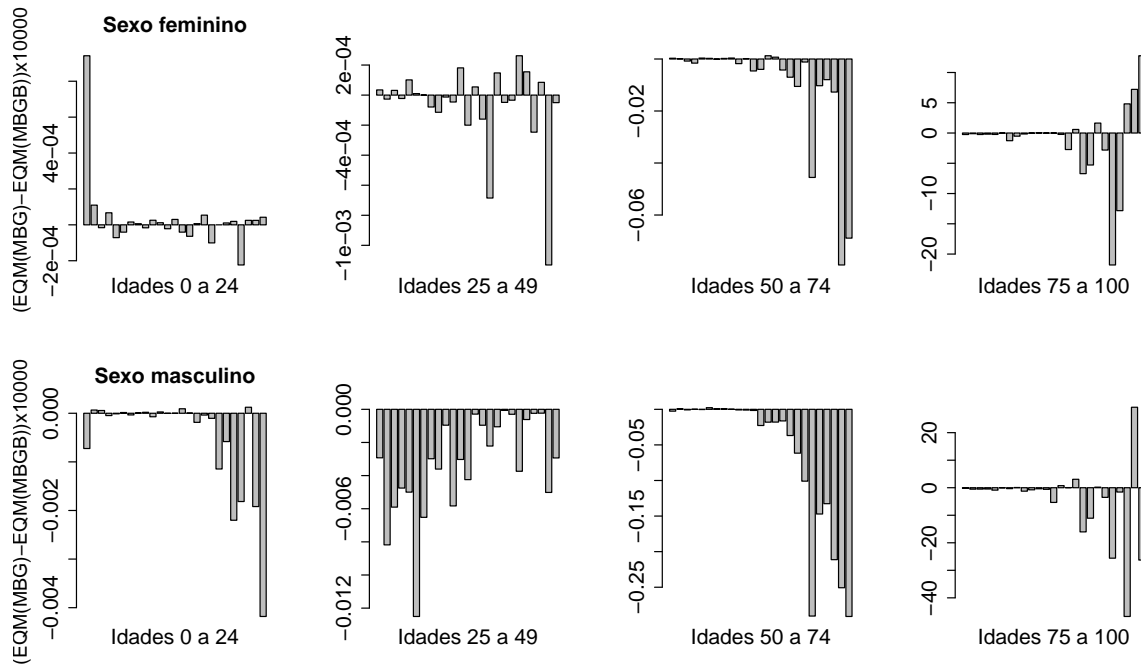


Figura 4.6: Diferença ($\times 10000$) entre os EQMs associados às previsões a LP das TBMs (de 2000 a 2009) do MBG e do MBGB, por idade; sexo feminino, em cima, e masculino, em baixo

Note-se que, tal como nas representação análogas do capítulo 3, nas figuras anteriores, dado o pequeno valor das diferenças entre os EQMs, alterámos a escala dos resultados representados, multiplicando os valores estimados para essa diferença por 10000. Se analisarmos, quer para os ajustamentos, quer para as previsões, a diferença entre os respetivos EQMs, para cada idade e por sexo, o MBG apresenta, globalmente, vantagens face ao MBGB, embora ligeira.

No caso do ajustamento (com exceção das primeiras 2 idades e das idades 25 a 49, em ambos os sexos), a diferença entre os erros aumenta com a idade. Adicionalmente, nas idades 25 a 49 o padrão da diferença é também contrário à tendência geral e claramente favorável ao modelo bidimensional. Essa vantagem é ainda expressiva na idade 91 do sexo masculino.

Já para as previsões, não se verifica uma tendência geral, quer por sexo, quer por idade, em termos da prestação dos dois modelos, se atendermos ao critério do EQM. Ao contrário dos resultados para o ajustamento, não há aqui uma vantagem clara do modelo unidimensional em todos os grupos de idades representados (note-se, desde logo, o resultado da diferença entre os erros na primeira idade, que sugere um melhor desempenho do modelo bidimensional para o sexo feminino e do modelo unidimensional para o sexo masculino). De facto, o único gráfico que ilustra, de modo mais evidente, uma vantagem em termos de prestação para o MBG diz respeito às idades 25 a 49 do sexo masculino (o gráfico sugere também que a diferença entre os EQMs associados aos dois modelos tem tendência decrescente com a idade). Para as idades entre a idade 25 e 49, mas agora do sexo feminino, as diferenças entre os erros alternam entre valores positivos e negativos (tal como já havia acontecido no grupo de idades anterior, para o mesmo sexo), o que não contribui para destacar nenhum dos modelos. Note-se que partir da idade 50, a diferença entre os erros aumenta bastante (com maior dimensão entre as idades 75 a 100 do sexo masculino), no entanto o padrão é semelhante por sexo (inclusive o melhor desempenho do modelo bidimensional nas últimas idades).

Considerando os resultados obtidos nesta secção, sobretudo do teste de razão de verosimilhanças, faz sentido continuar a apostar na modelação através de modelos de EDEs multidimensionais com recurso a processos de Wiener correlacionados (conforme já referimos, devido à informação obtida a partir da análise dos dados). Assim, baseando-nos, desta vez, no MGE, na secção seguinte, propomos um outro modelo bidimensional de EDEs, cujos resultados iremos comparar com os agora apresentados.

4.3 Modelo de Gompertz estocástico bidimensional com processos de Wiener correlacionados

À semelhança do exposto na introdução da secção anterior, para justificar a construção de um modelo bidimensional, recordamos que, quando observamos as TBMs de ambos os sexos para uma certa idade, identificamos uma correlação em todas as idades. Consideramos que se trata de correlações entre os efeitos dos fatores ambientais na mortalidade dos dois sexos, o que sugere que os processos de Wiener associados às EDEs que modelam as taxas estejam correlacionados.

Admitindo que grupos de indivíduos de ambos os sexos (fixando a idade) podem ser afetados pelo mesmo tipo de aleatoriedade ambiental no que diz respeito à mortalidade, usamos, neste caso, como modelo base para descrever a evolução temporal das TBMs de uma certa idade e para ambos os sexos, o MGE, apresentado na secção 3.3.

No final da secção, testamos também o efeito da correlação entre sexos, comparando o modelo bidimensional, com correlação entre os $W(t)$ unidimensionais associados aos modelos para as TBMs de uma certa idade, do sexo feminino e masculino, contra o modelo sem o efeito da correlação entre os dois sexos, tal como fizemos para o MBGB.

4.3.1 O MGE

À semelhança da secção anterior, consideremos a existência de uma variabilidade ambiental que influencia simultaneamente as TBMs dos dois sexos, para uma certa idade. Considerando ainda que as TBMs seguem o MGE, associamos, para cada idade, os modelos de ambos os sexos, obtendo o sistema de EDEs bidimensional

$$\begin{cases} dY_1(t) = b_1(A_1 - Y_1(t))dt + \sigma_1 dW_1(t) \\ dY_2(t) = b_2(A_2 - Y_2(t))dt + \sigma_2 dW_2(t), \end{cases} \quad (4.13)$$

com $Y_j(t) = \ln X_j(t)$ (recordamos $j = 1$ para o sexo feminino e $j = 2$ para o masculino), $Y_j(t_0) = y_{j,t_0}$ suposto conhecido, $X_j(t)$ as TBMs dessa idade, no ano t , $W_j(t)$ os processos de Wiener correlacionados e os parâmetros A_j (taxas médias de mortalidade assintóticas, em escala logarítmica), b_j (velocidades de aproximação ao regime assintótico) e σ_j (intensidades das flutuações ambientais). Ao modelo descrito pelo sistema de equações (4.13) chamamos modelo de Gompertz estocástico bidimensional (MGE).

Neste ponto refira-se que são válidas para esta secção, todas as propriedades (que mostrámos na secção anterior) relativas aos processos de Wiener correlacionados, $W_1(t)$ e $W_2(t)$, com correlação ρ entre os incrementos $W_1(t) - W_1(t-1)$ e $W_2(t) - W_2(t-1)$.

Dada a condição inicial $Y_1(t_0) = y_{1,t_0}$ e $Y_2(t_0) = y_{2,t_0}$, a solução de (4.13) é

$$Y_1(t) = A_1 + (y_{1,t_0} - A_1) \exp\{-b_1(t - t_0)\} + \sigma_1 \exp\{-b_1 t\} \int_{t_0}^t \exp\{b_1 s\} dW_1(s) \quad (4.14)$$

e

$$Y_2(t) = A_2 + (y_{2,t_0} - A_2) \exp\{-b_2(t - t_0)\} + \sigma_2 \exp\{-b_2 t\} \int_{t_0}^t \exp\{b_2 s\} dW_2(s), \quad (4.15)$$

Em [9], pode ver-se que um integral estocástico com função integranda determinística G tem distribuição normal com média 0 e variância $E[(\int_{t_0}^t G(s) dW(s))^2] = \int_{t_0}^t E[G^2(s)] ds = \int_{t_0}^t G^2(s) ds$.

Uma vez que o integral estocástico com função integranda determinística, tem distribuição normal, como $W_1(t)$ e $W_2(t)$ são processos de Wiener, as distribuições marginais normais têm médias

$$\mu_{Y_j(t)} = A_j + (y_{j,t_0} - A_j) \exp\{-b_j(t - t_0)\}$$

e variâncias

$$\sigma_{Y_j(t)}^2 = \sigma_j^2(1 - \exp\{-2b_j(t - t_0)\}/(2b_j)),$$

isto é,

$$Y_1(t) \sim \mathcal{N}(\mu_{Y_1(t)}, \sigma_{Y_1(t)}^2) \quad (4.16)$$

e

$$Y_2(t) \sim \mathcal{N}(\mu_{Y_2(t)}, \sigma_{Y_2(t)}^2). \quad (4.17)$$

A distribuição conjunta é normal bivariada com coeficiente de correlação, $r(t_0, t)$, que se obtém a partir da expressão

$$r(t_0, t) = \frac{Cov[Y_1(t), Y_2(t)]}{s_{Y_1(t)}s_{Y_2(t)}}. \quad (4.18)$$

A covariância entre $Y_1(t)$ e $Y_2(t)$ é

$$\begin{aligned} Cov[Y_1(t), Y_1(t)] &= \\ Cov \left[A_1 + (y_{1,t_0} - A_1) \exp\{-b_1(t - t_0)\} + \sigma_1 \exp\{-b_1 t\} \int_{t_0}^t \exp\{b_1 s\} dW_1(s), \right. \\ &\quad \left. A_2 + (y_{2,t_0} - A_2) \exp\{-b_2(t - t_0)\} + \sigma_2 \exp\{-b_2 t\} \int_{t_0}^t \exp\{b_2 s\} dW_2(s) \right] \\ &= E \left[\sigma_1 \exp\{-b_1 t\} \int_{t_0}^t \exp\{b_1 s\} dW_1(s) \sigma_2 \exp\{-b_2 t\} \int_{t_0}^t \exp\{b_2 s\} dW_2(s) \right] \\ &= \sigma_1 \sigma_2 \exp\{-(b_1 + b_2)t\} E \left[\left(\int_{t_0}^t \exp\{b_1 s\} \alpha dW_1^*(s) + \int_{t_0}^t \exp\{b_1 s\} \beta dW_2^*(s) \right) \right. \\ &\quad \left. \left(\int_{t_0}^t \exp\{b_2 s\} \beta dW_1^*(s) + \int_{t_0}^t \exp\{b_2 s\} \alpha dW_2^*(s) \right) \right] \\ &= \sigma_1 \sigma_2 \exp\{-(b_1 + b_2)t\} E \left[\alpha \beta \int_{t_0}^t \exp\{b_1 s\} dW_1^*(s) \int_{t_0}^t \exp\{b_2 s\} dW_1^*(s) + \right. \\ &\quad \left. \alpha \beta \int_{t_0}^t \exp\{b_1 s\} dW_2^*(s) \int_{t_0}^t \exp\{b_2 s\} dW_2^*(s) \right]. \end{aligned}$$

Como para as funções determinísticas G_1 e G_2 temos

$$E \left[\int_{t_0}^t G_1(s) dW(s) \int_{t_0}^t G_2(s) dW(s) \right] = \int_{t_0}^t E[G_1(s)G_2(s)] ds = \int_{t_0}^t G_1(s)G_2(s) ds,$$

vem

$$\begin{aligned} Cov[Y_1(t), Y_1(t)] &= 2\alpha\beta\sigma_1\sigma_2 \exp\{-(b_1 + b_2)t\} \int_{t_0}^t \exp\{(b_1 + b_2)s\} ds \\ &= 2\alpha\beta\sigma_1\sigma_2 \frac{1}{(b_1 + b_2)} (1 - \exp\{-(b_1 + b_2)(t - t_0)\}). \end{aligned}$$

Então, substituindo a expressão agora obtida para a covariância em (4.18), obtemos

$$r(t_0, t) = \frac{2\alpha\beta\sigma_1\sigma_2(b_1 + b_2)^{-1}(1 - \exp\{-(b_1 + b_2)(t - t_0)\})}{\sigma_1(1 - \exp\{-2b_1(t - t_0)\}/(2b_1))^{1/2}\sigma_2(1 - \exp\{-2b_2(t - t_0)\}/(2b_2))^{1/2}}.$$

Finalmente, simplificamos a expressão e, como sabemos que $\rho = 2\alpha\beta$, resulta que

$$r(t_0, t) = \rho \frac{1 - \exp\{-(b_1 + b_2)(t - t_0)\}}{(1 - \exp\{-2b_1(t - t_0)\})^{1/2}(1 - \exp\{-2b_2(t - t_0)\})^{1/2}} \frac{2(b_1 b_2)^{1/2}}{b_1 + b_2}. \quad (4.19)$$

4.3.2 Estimação

Consideremos agora $t_n = t_0 + n$ ($n = 0, 1, 2, \dots, N$) os anos em que as TBMs foram observadas, para cada idade e por sexo, e seja $Y_j(t_n) = \ln(X_j(t_n))$. Podemos escrever a f.d.p. conjunta de transição de $(Y_1(t), Y_2(t))$ entre t_{n-1} e t_n , que é normal, como

$$\begin{aligned} f(y_{1,n}, y_{2,n} | Y_1(t_{n-1}) = y_{1,n-1}, Y_2(t_{n-1}) = y_{2,n-1}) \\ = \frac{1}{2\pi s_{Y_{1,n}} s_{Y_{2,n}} \sqrt{1 - r_{n|n-1}^2}} \exp\left\{-\frac{1}{2} Q(y_{1,n}, y_{2,n} | y_{1,n-1}, y_{2,n-1})\right\} \end{aligned}$$

com

$$\begin{aligned} Q(y_{1,n}, y_{2,n} | y_{1,n-1}, y_{2,n-1}) = \frac{1}{1 - r_{n|n-1}^2} \left[\left(\frac{y_{1,n} - \mu_{Y_{1,n}|n-1}}{s_{Y_{1,n}|n-1}} \right)^2 \right. \\ \left. - 2r_{n|n-1} \frac{(y_{1,n} - \mu_{Y_{1,n}|n-1})(y_{2,n} - \mu_{Y_{2,n}|n-1})}{s_{Y_{1,n}|n-1} s_{Y_{2,n}|n-1}} + \left(\frac{y_{2,n} - \mu_{Y_{2,n}|n-1}}{s_{Y_{2,n}|n-1}} \right)^2 \right], \end{aligned}$$

em que

$$\begin{aligned} \mu_{Y_{j,n}|n-1} &= A_j + (y_{j,n-1} - A_j) \exp\{-b_j(t_n - t_{n-1})\}, \\ s_{Y_{j,n}|n-1} &= \sigma_j^2 (1 - \exp\{-2b_j(t_n - t_{n-1})\}) / (2b_j) \end{aligned}$$

e

$$r_{n|n-1} = \rho \frac{1 - \exp\{-(b_1 + b_2)(t_n - t_{n-1})\}}{(1 - \exp\{-2b_1(t_n - t_{n-1})\})^{1/2}(1 - \exp\{-2b_2(t_n - t_{n-1})\})^{1/2}} \frac{2(b_1 b_2)^{1/2}}{b_1 + b_2}. \quad (4.20)$$

A expressão (4.20) obtém-se como em (4.19) usando nos cálculos, nomeadamente da covariância, t_n e t_{n-1} em vez de t e t_0 .

Sejam $y_{1,n}$ e $y_{2,n}$ ($n = 1, \dots, N$), respetivamente, os valores observados de $Y_1(t_n)$ e de $Y_2(t_n)$, e seja $\mathbf{p} = (A_1, A_2, b_1, b_2, \sigma_1, \sigma_2, \rho)$ o vetor dos parâmetros. Usando a propriedade de Markov na solução da EDE, a função de log-verosimilhança, \mathcal{L} , dos logaritmos das TBMs, para uma certa idade, é dada por

$$\begin{aligned} \mathcal{L}(\mathbf{p} | Y_1(t_1), \dots, Y_2(t_1) \dots Y_1(t_N), \dots, Y_2(t_N)) \\ = \sum_{n=1}^N \ln(f(Y_1(t_n), Y_2(t_n) | Y_1(t_{n-1}), Y_2(t_{n-1}))). \end{aligned} \quad (4.21)$$

Também neste caso, as estimativas dos parâmetros, para cada idade, foram obtidas por métodos numéricos através da minimização do simétrico da função \mathcal{L} em (4.21), utilizando a função `nlm` do R (veja-se o código no apêndice D). Recordemos ainda que, em toda a computação, usamos $t_n - t_{n-1} = 1$ (já referido nos capítulos anteriores).

Para cada parâmetro, podemos ainda obter uma aproximação dos limites de um intervalo de confiança, para um certo nível de confiança, tal como descrito para os modelos apresentados anteriormente (ver, por exemplo, na secção 4.2).

Quanto a previsões, se tivermos observações até um certo instante t_N e quisermos estimar valores dos logaritmos das TBMs para $t > t_N$, usando a propriedade de Markov em $Y_j(t)$, temos que

$$E[Y_j(t)|Y_1(t_0), Y_2(t_0), Y_1(t_1), Y_2(t_1), \dots, Y_1(t_N), Y_2(t_N)] = E[Y_j(t)|Y_1(t_N), Y_2(t_N)].$$

Como

$$Y_j(t)|Y_1(t_N) = y_{1,N}, Y_2(t_N) = y_{2,N} \\ \sim \mathcal{N}(A_j + (y_{j,N} - A_j) \exp\{-b_j(t - t_N)\}, \sigma_j^2(1 - \exp\{-2b_j(t - t_N)\})/2b_j),$$

para cada idade, podemos usar para previsões a LP,

$$\hat{Y}_j(t) = \hat{E}[Y_j(t)|Y_1(t_N) = y_{1,N}, Y_2(t_N) = y_{2,N}] = \hat{A}_j + (y_{j,N} - \hat{A}_j) \exp\{-\hat{b}_j(t - t_N)\}, \quad (4.22)$$

onde \hat{E} representa o valor aproximado da esperança matemática substituindo os valores exatos de A_j e de b_j pelos dos seus estimadores de MV, respetivamente, \hat{A}_j e \hat{b}_j . As previsões PP são estimadas da mesma forma que em (4.22), mas atualizando anualmente t , $Y_1(t)$ e $Y_2(t)$, bem como as estimativas dos parâmetros (a computação é feita de modo análogo ao caso do modelo unidimensional, com as alterações resultantes da estimação do corrente modelo).

4.3.3 Resultados

Começamos por apresentar, na figura 4.7, as estimativas dos parâmetros do MGEB para todas as idades. Para efeitos de comparações com a versão unidimensional, podemos designar este modelo de modelo completo (pois no fundo é um MGE que considera, adicionalmente, a correlação entre sexos).

Note-se que as formas descritas nos gráficos dos valores estimados de a , b e σ são semelhantes às obtidas no caso do modelo unidimensional, as quais já interpretámos na subsecção 3.3.3. Quanto a ρ , os seus valores são semelhantes aos obtidos para o MBGB (ver figura 4.3 da secção anterior). Assim, os valores de ρ estimados são mais elevados nas primeiras idades, sendo que, relativamente às restantes idades adultas (e também às idades jovens), ocorre um aumento progressivo do valor deste parâmetro sensivelmente a partir da idade 50 e até à idade 80 (em que ocorre o valor mais elevado, superior a 0.8), quando se inicia uma diminuição dos valores, aproximando-se dos níveis estimados para as restantes idades (sensivelmente entre 0 e 0.4) muito perto da idade 100.

Foi também possível estimar os respetivos intervalos de confiança, \mathcal{IC} , assintóticos, associados aos parâmetros estimados, cujos valores ilustramos na tabela 4.3, para algumas idades (concretamente as idades 18, 40, 69, 82 e 95), para uma confiança de 95%. Na referida tabela, é visível que as estimativas dos parâmetros b_1 e b_2 apresentam as maiores margens de erro associadas, relativamente às estimadas para os restantes parâmetros, o que resulta, neste caso, numa maior amplitude dos intervalos de confiança.

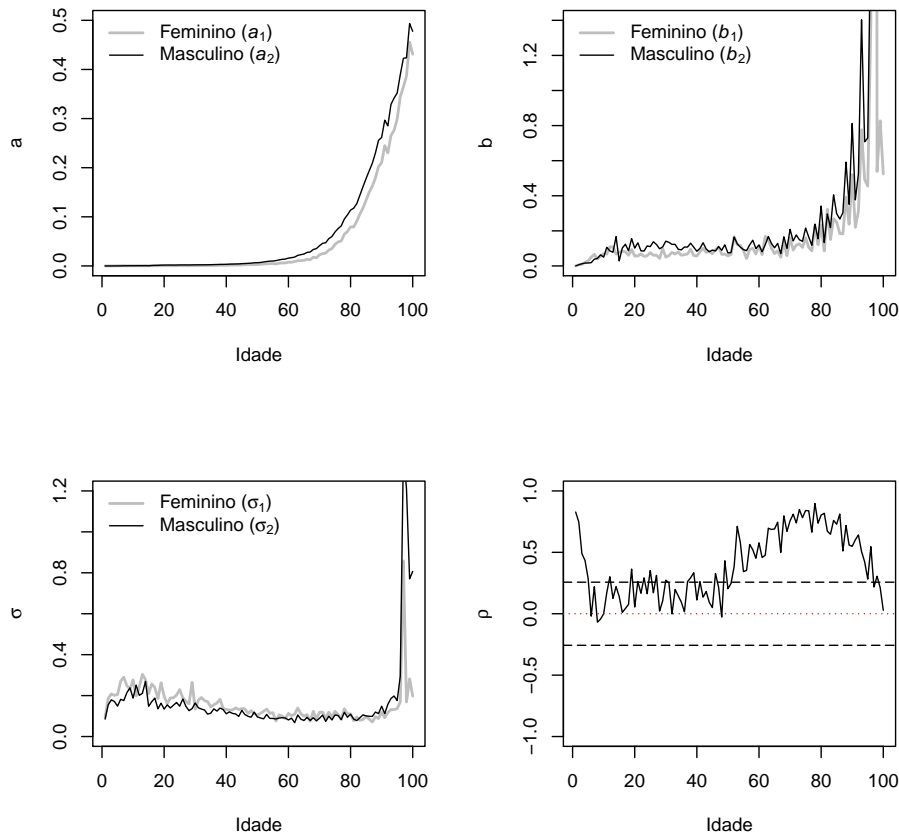


Figura 4.7: Estimativas dos parâmetros do MGEb, por idade e para ambos os sexos. No último gráfico, indica-se a região de aceitação (a tracejado) da hipótese $\rho = 0$ para o nível de significância de 5% (para o efeito usamos a transformação de Fisher como sugerido, por exemplo, em [82])

Tabela 4.3: Parâmetros estimados do MGEb com semi-amplitude dos $\mathcal{IC}_{95\%}$ assintóticos para as idades 18, 40, 69, 82 e 95

Idade	$A_1 = \ln(a_1)$	$b_1(\text{ano}^{-1})$	$\sigma_1(\text{ano}^{-1/2})$	ρ	$A_2 = \ln(a_2)$	$b_2(\text{ano}^{-1})$	$\sigma_2(\text{ano}^{-1/2})$
18	-7.5273 (± 0.8161)	0.0940 (± 0.0889)	0.2626 (± 0.0493)	0.3636 (± 0.2256)	-6.4082 (± 0.2861)	0.1554 (± 0.1197)	0.1626 (± 0.0315)
40	-6.3655 (± 0.4828)	0.0811 (± 0.0716)	0.1291 (± 0.0240)	0.2615 (± 0.2436)	-5.6030 (± 0.3376)	0.1000 (± 0.0922)	0.1206 (± 0.0227)
69	-3.7801 (± 0.1963)	0.1607 (± 0.1370)	0.1190 (± 0.0254)	0.7570 (± 0.1186)	-3.1916 (± 0.1118)	0.2085 (± 0.1690)	0.0893 (± 0.0193)
82	-2.2488 (± 0.1440)	0.1509 (± 0.1275)	0.0790 (± 0.0166)	0.6470 (± 0.1618)	-1.9378 (± 0.1126)	0.2184 (± 0.1738)	0.0918 (± 0.0195)
95	-1.0638 (± 0.0331)	1.4030 (± 0.8638)	0.1692 (± 0.0537)	0.5417 (± 0.1794)	-0.9432 (± 0.0438)	1.9655 (± 1.4771)	0.2949 (± 0.1188)

Na figura 4.8, mostramos ainda uma concretização, neste caso ilustrada com a primeira idade do sexo feminino, do ajustamento do modelo, previsões PP e a LP (no gráfico de cima, as previsões são para o período de 2000 a 2024 e nos abaixo, para 2000 a 2009).

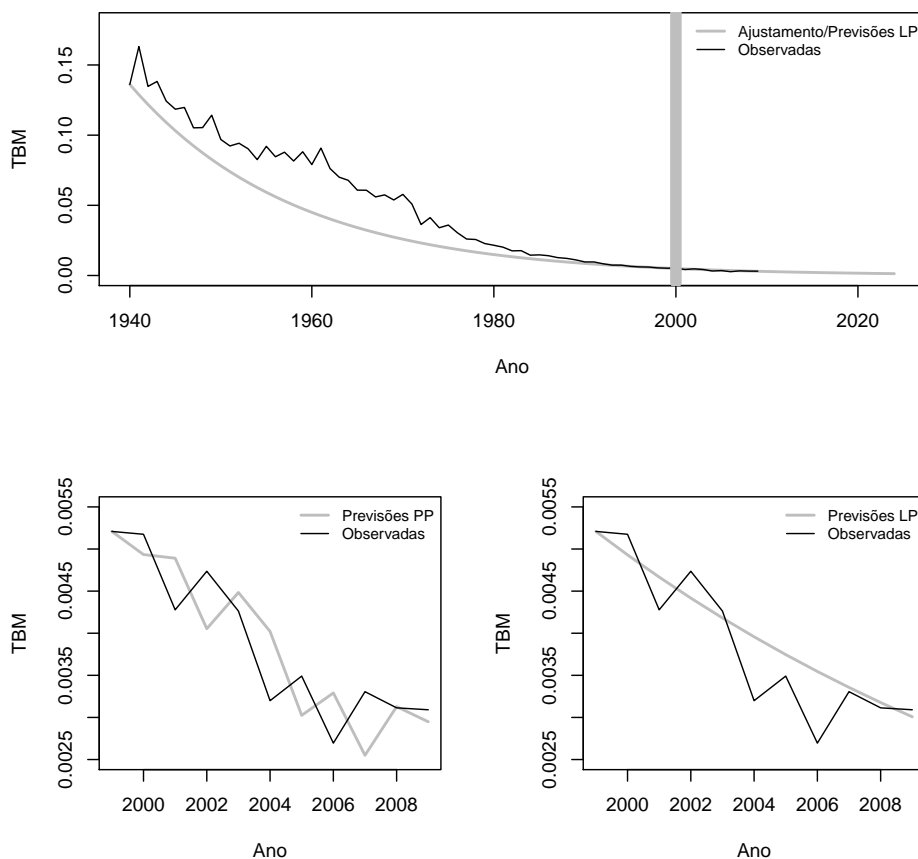


Figura 4.8: Ajustamento do MGEb com previsões a LP (25 anos: de 2000 a 2024) para a idade 0 do sexo feminino (em cima); ampliação das previsões (PP e a LP, no período de 2000 a 2009, respetivamente, à esquerda e à direita, em baixo)

Pelo exposto, considerarmos o MGEb adequado para este tipo de dados, apresentando ajustamentos e previsões que consideramos satisfatórias.

Vamos ainda comparar o modelo com o efeito das correlações por sexo com o modelo sem correlações. Para o efeito, podemos aplicar um teste de razão de verossimilhanças, à semelhança (e de modo análogo) ao que aplicámos em 4.2, para comparar a versão bidimensional do movimento browniano geométrico com a versão unidimensional. Na tabela 4.4, apresentamos os valores estimados da estatística de teste D para todas as idades. Os resultados sugerem, para um nível de significância $\leq 5\%$, que o modelo completo (com correlações entre os processos de Wiener que descrevem o efeito das flutuações ambientais) tem diferenças significativas quando comparado com o modelo sem o efeito das correlações, em quase todas as idades depois da idade 50 e também nas primeiras idades.

Nas figuras 4.9, relativamente aos ajustamentos, e 4.10, para as previsões a LP, representamos ainda (e à semelhança do que fizemos para o MBG) a os erros da versão uni e bidimensional do MGE, isto é, $EQM_{MGE} - EQM_{MGEb}$. Em ambos os casos, os gráficos sugerem um melhor desempenho global do modelo unidimensional (como também sucede para o MBG). Os padrões são semelhantes por sexo e a diferença dos erros tem uma tendência crescente com a idade (exceção para as primeiras idades, no caso dos ajustamentos e para as últimas idades no caso quer dos ajustamentos, quer das previsões, e em ambos os sexos). Note-se, no entanto, que muito embora os gráficos sugiram uma vantagem para os modelos unidimensionais, as diferenças entre os erros são muito reduzidas (mesmo multiplicadas por 10000, sendo que apenas nos últimos grupos de idades aumenta visivelmente a ordem de grandeza).

Tabela 4.4: Resultados do teste de razão de verosimilhanças entre o MGEb e o MGE para todas as idades
(* representa valores-p < 0.05; ** representa valores-p < 0.01)

Idade	D	Idade	D	Idade	D	Idade	D
0	68.33150**	25	1.67865	50	3.53230	75	68.47583**
1	47.86390**	26	5.88632*	51	8.85401**	76	32.92903**
2	15.98596**	27	0.02848	52	39.76502**	77	92.84271**
3	12.15687**	28	0.62421	53	22.76972**	78	43.81416**
4	5.00230*	29	4.43612*	54	6.96348**	79	56.39553**
5	0.02140	30	3.79528	55	7.50503**	80	63.33385**
6	2.80414	31	0.00019	56	21.43279**	81	30.82691**
7	0.26830	32	2.26401	57	17.48574**	82	29.31415**
8	0.10529	33	1.23678	58	12.78173**	83	44.10562**
9	0.00000	34	0.51013	59	22.33432**	84	40.63140**
10	1.66353	35	0.01452	60	13.31082**	85	59.34599**
11	5.34736*	36	4.05373*	61	14.36165**	86	15.86499**
12	0.86281	37	5.02216*	62	37.07712**	87	36.90376**
13	2.67163	38	6.73935**	63	35.64075**	88	19.50755**
14	1.14750	39	0.71664	64	36.53741**	89	17.42499**
15	0.00688	40	4.02279*	65	46.17152**	90	23.87252**
16	0.09950	41	1.02278	66	16.15175**	91	27.90838**
17	0.33019	42	1.88173	67	56.88815**	92	15.31689**
18	8.12858**	43	0.48002	68	34.06079**	93	10.46706**
19	0.17147	44	0.13951	69	46.69097**	94	3.86429*
20	4.09134*	45	6.39170*	70	61.06492**	95	20.12945**
21	1.62040	46	2.63369	71	44.56891**	96	2.68888
22	5.15558*	47	0.04081	72	72.83185**	97	0.81996
23	1.54344	48	11.49340**	73	53.77353**	98	2.06987
24	7.55196**	49	2.34264	74	68.18138**	99	0.03821

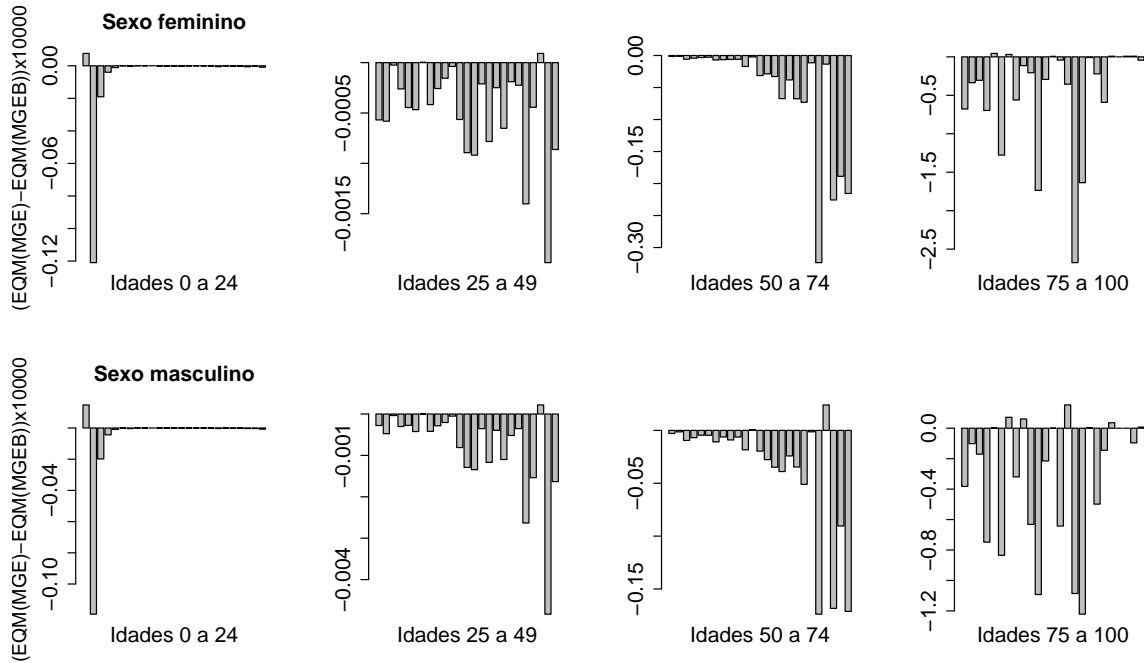


Figura 4.9: Diferença ($\times 10000$) entre os EQMs associados ao ajustamento às TBMs do MGE e do MGEB, por idade; sexo feminino, em cima, e masculino, em baixo

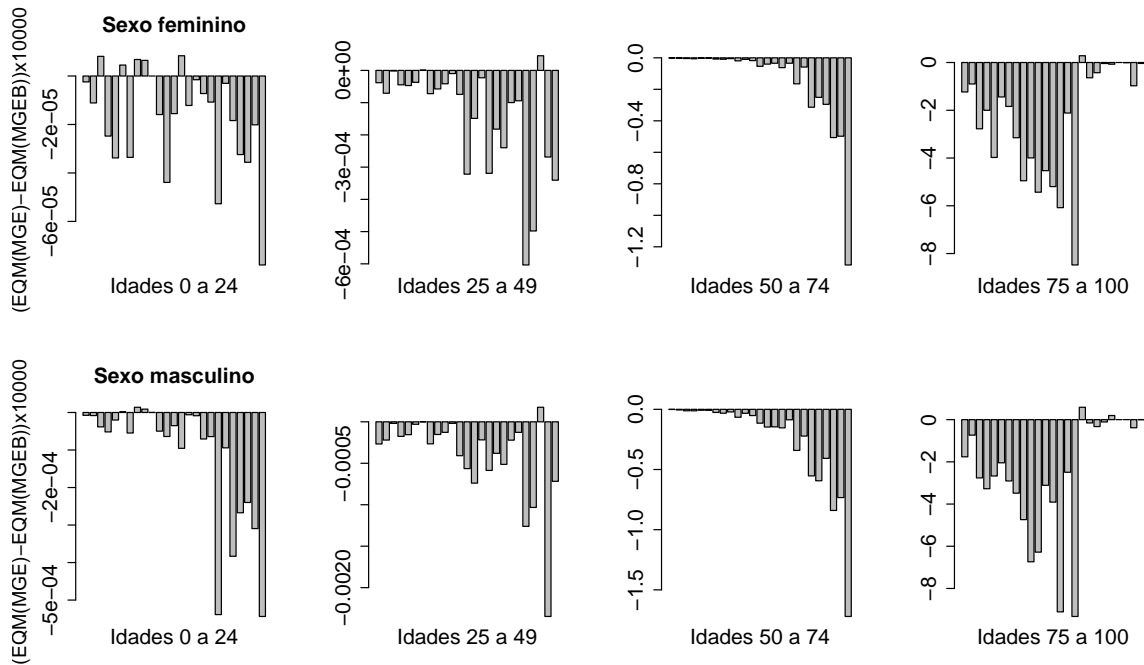


Figura 4.10: Diferença ($\times 10000$) entre os EQMs associados às previsões a LP das TBMs (de 2000 a 2009) do MGE e do MGEB, por idade; sexo feminino, em cima, e masculino, em baixo

Por último, vamos ainda comparar o desempenho dos dois modelos bidimensionais, MBGB e MGEb, no que diz respeito ao ajustamento e às previsões a LP, através da diferença entre os EQMs que lhe estão associados, isto é, $EQM_{MBGB} - EQM_{MGEb}$. Nas figuras 4.11 e 4.12, apresentamos os resultados.

Os gráficos para as diferenças dos erros entre os dois modelos bidimensionais de EDEs sugerem um melhor desempenho global do MBGB. No caso dos ajustamentos, no entanto, em algumas das idades entre os 25 e os 49 anos (sobretudo no sexo masculino) e também depois da idade 90, a diferença entre os EQMs dos dois modelos é favorável ao MGEb.

Note-se que, de um modo geral, neste caso, as diferenças dos erros são superiores às estimadas para os modelos unidimensionais, tanto no caso dos ajustamentos como das previsões a LP. Com exceção da primeira idade (e do grupo de idades entre os 25 e 49, para os ajustamentos), essas diferenças aumentam com a idade. Os gráficos mostram também que as diferenças entre os erros são maiores para o sexo masculino (exceção para o grupo de idades entre os 75 e os 99, para os ajustamentos), o que sugere, para as idades mais avançadas, as maiores diferenças entre os ajustamentos dos dois modelos, quer para o mesmo sexo, quer entre sexos.

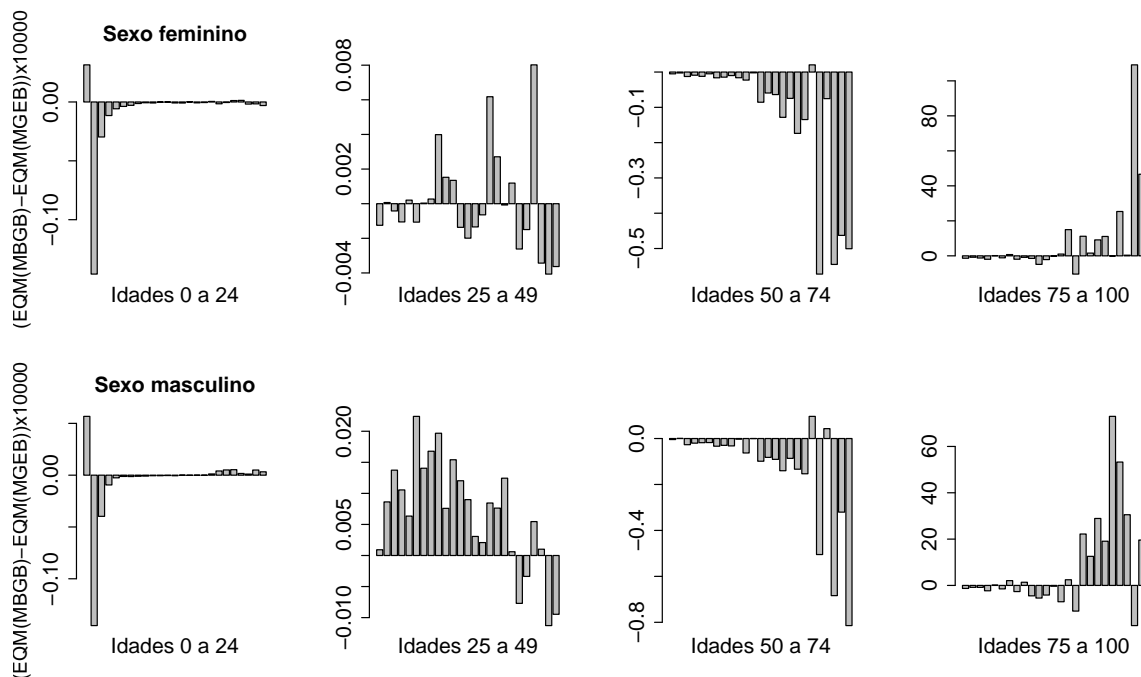


Figura 4.11: Diferença ($\times 10000$) entre os EQMs associados ao ajustamento às TBMs do MBGB e do MGEb, por idade; sexo feminino, em cima, e masculino, em baixo

Na secção seguinte, vamos comparar o desempenho de um dos modelos bidimensionais agora propostos com um modelo VARMA (à semelhança do que fizemos para o caso unidimensional no capítulo anterior para os modelos de EDEs e os modelos ARIMA).

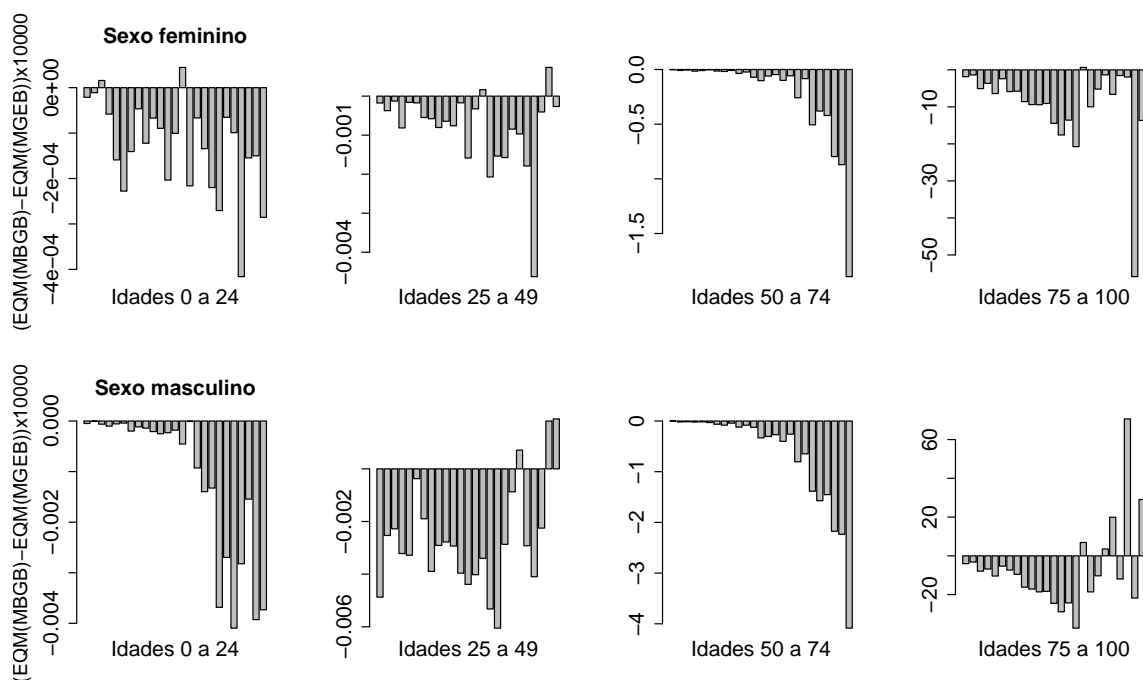


Figura 4.12: Diferença ($\times 10000$) entre os EQMs associados às previsões a LP das TBMs (de 2000 a 2009) do MBGB e do MGEB, por idade; sexo feminino, em cima, e masculino, em baixo

4.4 Comparação entre o movimento browniano geométrico bidimensional e um modelo de vetores autorregressivos e/ou de médias móveis

Conforme exposto no início deste capítulo, para explicar a evolução da mortalidade da população humana e para prever o seu comportamento a médio/longo prazo, podemos considerar uma análise bidimensional, transversal ao longo do tempo, à custa da informação das séries temporais e das correlações entre sexos, fixando uma certa idade. Esta abordagem foi sugerida pela análise exploratória dos dados, conforme já referido anteriormente e ilustrado na figura 4.2.

Nas secções anteriores, aplicámos os modelos bivariados de EDEs propostos. No final desta secção iremos comparar os resultados das previsões obtidos pelo movimento browniano geométrico bidimensional (MBGB) com os agora obtidos através da modelação com um modelo de vetores autorregressivos e de médias móveis (VARMA), em ambos os casos aplicados aos logaritmos dos retornos das TBMs da população portuguesa no mesmo período de tempo. A primeira metodologia apresenta uma computação e construção mais complexas (conforme exposto na subsecção 4.2.2), a segunda, cuja implementação ilustramos nesta subsecção, com um esforço computacional menor (uma vez que usamos funções já implementadas no R). No final, comparamos o desempenho dos dois tipos de metodologias, no que se refere às previsões, por idade e por sexo.

O pressuposto de partida para a aplicação dos modelos do tipo VARMA é o mesmo que o

já referido para o MBGB descrito pela equação (4.1): considera-se que existe uma variabilidade ambiental (em sentido lato) que afeta de forma relacionada indivíduos de ambos os sexos.

De facto, e para clarificar qualquer questão que possa ser nesta altura colocada acerca de uma possível causalidade entre as séries temporais (questão habitual aquando da aplicação de modelos do tipo VARMA), refira-se que as causas para a mortalidade humana são externas às variáveis consideradas (ou, de outra forma, exteriores ao sistema composto pelas TBMs). Deste modo, o modelo bivariado de séries temporais que vamos aplicar não traduz definitivamente uma relação causal, para uma idade, entre a mortalidade do sexo feminino e a do masculino (ou vice versa), mas sim uma inter-relação, o que nos parece apropriado para a aplicação da metodologia. Como não conhecemos, neste ponto da investigação, a/as variável/eis ambiental/ais que causam esse efeito comum, faz sentido usar a informação das séries das TBMs que possa trazer vantagens para a modelação. Começamos por usar as correlações entre sexos diferentes, para cada idade, e aplicar um modelo bidimensional (em alternativa a trabalhar cada série individualmente).

Dado que, previamente, na modelação das séries unidimensionais, obtivemos resultados bastante satisfatórios com modelos do tipo autorregressivo, nesta primeira abordagem à metodologia com recurso a dados da mortalidade de uma população, mostramos a aplicação de um modelo de vetores autorregressivos (VAR) que, ainda que não diretamente comparável (embora, por construção, o MBGB seja equivalente), tem um número de parâmetros próximo dos dos modelos de EDEs propostos nas secções anteriores. Adicionalmente, e por estes modelos serem modelos dinâmicos, que traduzem as inter-relações entre as séries, podemos ainda extrair informações adicionais dos resultados, nomeadamente testando a *causalidade de Granger*, analisando as *funções de impulso-resposta* (ou de resposta ao impulso) e também estudando a *decomposição de variância dos erros*. A combinação destes três tópicos é frequentemente designada por *análise estrutural*. Este tipo de abordagem surgiu nos anos 80 do século passado com Sims (ver, por exemplo, [80]), essencialmente para medir o efeito de choques na política monetária em variáveis macroeconómicas e dando origem a modelos estruturais mais complexos em que se tira partido das inter-relações dinâmicas entre as variáveis e a informação que se pode daí extrair, nomeadamente, identificando períodos de estabilidade ao longo do tempo.

Os aspetos metodológicos relativos a estes modelos, à sua implementação bem como à análise estrutural complementar que lhe está associada estão exhaustivamente expostos na bibliografia de referência (veja-se, por exemplo, [54, 73, 76]). Para ilustrar a sua aplicação, apresentamos, de seguida, de modo sucinto, os pressupostos do modelo bivariado e os critérios de implementação, aplicados à idade 2, e a interpretação dos resultados. No final da subsecção, comparamos as previsões obtidas através do modelo do tipo VARMA com as obtidas pelo MBGB.

4.4.1 Modelos VARMA

À partida, sabemos que as TBMs, em cada idade, estão correlacionadas por sexo e que essas correlações são estatisticamente significativas para a maior parte dos pares das TBMs da mesma

idade de ambos os sexos. Seja $\mathbf{Z}(t) = (Z_{1t}, \dots, Z_{kt}, \dots, Z_{Kt})$, com $k = 1, \dots, K$, um processo multivariado de VARMA ou VARMA(p, q) (com p e q as ordens dos polinómios autoregressivos e de médias móveis, respetivamente), estacionário. Pode escrever-se na forma

$$\mathbf{Z}(t) = \boldsymbol{\nu} + \mathbf{A}_1 \mathbf{Z}(t-1) + \dots + \mathbf{A}_p \mathbf{Z}(t-p) + \boldsymbol{\epsilon}(t) - \mathbf{M}_1 \boldsymbol{\epsilon}(t-1) - \dots - \mathbf{M}_q \boldsymbol{\epsilon}(t-q), \quad (4.23)$$

com $t = 0, \pm 1, \pm 2, \dots$, $\boldsymbol{\nu} = (\nu_1, \dots, \nu_k)$ um vetor de termos constantes com dimensão $K \times 1$, de modo a considerar-se um processo de média não nula. \mathbf{A}_a ($a = 1, \dots, p$) e \mathbf{M}_m ($m = 1, \dots, q$), ambas de dimensão $K \times K$, são, respetivamente, as matrizes dos parâmetros autorregressivos e de médias móveis, $\boldsymbol{\epsilon}(t)$ é um processo ruído branco (WN) K -dimensional, isto é, $E[\boldsymbol{\epsilon}(t)] = 0$, com matriz de variância-covariância não singular $\boldsymbol{\Sigma}_\epsilon = E[\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}(t)']$ e com $E[\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}(s)'] = 0$ para $s \neq t$.

Em termos práticos, e à semelhança do procedimento para os modelos unidimensionais ARIMA (cujos resultados da aplicação mostrámos em 3.4), começamos por identificar, para cada idade, o tipo de modelo (autoregressivo e/ou de médias móveis) e a respetiva ordem p , através da combinação de métodos empíricos e analíticos.

Para escolher o modelo de vectores autoregressivos para as séries selecionadas que vamos estudar conjuntamente, fazemos, como é habitual, a sua representação gráfica, de modo a ter uma perceção visual sobre a existência de tendência e mudanças na variância (neste caso, não há sazonalidade), bem como da existência de *outliers* (também não se admite a sua existência), descontinuidades ou mudanças de regime. Necessariamente, temos também que verificar a estacionaridade conjunta das séries (caso isso não aconteça, temos que transformar as séries por diferenciação, uma vez que já estamos a trabalhar com os dados numa escala transformada). Para selecionar a ordem p de um modelo multivariado utilizamos os critérios usuais de seleção deste tipo de modelos, em que é útil, por exemplo, a informação fornecida pelas funções de autocorrelação (FACs) e pelas funções de autocorrelação parciais (FACPs), mas também as funções de correlação cruzadas (FCCs), que podem dar informação acerca da intensidade da relação entre as séries temporais. Neste processo, usamos ainda, a título indicativo, a informação dos modelos univariados aplicados (cujos resultados ilustrámos no capítulo anterior, em 3.4).

Determinado o tipo de modelo e selecionada a ordem p , podemos estimar os parâmetros e executar os testes de diagnóstico aos resíduos (usuais neste tipo de metodologia) e aferir acerca da qualidade das previsões.

Particularizando a equação (4.23) para, por exemplo, um modelo VAR(1), temos

$$\mathbf{Z}(t) = \boldsymbol{\nu} + \mathbf{A}_1 \mathbf{Z}(t-1) + \boldsymbol{\epsilon}(t). \quad (4.24)$$

Estes processos admitem a decomposição em médias móveis com condições de estabilidade semelhantes aos modelos autorregressivos unidimensionais. Assim, assumindo que se verifica a condição de estabilidade $\det(I_K - A_1 z) \neq 0$ (para $|z| \leq 1$), podemos reescrever a equação (4.24) na forma

$$\mathbf{Z}(t) = \boldsymbol{\mu} + \sum_{l=0}^{\infty} \mathbf{A}_1^l \boldsymbol{\epsilon}(t-l), \quad (4.25)$$

com $\boldsymbol{\mu} = (\mathbf{I}_K - \mathbf{A}_1)^{-1} \boldsymbol{\nu}$.

A partir da expressão apresentada em (4.25), facilmente se deduzem também as expressões para as FACs e FACPs dos $\mathbf{Z}(t)$ e os primeiro e segundo momentos centrais. A estimação faz-se pelo *método dos mínimos quadrados*, para cada equação. O preditor de $\mathbf{Z}(t + l)$, que minimiza o EQM, pode ser obtido, para $l > 0$, a partir da esperança matemática de $\mathbf{Z}(t + l)$ condicionada aos valores de \mathbf{Z} até ao instante t , isto é,

$$\widehat{\mathbf{Z}}(t + l) = E[\mathbf{Z}(t + l) | \mathbf{Z}(t), \mathbf{Z}(t - 1), \dots].$$

No caso da aplicação às TBMs, cujos resultados apresentamos a seguir, para favorecer a estabilidade dos processos, optámos por aplicar uma transformação aos dados, trabalhando com os logaritmos dos retornos das TBMs, que representamos por $Z_j(t)$ e que podemos definir como $Z_j(t) = \ln(X_j(t)/X_j(t - 1))$, com $X_j(t)$ as TBMs ($j = 1$ para o sexo feminino; $j = 2$ para o masculino, ambas para uma mesma idade i). Recorde-se que já havíamos usado esta transformação (que corresponde a usar uma transformação logarítmica seguida de uma diferenciação simples) na secção 4.2, a propósito da estimação do MBGB. O objectivo é garantir, à partida, a estabilidade individual de cada série componente do processo.

Retomando o caso particular do modelo VAR(1), como em (4.24), mas considerando já a aplicação aos dados, podemos escrevê-lo mais explicitamente na forma

$$\begin{cases} Z_1(t) = \nu_1 + \alpha_{11}Z_1(t - 1) + \alpha_{12}Z_2(t - 1) + \epsilon_1(t) \\ Z_2(t) = \nu_2 + \alpha_{21}Z_1(t - 1) + \alpha_{22}Z_2(t - 1) + \epsilon_2(t), \end{cases} \quad (4.26)$$

em que os ν_j são os regressores determinísticos, α_{j1} e α_{j2} estão associados aos regressores dinâmicos, neste caso, ao polinómio autoregressivo. Assume-se que se tem para os resíduos $\epsilon_j(t)$, $E[\epsilon_j(t)] = 0$ e $E[\epsilon_j(t)\epsilon_l(\tau)] = 0$ (com $t \neq \tau$). Os processos de ruído branco $\epsilon_1(t)$ e $\epsilon_2(t)$ podem estar contemporaneamente correlacionados.

4.4.2 Análise estrutural

Consideremos agora as possíveis ligações existentes entre as variáveis do sistema composto pelas TBMs da população portuguesa, que estamos a estudar, e a informação que se pode daí extrair e que nos ajude a aumentar o conhecimento acerca da evolução do fenómeno ao longo do tempo (neste âmbito, seguimos a abordagem metodológica exposta em [54, 71, 72, 76]).

Neste sentido, um dos conceitos que estudamos é o de *causalidade de Granger*, que tem a ver com a capacidade de uma variável ou grupo de variáveis ser útil na previsão de outra ou de outras variáveis. Trata-se como que de uma precedência de causalidade e não da causalidade no sentido habitual (supõe-se que as informações relevantes para as previsões das variáveis estão nas séries temporais). Por outras palavras, se uma variável x afecta uma variável y , então, a primeira deverá ajudar a obter melhores previsões para a segunda e diz-se que x *causa à Granger* y .

Por sua vez, se considerarmos as interações entre as variáveis do sistema, podemos, ao analisar as *funções de impulso-resposta*, estudar o efeito no processo de choques simulados e detetar

relações dinâmicas ao longo do tempo, através do estudo da resposta de uma variável a choques ou impulsos noutras variáveis do sistema.

Por último, a terceira componente usual nesta análise estrutural diz respeito à *decomposição da variância dos erros* de previsão e visa determinar a importância relativa de cada variável de um modelo do tipo VARMA na explicação da variância dos resíduos das demais variáveis, ou qual a proporção de erros que é devida a choques no processo, em termos de perturbações ortogonais.

4.4.3 Exemplo ilustrativo de um modelo VAR(p)

O nosso objetivo nesta subsecção é mostrar a aplicação do modelo (4.26) (ou (4.24), na forma matricial). Para efeitos de representação gráfica e programação, usamos $F = Z_1(t)$ e $M = Z_2(t)$. A variável F representa aqui os logaritmos dos retornos das TBMs para a idade 2 dos indivíduos do sexo feminino e a M os do masculino. O seu comportamento pode ser observado nos gráficos da figura 4.13.

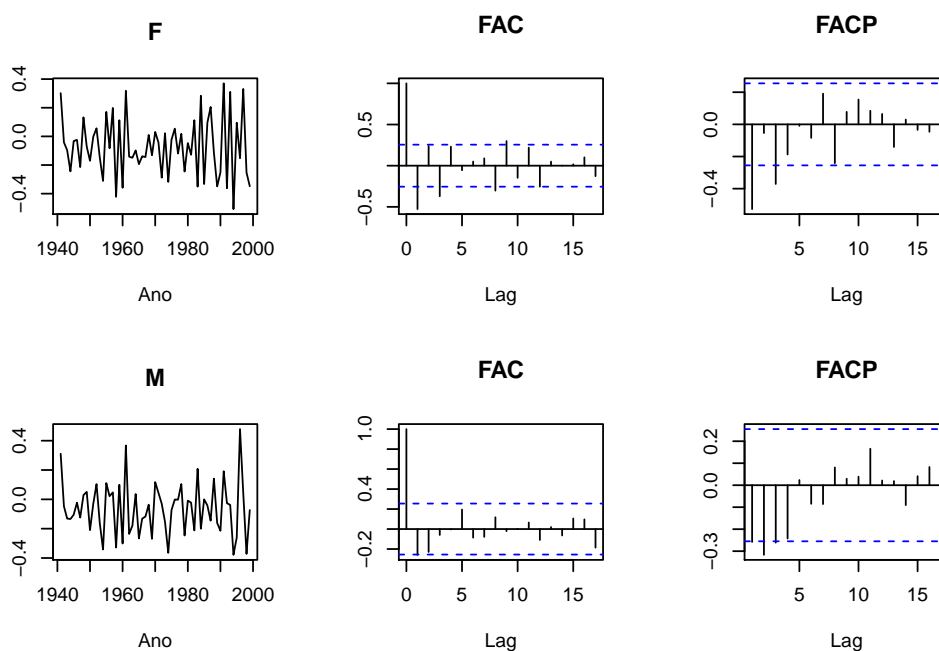


Figura 4.13: FACs e FACP dos logaritmos dos retornos das TBMs da idade 2 de ambos os sexos

A condição para a aplicação de um modelo VAR(p) é que as séries sejam conjuntamente estacionárias. Um processo multivariado é conjuntamente estacionário se as médias das séries temporais e todos os elementos da matriz de variância-covariância forem independentes do tempo para todos os desfasamentos ou *lags* (conforme definido em [17]). Para garantir que as séries são conjuntamente estacionárias (ver [55]), é condição necessária (mas não suficiente) que cada uma das séries seja estacionária.

Para testar a estacionaridade individual, usamos o teste ADF, de *augmented Dickey-Fuller test*, que verifica se existe tendência, *trend*, ou alterações na variância, *drift* (já que, à partida, nas TBMs

que traduzem médias anuais, não ocorrem fenómenos de variação sazonal). O teste admite na hipótese nula que o polinómio característico do processo VAR tem raízes dentro do círculo unitário contra a hipótese alternativa de a série ser estacionária.

Podemos ver nas caixas 4.1 e 4.2, os resultados do teste ADF, para a variável F . Os testes, que foram aplicados aos logaritmos dos retornos das TBMs sugerem que ambas as séries individuais são estacionárias (os valores-p obtidos para as séries individuais, F e M , são inferiores a 0.001), o que é também corroborado com o facto de não haver decaimento lento para zero das respetivas FACs e FACPs.

Caixa 4.1 Teste da raiz unitária de Dickey Fuller (*trend*) para a idade 2 do sexo feminino

```
# TESTAR ESTACIONARIDADE DA SÉRIE - TREND
> library(urca)
> adfF<-summary(ur.df(F,type="trend",lags=2))
> adfF

Call: lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
Residuals:
    Min       1Q   Median       3Q      Max
-0.53281 -0.08948  0.01922  0.09329  0.35399

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.182554   0.056303  -3.242  0.00209 **
z.lag.1      -1.883025   0.313667  -6.003 2.03e-07 ***
tt           0.003545   0.001509   2.349  0.02273 *
z.diff.lag1  0.421316   0.233888   1.801  0.07756 .
z.diff.lag2  0.127881   0.135705   0.942  0.35046

---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.1741 on 51 degrees of freedom
Multiple R-squared:  0.6984,    Adjusted R-squared:  0.6747
F-statistic: 29.52 on 4 and 51 DF,  p-value: 1.003e-12

Value of test-statistic is: -6.0033 12.1356 18.199
Critical values for test statistics:
      1pct  5pct 10pct
tau3  -4.04 -3.45 -3.15
phi2   6.50  4.88  4.16
phi3   8.73  6.49  5.47
```

A propósito da estacionaridade, refira-se que, ao aplicarmos a transformação às séries das TBMs designada por logaritmo dos retornos, estabilizámos simultaneamente quer a variância quer a média (pois estamos também a aplicar uma diferença simples) e podemos considerar todas as séries individualmente estacionárias. Recordamos que nos modelos ARIMA unidimensionais, usámos também a transformação logarítmica nas TBMs, na maior parte das séries. Nos casos em que essa transformação não era suficiente para garantir a estacionaridade (as séries exibem, frequentemente,

trajetórias que seguem uma evidente tendência negativa ao longo do tempo) foi aplicada também uma diferenciação simples, o que equivale a modelar os logaritmos dos retornos.

Caixa 4.2 Teste da raiz unitária de Dickey Fuller (*drift*) para a idade 2 do sexo feminino

```
# TESTAR ESTACIONARIDADE DA SÉRIE - DRIFT
> library(urca)
> adfdF<-summary(ur.df(F,type="drift",lags=2))
> adfdF

Call: lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
Residuals:
    Min       1Q   Median       3Q      Max
-0.48869 -0.12561  0.02129  0.09670  0.45328

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.06556   0.02738  -2.395  0.0203 *
z.lag.1     -1.66491   0.31236  -5.330 2.15e-06 ***
z.diff.lag1  0.28333   0.23602   1.200  0.2354
z.diff.lag2  0.07154   0.13925   0.514  0.6096
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.1815 on 52 degrees of freedom
Multiple R-squared:  0.6657,    Adjusted R-squared:  0.6464
F-statistic: 34.52 on 3 and 52 DF,  p-value: 2.036e-12

Value of test-statistic is: -5.3302 14.2095
Critical values for test statistics:
      1pct  5pct 10pct
tau2  -3.51 -2.89 -2.58
phi1   6.70  4.71  3.86
```

Como, na prática, os resultados dos testes ADF refletem apenas que os elementos da diagonal da matriz de variância-covariância são independentes do tempo, temos que testar o comportamento dos elementos fora da diagonal que correspondem às correlações cruzadas, para assegurar a estabilidade conjunta.

Para estudar as correlações cruzadas, as FCC devem ser estimadas nos mesmos *lags* sugeridos pelas FACs e FACPs individuais. Duas séries não são consideradas conjuntamente estacionárias se a FCC for significativamente diferente de zero nos *lags* em que as FAC e as FACP das séries individuais são significativos. Tal não se verifica, como se pode observar, na FCC relativa, por exemplo, à idade 2 (esta é apenas significativamente diferente de zero no *lag* 0, tal como se observa na figura 4.14), pelo que aceitamos a hipótese nula de as séries serem conjuntamente estacionárias. Obviamente que, se tivéssemos mais que duas séries (se considerássemos, por exemplo, as correlações entre as TBMs de idades diferentes), teríamos que testar todas as combinações possíveis de pares de séries e todos os pares teriam que ser conjuntamente estacionários. Neste

caso, como fixamos a idade, só testamos um par de cada vez.

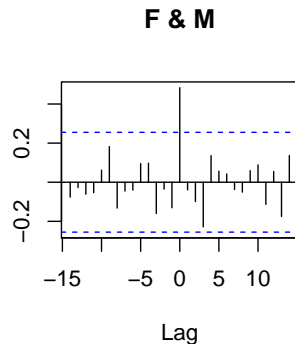


Figura 4.14: FCC entre os logaritmos dos retornos das TBMs da população portuguesa da idade 2 de ambos os sexos

Ainda relativamente à estacionaridade (pois é o principal pressuposto para a implementação da metodologia), se as séries não fossem estacionárias, poderiam voltar a diferenciar-se, ou, em alternativa, usar-se a teoria da cointegração (se duas variáveis são cointegradas significa que possuem um relacionamento estável e constante ao longo do tempo, por exemplo, apresentam uma tendência comum e geram resíduos estacionários).

$Z(t)$ é cointegrado se existe um vetor constante $\beta \neq 0$, tal que $\beta Z(t)$ é integrado de ordem $p < d$. Para o teste de cointegração, pode usar-se o procedimento de Johansen (ver, por exemplo, [72]). Assim, no caso de as séries não serem estacionárias, mas quando se verifica a existência de cointegração, aplica-se um modelo com correção de erros (do inglês *vector error correction model* ou VECM), ou um modelo para séries não estacionárias e não cointegradas (do inglês *structural vector correction error* ou SVEC). A implementação destes modelos e a análise estrutural é em tudo análoga aos modelos VAR (ver, por exemplo, [71]).

Na caixa 4.3 ilustramos a escolha da ordem p do modelo VAR(p) (com $1 \leq p \leq 8$). Considerando que para o exemplo da idade 2 se opta por um modelo do tipo autoregressivo, a ordem p foi determinada com recurso também à informação dada pelos critérios usuais neste tipo de metodologia, nomeadamente dos critérios de informação de Akaike (AIC), de Hannan-Quinn (HQ), Bayesiano de Schwarz (SC) e do erro de predição final de Akaike (FPE). Os critérios sugerem as ordens 1 ou 3. Neste caso, optou-se pela aplicação de um modelo VAR(1), uma vez que os parâmetros associados às ordens 2 e 3 não se revelaram estatisticamente significativos (com a vantagem adicional de o modelo VAR(1) necessitar de um menor número de parâmetros).

Testada a estacionaridade das séries e determinada a ordem p mais adequada, estimam-se os parâmetros do modelo (para o efeito, utilizámos a função VAR do R que usa o método dos mínimos quadrados como método de estimação). Os resultados do ajustamento podem ser observados na caixa 4.4.

Caixa 4.3 VAR(p): determinação da ordem p para a idade 2

```
# SELECIONAR MELHOR ORDEM P (LAG) PARA MODELO VAR
> library(vars)
> VARselect(Z,lag.max=8,type="both")

$selection
AIC(n)  HQ(n)  SC(n)  FPE(n)
   3     3     1     3

$criteria
          1          2          3          4          5          6          7          8
AIC(n) -7.13016461 -7.20737357 -7.29136950 -7.2857478 -7.21075636 -7.25372078 -7.26107703 -7.25302843
HQ(n)   -7.01436733 -7.03367764 -7.05977493 -6.9962545 -6.86336450 -6.84843027 -6.79788788 -6.73194064
SC(n)   -6.82713314 -6.75282636 -6.68530656 -6.5281691 -6.30166195 -6.19311062 -6.04895114 -5.88938680
FPE(n)  0.00080110  0.00074272  0.00068496  0.0006923  0.00075199  0.00072828  0.00073379  0.00075439
```

Caixa 4.4 VAR(1): ajustamento para a idade 2

```
# AJUSTAR MODELO VAR(1)
> library(vars)
> ajustVAR1 <- VAR(Z,p=1,type="const")
> ajustVAR1

Estimated coefficients for equation F (Call: F = F.l1 + M.l1 + const):
      F.l1      M.l1      const
-0.6889804  0.3407135 -0.1027108

Estimated coefficients for equation M (Call: M = F.l1 + M.l1 + const):
      F.l1      M.l1      const
-0.006310578 -0.2545965 -0.08539561

# TESTAR SIGNIFICÂNCIA DOS PARÂMETROS ESTIMADOS
> library(lmtest)
> coeftest(ajustVAR1)

t test of coefficients:
              Estimate Std. Error t value Pr(>|t|)
F:(Intercept) -0.1027108  0.0232957 -4.4090 4.868e-05 ***
F:F.l1        -0.6889804  0.1209438 -5.6967 4.945e-07 ***
F:M.l1         0.3407135  0.1384547  2.4608 0.0170289 *
M:(Intercept) -0.0853956  0.0240990 -3.5435 0.0008139 ***
M:F.l1        -0.0063106  0.1251138 -0.0504 0.9599555
M:M.l1        -0.2545965  0.1432285 -1.7776 0.0810060 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
```

Antes de avançarmos, é necessário verificar os pressupostos do modelo que têm a ver com os resíduos gerados pela aplicação aos dados. Para o efeito, foram efetuados testes de diagnóstico:

para a autocorrelação das séries (Portmanteau), de normalidade (Jarque Bera, achatamento e curtose) e de heterocedasticidade (*ARCH*), cujos resultados mostramos na caixa 4.5.

Caixa 4.5 VAR(1): testes de diagnóstico aos resíduos da idade 2

```
# TESTES DE DIAGNÓSTICO AOS RESÍDUOS DO MODELO VAR(1) AJUSTADO
> library(vars)
> Port.test <- serial.test(ajustVAR1,lags.pt=16,type="PT.asymptotic")
> Port.test$serial
      Portmanteau Test (asymptotic)
data:  Residuals of VAR object ajustVAR1
Chi-squared = 68.5962, df = 60, p-value = 0.2089

> Norm.test <- normality.test(ajustVAR1)
> Norm.test$jb.mul
$JB
      JB-Test (multivariate)
data:  Residuals of VAR object ajustVAR1
Chi-squared = 2.9889, df = 4, p-value = 0.5597

$Skewness
      Skewness only (multivariate)
data:  Residuals of VAR object ajustVAR1
Chi-squared = 0.4557, df = 2, p-value = 0.7962

$Kurtosis
      Kurtosis only (multivariate)
data:  Residuals of VAR object ajustVAR1
Chi-squared = 2.5332, df = 2, p-value = 0.2818

> Arch.test <- arch.test(ajustVAR1,lags.multi=5)
> Arch.test$arch.mul
      ARCH (multivariate)
data:  Residuals of VAR object ajustVAR1
Chi-squared = 62.8223, df = 45, p-value = 0.04058
```

Neste ponto, faremos uma análise conjunta dos resultados dos testes apresentados na caixa 4.5 e dos gráficos das figuras 4.15 e 4.16. Assim, a propósito da independência dos resíduos do modelo VAR(1), podemos observar nas FAC e nas FACP representadas nas figuras 4.15 e 4.16 que, efetivamente, para o nível de significância usual de 5%, ambas as funções não são estatisticamente nulas em alguns desfasamentos (caso do *lag* 3 na FACP dos resíduos de *F* e do *lag* 2, tanto na FAC como na FACP dos de *M*). Não obstante, consideramos que, por se tratarem de casos isolados, não estamos na presença de um padrão de correlação. Na realidade, o valor-p do teste de Portmanteau (apresentado na caixa anterior) não rejeita a hipótese nula de ausência de correlação. Note-se que usámos o desfasamento 16, *lags.pt*=16, sugerido usualmente na literatura para dados com esta dimensão, mas, para que não houvessem dúvidas, testámos também para o valor 8, ou mesmo 5, e o resultado do teste ainda é favorável neste último caso (*Chi-squared*=24.1653, *df*=16, *p-value*=0.08596, para um nível de significância de 5%). Pelo exposto, aceitamos a hipótese nula de

independência dos resíduos do modelo estimado.

Nas figuras 4.15 e 4.16, representamos as séries dos resíduos, conjuntamente com os respectivos histogramas, FAC e FACP, para complementar a informação dos resultados dos testes anteriores. Como se pode observar, os resíduos distribuem-se aleatoriamente em torno de um valor médio (próximo de zero), com alguns picos, o que eventualmente traduz um ajustamento menos bom a alguns valores mais elevados ou mais reduzidos (não têm necessariamente de ser valores extremos) das séries transformadas. Estas variações não causam, no entanto, uma alteração nos coeficientes de curtose ou de achatamento que ponham em causa a normalidade (sugerida, de resto, pela aparência do histograma). Na caixa 4.5, já referida, apresentámos os resultados do teste de Jarque-Bera multivariado para os resíduos do modelo VAR(1) estimado), que confirmaram a não rejeição da hipótese nula de normalidade.

Quanto ao resultado do teste *ARCH*, com um valor-p de 0.04058, reflete o que é visível nas FACs e FACP dos quadrados dos resíduos de F do modelo estimado: ambas as funções são significativamente diferentes de zero, para um nível de significância de 5%, nos *lag* 2 na FAC e 2 e 5 na FACP. Não obstante, decidimos continuar a trabalhar com este modelo VAR(1), pois consideramos aqueles resultados exceções. Na base dessa decisão está ainda o facto de não haver nenhuma sugestão de alteração no nível das variâncias, quer nas FACs, quer na FACP dos quadrados dos resíduos de M até ao *lag* 12 e porque o valor-p (0.04058) do teste global está muito perto de não rejeitar a hipótese nula de homocedasticidade dos resíduos para os 5%.

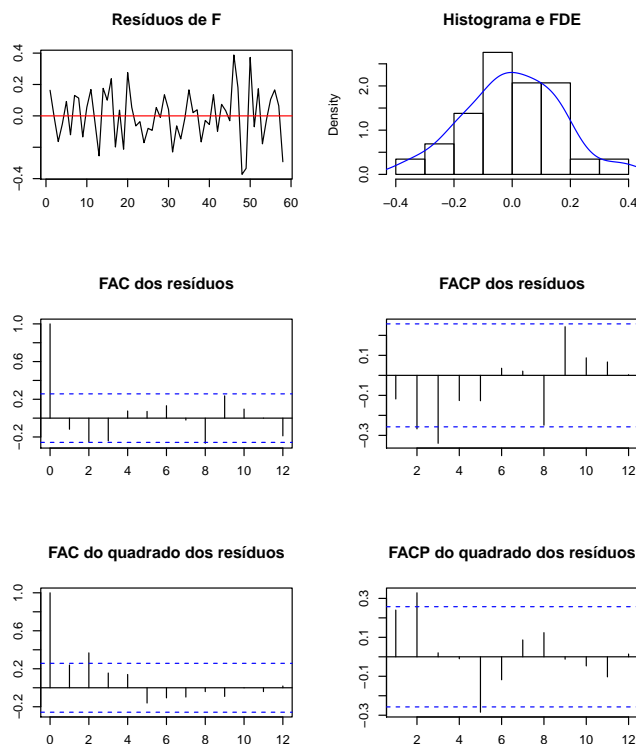


Figura 4.15: Resíduos do modelo VAR(1) ajustado à idade 2 do sexo feminino

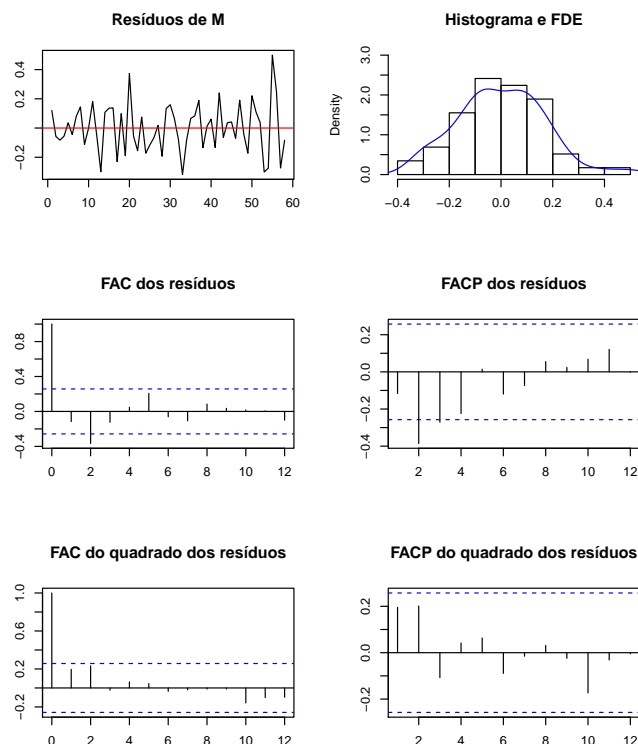


Figura 4.16: Resíduos do modelo VAR(1) ajustado à idade 2 do sexo masculino

Foram também efetuados testes de estabilidade do tipo *CUSUM* (como em [88]) aos resíduos do modelos VAR(1) aplicado. Os resultados não mostraram mudanças estruturais nas séries (ver figura 4.17, a seguir).

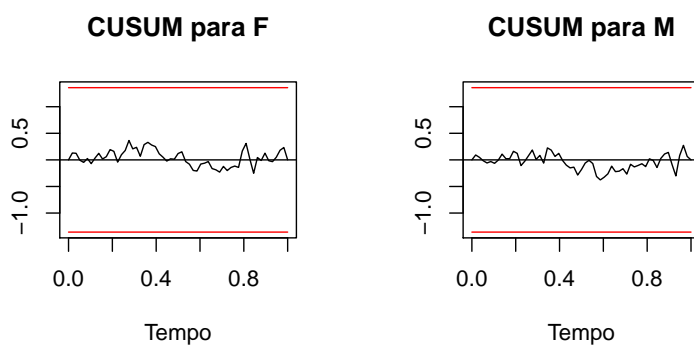


Figura 4.17: Carta de controlo tipo *CUSUM* para os resíduos do modelo VAR(1) ajustado

Considerando que os resultados dos testes aos resíduos vão ao encontro dos pressupostos do modelo, aceitamos o modelo VAR(1) estimado. Na figura 4.18, mostramos o ajustamento e, na figura 4.19, as previsões a LP (a 10 anos), com as respetivas bandas de confiança a 95%. Na caixa 4.6, mostramos os valores das previsões (para o período 2000 a 2009) e os limites dos $IC_{95\%}$.

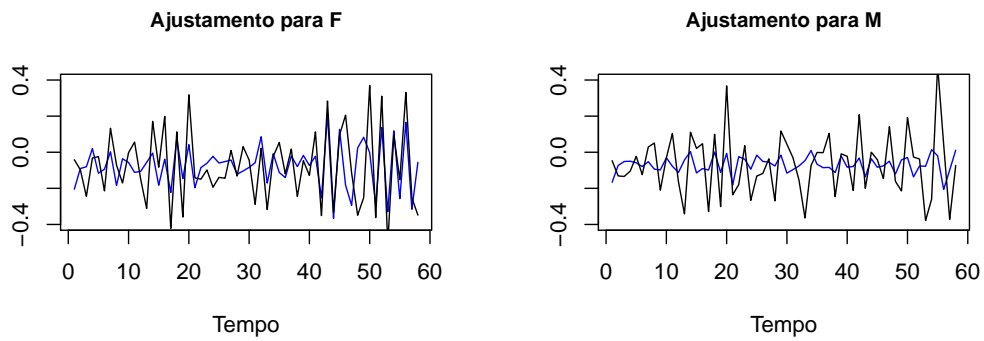


Figura 4.18: Ajustamento (a azul) do modelo VAR(1) aos logaritmos dos retornos das TBMs da idade 2 de ambos os sexos (no eixo do *Tempo*, o ano 0 corresponde a 1941 e o 60 ao ano 1999)

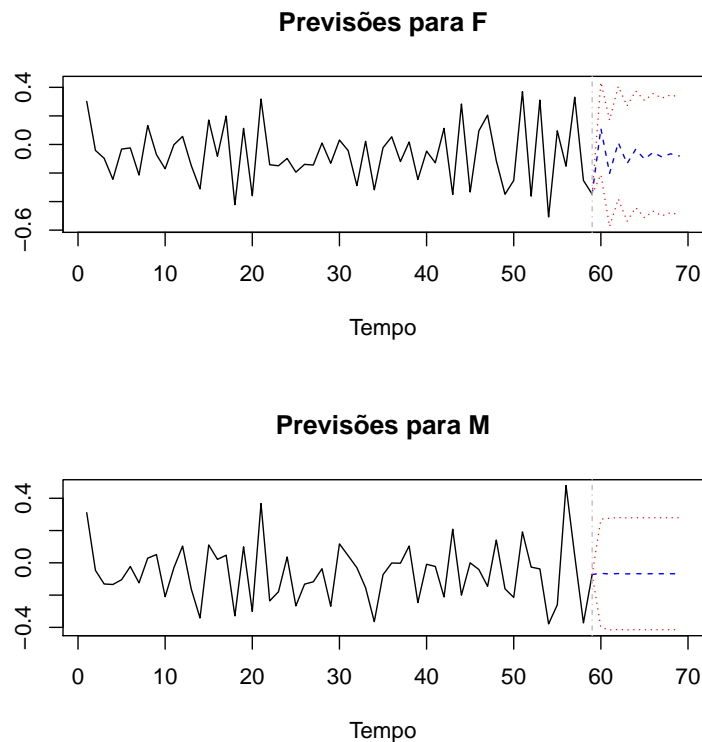


Figura 4.19: Previsões a LP (no período 2000 a 2009, a azul) do modelo VAR(1) dos logaritmos dos retornos das TBMs, com $IC_{95\%}$ (a vermelho), da idade 2 de ambos os sexos (no eixo *Tempo* aplica-se a mesma correspondência que para a figura 4.18, com o período das previsões do ano 61 ao 70)

Obtidas as previsões, procede-se à análise estrutural. Relativamente à causalidade de Granger, cujos resultados do teste conjunto mostramos na caixa 4.7, na primeira hipótese nula, isto é, “F não causa à Granger M”, há evidência estatística para não rejeitar a hipótese nula, ou seja, somos levados a concluir que não há vantagens em usar *F* para prever *M*. Já a segunda hipótese nula, “M não causa à Granger F”, é rejeitada, isto é, o teste sugere que há vantagens em usar *M* para prever

F. Recordamos que este resultado só é válido, porque, neste ponto da investigação, consideramos não conhecer as causas potenciais que influenciam a evolução das TBMs. A partir do momento em que identificarmos uma variável que influencie ambas *F* e *M*, a conclusão do teste pode significar que essa influência é primeiramente sentida em *M* que em *F* (como já referimos, a causalidade de Granger não traduz uma “verdadeira causalidade”, mas antes uma melhoria da capacidade de previsão do modelo e só deve por isso ser usada neste âmbito restrito).

Caixa 4.6 VAR(1): previsões a LP (2000 a 2009) dos logaritmos dos retornos das TBMs da idade 2 e por sexo

```
# PREVISÕES
> library(MSBVAR)
> Prev <- predict(ajustVAR1, 10)
> Prev$F
      fcst      lower      upper      CI
[1,]  0.11251211 -0.2114639  0.4364882  0.3239760
[2,] -0.20227382 -0.5749704  0.1704228  0.3726966
[3,]  0.01292696 -0.3810691  0.4069230  0.3939961
[4,] -0.13423748 -0.5381736  0.2696987  0.4039361
[5,] -0.03358801 -0.4421722  0.3749962  0.4085842
[6,] -0.10242770 -0.5131837  0.3083283  0.4107560
[7,] -0.05534369 -0.4671146  0.3564272  0.4117709
[8,] -0.08754775 -0.4997931  0.3246976  0.4122454
[9,] -0.06552108 -0.4779883  0.3469462  0.4124672
[10,] -0.08058672 -0.4931577  0.3319843  0.4125710
> Prev$M
      fcst      lower      upper      CI
[1,] -0.06470061 -0.3998470  0.2704458  0.3351464
[2,] -0.06963308 -0.4157630  0.2764969  0.3461300
[3,] -0.06639081 -0.4132291  0.2804475  0.3468383
[4,] -0.06857432 -0.4154582  0.2783095  0.3468839
[5,] -0.06708971 -0.4139769  0.2797975  0.3468872
[6,] -0.06810284 -0.4149906  0.2787849  0.3468878
[7,] -0.06741048 -0.4142985  0.2794775  0.3468880
[8,] -0.06788388 -0.4147720  0.2790042  0.3468881
[9,] -0.06756013 -0.4144483  0.2793280  0.3468882
[10,] -0.06778156 -0.4146698  0.2791066  0.3468882
```

Caixa 4.7 VAR(1): teste à causalidade de Granger para a idade 2

```
# TESTAR A CAUSALIDADE DE GRANGER
> library(vars)
> Granger.test.F <- causality(ajustVAR1, cause = "F")
> Granger.test.F$Granger
      Granger causality H0: F do not Granger-cause M
data:  VAR object ajustVAR1
F-Test = 0.0025, df1 = 1, df2 = 110, p-value = 0.9599
> Granger.test.M <- causality(ajustVAR1, cause = "M")
> Granger.test.M$Granger
      Granger causality H0: M do not Granger-cause F
data:  VAR object ajustVAR1
F-Test = 6.0557, df1 = 1, df2 = 110, p-value = 0.01542
```

Na figura 4.20, mostramos as funções de impulso-resposta que oscilam entre valores positivos e negativos ao longo do tempo, o que, embora sugira que o comportamento das variáveis face aos choques não é claro (dado a ausência de uma resposta com tendência crescente ou decrescente ao

longo do tempo), pode significar que a influência de choques no comportamento da variável não tem significado. Recordemos que a função de impulso-resposta ilustra a resposta de uma variável a um choque exógeno. Esse choque é normalmente uma combinação linear de pequenas perturbações, dificilmente observáveis.

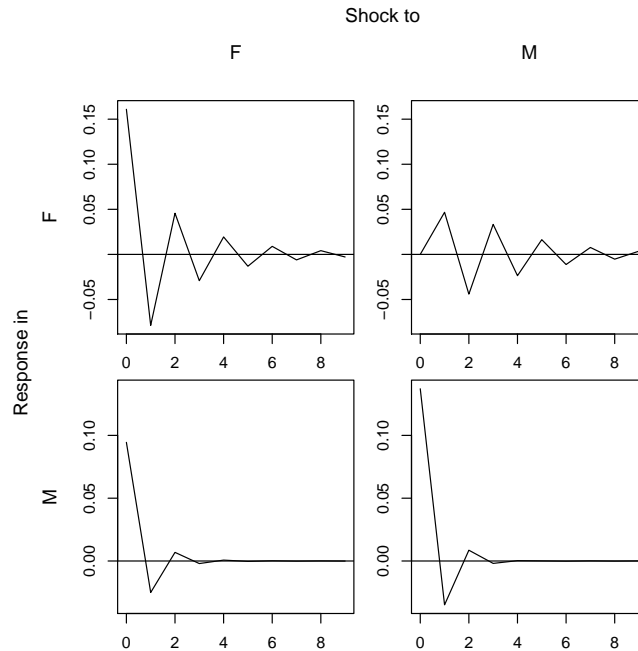


Figura 4.20: Funções de impulso-resposta do modelo VAR(1) ajustado aos logaritmos dos retornos das TBMs da idade 2

Consideremos $D(t)$ a variável que vai introduzir o choque (supomos que toma os valores 0 e 1) num modelo do tipo AR(1), que podemos escrever na forma $Z(t) = \mu + \phi Z(t - 1) + \xi D(t) + \epsilon(t)$. Queremos ilustrar, ao longo de um certo horizonte temporal t , como é que as previsões de $Z(t)$ variam quando $D(t)$ toma o valor 1 em vez de 0. Temos que ter em conta então o sinal de ϕ , pois, se for positivo, a resposta deverá aproximar-se de modo monótono de zero, ao passo que, se for negativo, oscila entre valores positivos e negativos, eventualmente com comportamento cíclico. Assim, e à semelhança do ilustrado na figura 4.20, a função impulso-resposta pode ser considerada um ruído e ser desprezável ou sem significado. Da figura 4.21, relativa à decomposição do erro das previsões, conclui-se todavia que cada variável explica uma proporção da variância do erro de previsões da outra, o que significa que não evoluem de modo independente ao longo do tempo.

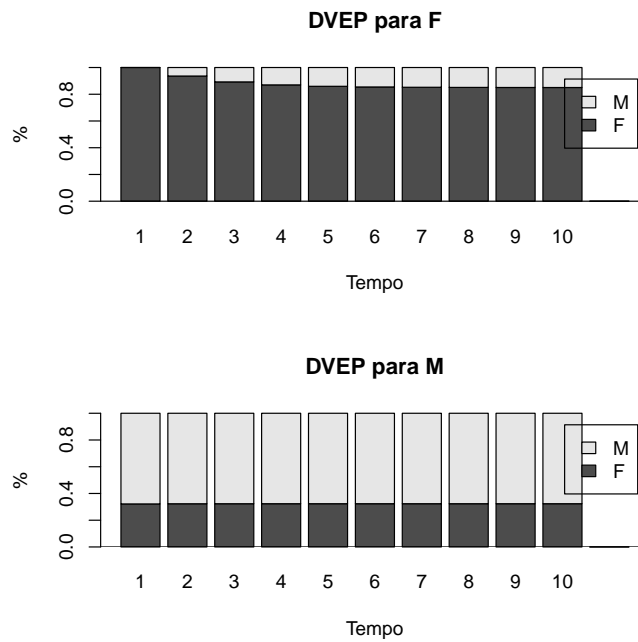


Figura 4.21: Decomposição do erro das previsões do modelo VAR(1) ajustado aos logaritmos dos retornos das TBMs da idade 2

4.4.4 MBGB vs VAR(1)

No início deste capítulo, dedicado aos modelos bivariados de EDEs, mostrámos, na secção 4.2, a conceção e a aplicação aos logaritmos das TBMs da população portuguesa, do modelo de EDEs que designámos por movimento browniano geométrico bidimensional (MBGB). Na corrente secção, temos vindo a mostrar a aplicação de um modelo estocástico bivariado do tipo VARMA, que concretizámos ilustrando com um exemplo de um modelo VAR(1), com a finalidade de fazermos agora uma comparação do desempenho das duas metodologias e aferir acerca das vantagens e inconvenientes da sua aplicação (à semelhança do que já havíamos feito no capítulo 3 para o caso unidimensional).

Na tabela 4.5, apresentamos os valores dos parâmetros estimados para o MBGB para a idade 2 (que tem vindo a servir de exemplo ilustrativo da aplicação da metodologia VARMA) e por sexo.

Tabela 4.5: Estimativas dos parâmetros do MBGB e semi-amplitude do $IC_{95\%}$ para a idade 2

MBGB	F	M
\tilde{R}	-0.071592 ± 0.084704	-0.062656 ± 0.072838
$\tilde{\sigma}$	0.650624 ± 0.153495	0.559486 ± 0.1319852
$\tilde{\rho}$	0.483473 ± 0.313110	

Na figura 4.22, representamos as previsões dos logaritmos dos retornos das TBMs obtidas pelos dois modelos, para os dois sexos. Conforme se pode observar não é óbvio, neste caso, qual o

modelo que melhor capta a variabilidade das séries. Tão pouco ocorre um padrão comum aos dois sexos, em termos de desempenho. Enquanto que no caso do sexo feminino há um evidente afastamento entre as previsões dos dois modelos, a figura sugere, no caso do sexo masculino, um desempenho semelhante dos mesmos (uma vez que a representação das previsões, do MBGB e do modelo VAR(1) pouco diferem).

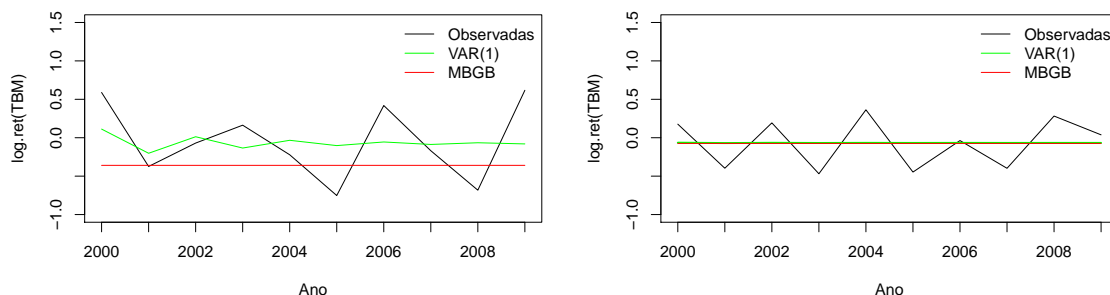


Figura 4.22: Comparação das previsões a LP (no período de 2000 a 2009) dadas pelo modelo VAR(1) e pelo MBGB, aplicados aos logaritmos dos retornos das TBMs, $\log.\text{ret}(\text{TBM})$, da idade 2 de ambos os sexos (feminino, à esquerda e masculino, à direita)

Na caixa 4.8, mostramos os valores estimados para as previsões a LP (no período 2000 a 2009) dos logaritmos dos retornos das TBMs (ilustradas na figura 4.22), obtidos pelos dois modelos: MBGB e VAR(1), para a idade 2 e por sexo.

Caixa 4.8 Previsões a LP (período 2000 a 2009), dos logaritmos dos retornos das TBMs, obtidas pelos modelos VAR(1) e MBGB para a idade 2 de ambos os sexos e estimativas de EQM associado

```
> prev.mod.bi<-cbind(Z[60:69,3],VAR1.F[,3],MBGB[,3],Z[60:69,103],VAR1.M[,3],MBGB[,103])
> #Z = logaritmos dos retornos das TBMs
      Z.F    VAR1.F    MBGB.F    Z.M    VAR1.M    MBGB.M
[1,] 0.58894340 0.11251211 -0.35947929 0.17687566 -0.06470061 -0.07391565
[2,] -0.37416810 -0.20227382 -0.35947929 -0.39696888 -0.06963308 -0.07391565
[3,] -0.06959329 0.01292696 -0.35947929 0.19325982 -0.06639081 -0.07391565
[4,] 0.16311935 -0.13423748 -0.35947929 -0.46916716 -0.06857432 -0.07391565
[5,] -0.22186885 -0.03358801 -0.35947929 0.36334164 -0.06708971 -0.07391565
[6,] -0.75218071 -0.10242770 -0.35947929 -0.44701411 -0.06810284 -0.07391565
[7,] 0.41888813 -0.05534369 -0.35947929 -0.03704127 -0.06741048 -0.07391565
[8,] -0.16907633 -0.08754775 -0.35947929 -0.39794628 -0.06788388 -0.07391565
[9,] -0.68267588 -0.06552108 -0.35947929 0.28204825 -0.06756013 -0.07391565
[10,] 0.61743536 -0.08058672 -0.35947929 0.03742641 -0.06778156 -0.07391565

> EQM.prev.modelos.bi
[1] - 0.025532 0.117717 - 0.017344 0.017824
```

Neste caso, o EQM das previsões do MBGB é superior aos EQMs associados ao modelo VAR(1),

para ambos os sexos. Por sua vez, conjuntamente, os erros das previsões, dos dois modelos, para o sexo masculino (que tomam valores muito próximos) são menores que os do sexo feminino e dentro de cada sexo.

De um modo geral, consideramos que ambos os modelos apresentam bons ajustamentos aos dados transformados. Para a maioria das idades, o EQM das previsões do modelo VAR(1) é superior ao do EQM do MGEB (embora próximos, pois, frequentemente, as previsões obtidas por via deste dois modelos têm tendência semelhante), mas inferior ao do MBGB. Estes resultados são contrários aos que apresentámos na subsecção 3.4.3, nomeadamente no comparativo entre os modelos unidimensionais do tipo ARIMA e o modelo de EDEs (no caso, o MGE) e sugerem que para mais que uma dimensão os modelos de EDEs têm uma melhor prestação, pelo menos no que se refere às previsões a LP, cuja estimação foi o objetivo nesta secção. Outras desvantagens da aplicação dos modelos VAR(p), prendem-se, por exemplo, com o número potencialmente elevado de parâmetros que aumenta com a ordem p (se esta for muito pequena, os modelos podem, por seu turno, ser menos precisos) e com as restrições ao nível da computação e métodos numéricos. Ao contrário do que sucedeu para os modelos de EDEs, não foi possível aplicar um modelo do tipo VAR(p), ou VARMA(p,q), a algumas idades (devido, em parte, aos resultados obtidos na fase de análise de diagnóstico dos resíduos, que não viabilizam o modelo escolhido). Nesta comparação, não estamos, todavia, alheios ao facto de os resultados diferentes poderem estar a acontecer porque usamos métodos de estimação distintos, no caso do MGEB, máxima verosimilhança e, no caso do modelo VAR(1), mínimos quadrados. Os modelos do tipo VAR(p) apresentam, não obstante, vantagens face aos modelos de EDEs, pois permitem interpretar relações dinâmicas entre as séries e são, ainda assim, bons modelos para previsões, como já referido.

No próximo capítulo, e na sequência do exposto no início deste capítulo acerca dos “comportamentos” observáveis nos dados (que sugerem uma influência simultânea, na maior parte das idades de ambos os sexos, dos fatores ambientais sobre a mortalidade), retomamos a construção, e também aplicação às taxas de mortalidade, de modelos de EDEs multidimensionais, neste caso, com $W(t)$ correlacionados por sexo e também por idade.

5

Modelos multivariados de equações diferenciais estocásticas para taxas de mortalidade

5.1	Introdução	98
5.2	Movimento browniano geométrico multidimensional com processos de Wiener correlacionados	100

5.1 Introdução

Considerámos, à partida, que existem fatores ambientais (em sentido lato) que podem afetar a mortalidade da população portuguesa. Por esse motivo, no capítulo 3, aplicámos modelos de EDEs (que incorporam o efeito da variabilidade ambiental na evolução das taxas com recurso a processos de Wiener) a cada série temporal das TBMs, por idade e por sexo. Observando os dados, considerámos também que esses fatores ambientais podem afetar de forma semelhante as TBMs dos indivíduos de ambos os sexos de uma certa idade (caso dos modelos bidimensionais propostos no capítulo 4) e/ou de idades diferentes.

Faz assim sentido a aplicação de modelos multidimensionais de EDEs com uma estrutura de correlação que ajude a explicar o fenómeno da mortalidade. Para o efeito, estudámos os padrões da mortalidade e as correlações associadas, quer por sexo, quer por idade (dentro e entre sexos).

Na figura 5.1, apresentamos a matriz das TBMs, discretizando os dados em 10 níveis de mortalidade (a gradação associa tons verdes a valores mais elevados de mortalidade, cinzentos a valores médios e roxos aos valores mais reduzidos). A categorização é feita baseada nos dados das séries temporais de todas as idades de ambos os sexos. Esta normalização permite identificar semelhanças e diferenças entre idades e entre sexos, dando pistas para a construção de uma possível estrutura de correlação. As variações entre sexos são maiores nas idades jovens-adultas, sensivelmente entre as idades 15 a 40, aumentando a dispersão (exceção para as TBMs associadas aos primeiros 2 anos de vida), em ambos os sexos, nas idades mais avançadas (a partir da idade 70).

Propomos, de seguida, um modelo multivariado de EDEs em que consideramos a existência de correlações entre idades do mesmo sexo e de sexos diferentes, que designamos por movimento browniano geométrico multidimensional (MBGM), usando processos de Wiener correlacionados por idade e também por sexo.

No final do capítulo, comparamos diferentes versões do modelo, consoante a estrutura de correlação definida.

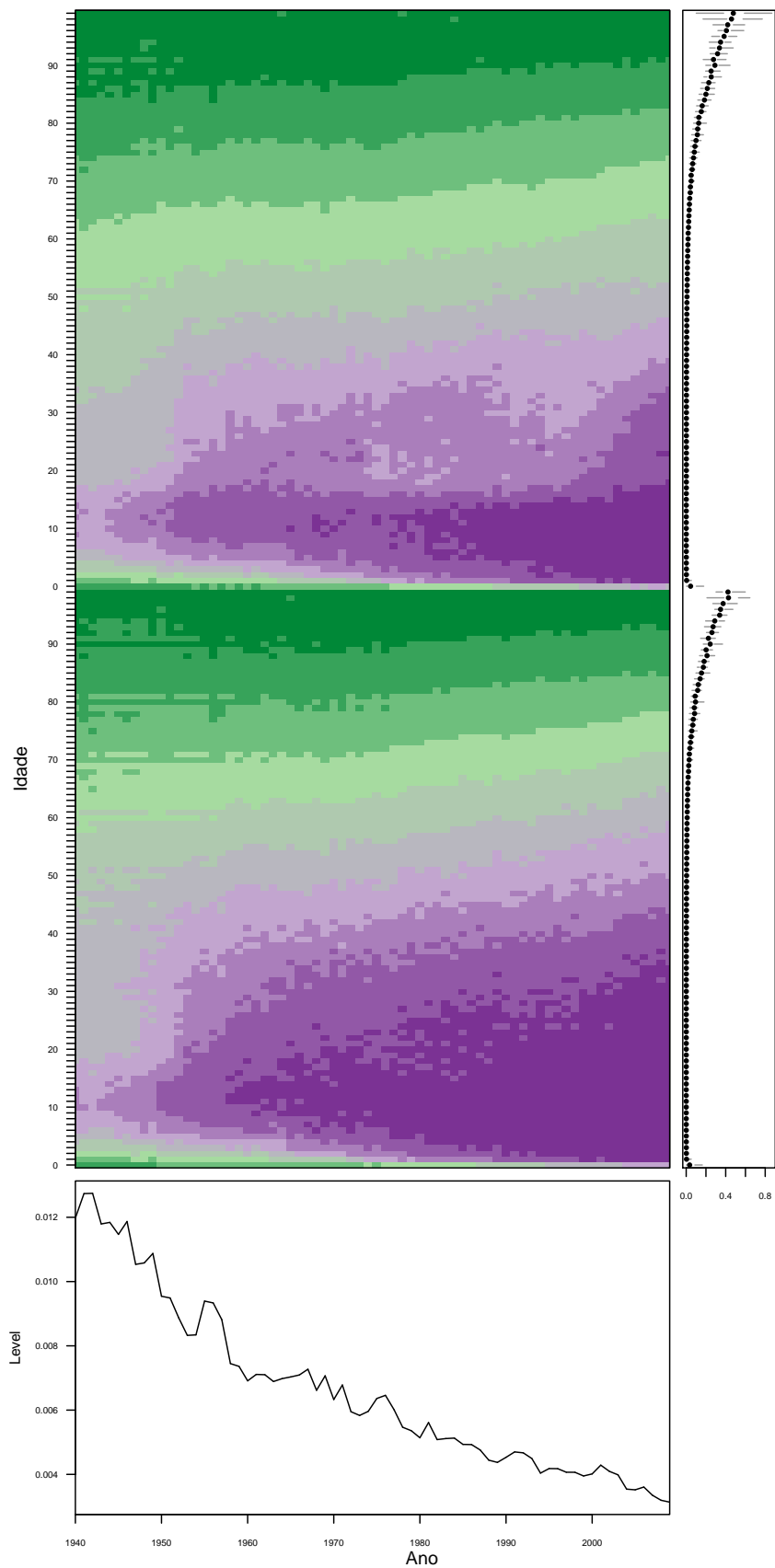


Figura 5.1: TBMs da população portuguesa (de 1940 a 2009) de todas as idades e de ambos os sexos (metade superior para o sexo masculino e inferior para feminino); na margem direita, temos a representação em caixa de bigodes das TBMs, para cada idade, e, em baixo, os valores medianos das TBMs de todas as idades e de ambos os sexos em cada unidade temporal

5.2 Movimento browniano geométrico multidimensional com processos de Wiener correlacionados

No capítulo 3, descrevemos a variabilidade das TBMs, aplicando modelos de EDEs a cada idade, por sexo. Na sequência dos resultados obtidos a partir da aplicação do modelo unidimensional do tipo movimento browniano geométrico (MBG), que apresentámos na secção 3.2, usamo-lo agora como modelo base para a construção de um modelo multidimensional de EDEs, com o objetivo de explicar a evolução temporal das TBMs e fazer previsões a longo prazo, admitindo que existem fatores ambientais que, ao longo do tempo, podem afetar de forma semelhante os indivíduos de ambos os sexos e de idades diferentes.

No capítulo anterior, apresentámos já dois modelos bidimensionais com correlação entre os processos de Wiener padrão associados às EDEs que modelam as TBMs de cada sexo, para uma certa idade: o MBGB e o MGEB.

Neste capítulo, consideramos como ponto de partida que cada idade, por sexo, tem a sua própria taxa de mortalidade e que a esta está associado um processo de Wiener ($W(t)$) que mede as flutuações ambientais sobre a mortalidade. Adicionalmente, notamos, quando observamos os dados, que existem picos de mortalidade que correspondem não só a aumentos, mas também a quebras nas TBMs, ao longo do tempo, que são comuns não só aos dois sexos, mas também às várias idades. Na figura 5.2, por exemplo, reparamos que, quando a TBM de uma das idades aumenta, ou diminui, de um ano para o outro, há uma forte tendência para que as TBMs das outras idades, quer do mesmo sexo, quer de sexos diferentes, tenham idêntico comportamento.

Pelo exposto, é natural admitirmos ainda que os $W(t)$ que medem a variabilidade ambiental em cada idade estão também correlacionados. Para captar essa estrutura de correlação num modelo multidimensional, a matriz de variâncias-covariâncias do vetor de processos de Wiener vai incorporar as correlações entre todos os pares de idades possíveis, dentro de cada sexo e entre sexos.

O modelo que a seguir propomos pode abarcar a mortalidade associada a todas as idades do arco de vida e de ambos os sexos, traduzindo uma explicação global dos efeitos dos fatores ambientais, neste caso, sobre as taxas de mortalidade. Esta conceção representa, na nossa opinião, uma enorme vantagem face à modelação individual das TBMs de cada idade e sexo, quer em termos computacionais, quer na redução do erro associado ao cálculo, com reflexos, nomeadamente, ao calcular grandezas derivadas das TBMs.

5.2.1 O MBGM

Consideremos o MBG que aplicámos a cada idade (designação que temos vindo a usar para cada um dos grupos etários anuais), por sexo, conforme enunciado na equação (3.6) da secção 3.2.

Seja K , o número de idades dos dois sexos, das quais $K/2$ relativas ao sexo feminino e $K/2$ ao masculino. No caso de usarmos as idades como representação dos grupos etários anuais dos 0 aos 99 anos, será $K = 200$. Se for conveniente (como é usual, por exemplo, em demografia)

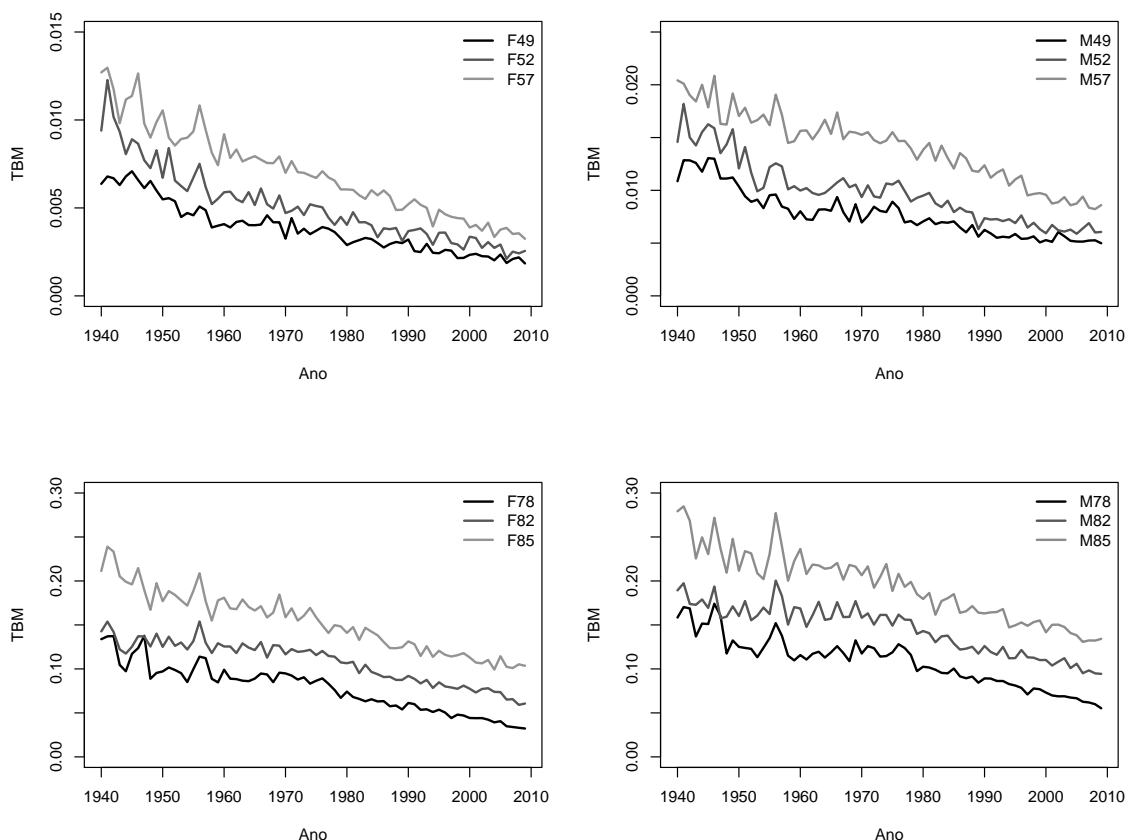


Figura 5.2: TBMs da população portuguesa, no período de 1940 a 2009, das idades 49, 52, 57, 78, 82 e 85 de ambos os sexos (respetivamente, F49, F52 e F57, para o sexo feminino, em cima e à esquerda, e M49, M52 e M57, para o sexo masculino, em cima e à direita; idem, em baixo, para as idades 78, 82 e 85)

usar grupos etários quinquenais (concretamente, os grupos etários 0-4,5-9,...,95-99), continuaremos a usar a designação simplificada idade, mas será $K = 40$ (a idade 1 corresponde ao grupo etário 0-4 do sexo feminino, a idade 2 ao grupo etário 5-9 também do sexo feminino e assim sucessivamente, a idade 21 corresponde ao grupo etário 0-4 do sexo masculino, a idade 22 ao grupo etário 5-9 também do sexo masculino e assim sucessivamente, até à idade 40 que corresponde ao grupo etário 95-99 do sexo masculino).

Seja $X_k(t)$ ($k = 1, \dots, K$) a TBM dos indivíduos de uma certa idade i ($i = 1, \dots, K/2$) e sexo j ($j = 1$ para o sexo feminino; $j = 2$ para o sexo masculino), no ano t , com $k = i + K/2(j - 1)$ para abranger todas as idades do arco da vida e ambos os sexos. Assumimos que cada $X_k(t)$ segue um MBG e designemos por $Z_k(t) = \ln(X_k(t)/X_k(t-1))$ os logaritmos dos retornos das TBMs (recorde-se que já havíamos utilizado esta mudança de variável no capítulo anterior, a propósito da estimação do MBGB ou dos modelos do tipo VARMA), tal que

$$Z_k(t) = R_k + \sigma_k(W_k(t) - W_k(t-1))$$

com R_k , σ_k e $W_k(t)$ como R , σ e W_t em (3.4) e que $\mathbf{Z}(t) = [Z_1(t), \dots, Z_k(t), \dots, Z_K(t)]'$ satisfaz

$$\mathbf{Z}(t) = \mathbf{R} + \mathbf{\Lambda}^{1/2}(\mathbf{W}(t) - \mathbf{W}(t-1)), \quad (5.1)$$

com

$$\mathbf{R} = \begin{bmatrix} R_1 \\ \vdots \\ R_K \end{bmatrix}, \quad \mathbf{\Lambda} = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_K^2 \end{bmatrix} \quad \text{e} \quad \mathbf{W}(t) = \begin{bmatrix} W_1(t) \\ \vdots \\ W_K(t) \end{bmatrix}.$$

Ao modelo descrito na equação matricial (5.1) chamamos movimento browniano geométrico multidimensional (MBGM). Note-se que o vetor \mathbf{R} representa as médias, R_k , dos $Z_k(t)$ e que, por sua vez, a matriz $\mathbf{\Lambda}$ é uma matriz diagonal com as variâncias dos $Z_k(t)$, σ_k^2 . $\mathbf{W}(t)$ é o vetor dos processos de Wiener, associados a cada um dos $Z_k(t)$, com incrementos não independentes.

Por construção, sabemos que $\mathbf{Z}(t)$ tem uma distribuição multinormal (veja-se, por exemplo, [84] acerca das propriedades genéricas e aspetos estatísticos desta distribuição), com média \mathbf{R} e variância $\mathbf{\Sigma}$, isto é, $\mathbf{Z}(t) \sim \mathcal{N}(\mathbf{R}, \mathbf{\Sigma})$ com $\mathbf{\Sigma} = \mathbf{\Lambda}^{1/2} \mathbf{S} \mathbf{\Lambda}^{1/2}$ e \mathbf{S} a matriz de correlações de $\mathbf{W}(t) - \mathbf{W}(t-1)$, dada por

$$\mathbf{S} = \begin{bmatrix} \rho_{1,1} & \cdots & \rho_{1,K} \\ \vdots & \ddots & \vdots \\ \rho_{K,1} & \cdots & \rho_{K,K} \end{bmatrix},$$

em que

$$\rho_{k_a, k_b} = \text{Corr}[W_{k_a}(t) - W_{k_a}(t-1), W_{k_b}(t) - W_{k_b}(t-1)], \quad (5.2)$$

com $k_a, k_b = 1, \dots, K$. Naturalmente, $\rho_{k_a, k_b} = 1$, se $k_a = k_b$, isto é, os elementos da diagonal da matriz \mathbf{S} são iguais a um.

Com o objetivo de reduzir o número global de parâmetros (pois, de outro modo, o modelo ficaria sobreparametrizado relativamente aos dados disponíveis, nomeadamente se tivéssemos uma média e um desvio padrão autónomos para cada $k \in K$ e coeficientes de correlação para cada par de valores $k_a, k_b \in K$, com $k_a \neq k_b$) e abranger, num único modelo, todas as idades e ambos os sexos, vamos representar os R_k , os σ_k e os ρ_{k_a, k_b} como funções da(s) idade(s).

Supomos que os R_k são funções polinomiais da idade i , tal que

$$R_k = \alpha_{0,j} i^0 + \alpha_{1,j} i^1 + \cdots + \alpha_{m-1,j} i^{m-1} + \alpha_{m,j} i^m,$$

em que $m \in \mathbb{N}^+$ e $\alpha_{0,j}, \alpha_{1,j}, \dots, \alpha_{m-1,j}, \alpha_{m,j}$ ($j = 1, 2$) são os coeficientes do polinómio R_k .

Procedemos analogamente para modelar a variância amostral dos logaritmos dos retornos das TBMs. Simplesmente, em vez de aplicar a função polinomial aos σ_k^2 usamos uma transformação que permita uma aproximação à distribuição normal. Das várias opções disponíveis na literatura para normalizar a variância, optámos pela transformação logarítmica, como sugerido, por exemplo, em [15]. Podemos então representar os $\ln(\sigma_k^2)$ como funções polinomiais da idade i , tal que

$$\ln(\sigma_k^2) = \beta_{0,j} i^0 + \beta_{1,j} i^1 + \cdots + \beta_{m-1,j} i^{m-1} + \beta_{m,j} i^m,$$

com $\beta_{0,j}, \beta_{1,j}, \dots, \beta_{m-1,j}, \beta_{m,j}$ os coeficientes do polinómio $\ln(\sigma_k^2)$.

Podemos também escrever as correlações em (5.2) em função da idade (para cada sexo e entre sexos). À semelhança do que fizemos para os σ_k podemos também normalizar a variável. Para as

correlações, usamos a transformação de Fisher (como sugerido, por exemplo, em [82]), e passamos a modelar a variável transformada $z_{\rho_{k_a, k_b}} = \frac{1}{2} \ln \frac{1+\rho_{k_a, k_b}}{1-\rho_{k_a, k_b}}$, que tem uma distribuição aproximadamente normal. Assim, para cada par (k_a, k_b) , com $k_a \neq k_b$, podemos ter

$$z_{\rho_{k_a, k_b}} = f((k_a, k_b); p^*), \quad (5.3)$$

com f função de uma variável explicativa, neste caso, a idade aqui representada pelo par (k_a, k_b) , e que resulta da combinação, que pode ser não linear, dos parâmetros do modelo, p^* . Neste ponto, não vamos assumir um número de parâmetros fixo para o modelo de correlações, já que estes poderão ser diferentes consoante o sexo e/ou entre sexos. Analogamente, não assumimos uma única forma genérica para f , sendo que podem ser, entre outras, uma função exponencial, uma potência, ou uma função polinomial da idade, de modo a estimar a correlação para todas as combinações de pares de idades possíveis, tal que podemos ter

$$z_{\rho_{k_a, k_b}} = f(j_a, j_b, \min(i_a, i_b), \text{média}(i_a, i_b), \text{abs}(i_a - i_b) | \text{parâmetros})$$

com \min , média e abs com os significados usuais. Em termos práticos, como ilustraremos nas subsecções seguintes, serão testados vários modelos.

Esta técnica tem por objetivo apenas perceber qual a melhor forma de introduzir as correlações no modelo global, que designamos por MBGM, tendo em vista a utilização do menor número de parâmetros possível.

Na sequência do exposto, a matriz S pode escrever-se com recurso a quatro blocos, cada um dos quais com dimensão $K/2 \times K/2$, tal que

$$\begin{array}{c} 1 \\ \vdots \\ K/2 \\ \hline (K/2) + 1 \\ \vdots \\ K \end{array} \left[\begin{array}{c|c} 1 \dots K/2 & (K/2)+1 \dots K \\ \hline FF & FM \\ \hline MF & MM \end{array} \right]. \quad (5.4)$$

Os blocos da diagonal de (5.4) correspondem às correlações entre idades diferentes, mas dentro do mesmo sexo, respetivamente entre idades do sexo feminino (FF) e entre idades do sexo masculino (MM). Fora da diagonal, temos as correlações entre idades e sexos diferentes (FM e o seu transposto, que representamos por MF).

5.2.2 Estimação

Sejam $t_n = t_0 + n$ ($n = 0, 1, 2, \dots, N$) os anos em que as TBMs foram observadas. Quando a matriz simétrica de covariâncias Σ é definida positiva, a f.d.p. de transição de $Z(t)$ entre t_{n-1} e t_n existe, é normal e não depende de n . Dada a independência entre os $Z(t_n)$ para diferentes valores de n , a referida f.d.p. pode escrever-se como

$$f(z_n | Z(t_{n-1}) = z_{n-1}) = f(z_n) = \frac{1}{(2\pi)^{K/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (z_n - \mathbf{R})' \Sigma^{-1} (z_n - \mathbf{R}) \right\}, \quad (5.5)$$

com Σ e \mathbf{R} como exposto anteriormente, $|\Sigma|$ o determinante de Σ e $(z_n - \mathbf{R})'$ o transposto de $(z_n - \mathbf{R})$.

Seja \mathbf{p} o vetor de parâmetros do modelo MBGM. Para ilustrar a estimação do mesmo, consideramos que o vetor \mathbf{R} se divide em duas partes (a primeira metade diz respeito às esperanças matemáticas dos logaritmos dos retornos da TBMs das idades do sexo feminino e a segunda parte das idades do masculino) e que cada R_k pode ser descrito por um polinómio de grau três na idade i . Procedemos de forma análoga relativamente à diagonal da matriz Λ , supondo que os logaritmos dos σ_k^2 são também polinómios de grau 3 na idade i . Temos até agora um vetor de 16 parâmetros. Quanto às correlações, consideremos que temos parâmetros diferentes por sexo e entre sexos e vamos admitir (apenas para ilustrar a configuração de um possível vetor de parâmetros do modelo) que $z_{\rho_{k_a, k_b}} = \phi_{j_a, j_b} \min(i_a, i_b) + \gamma_{j_a, j_b} \text{abs}(i_a - i_b)$. Neste caso concreto, temos

$$\begin{aligned} \mathbf{p} = & (\alpha_{0,1}, \alpha_{1,1}, \alpha_{2,1}, \alpha_{3,1}, \\ & \alpha_{0,2}, \alpha_{1,2}, \alpha_{2,2}, \alpha_{3,2}, \\ & \beta_{0,1}, \beta_{1,1}, \beta_{2,1}, \beta_{3,1}, \\ & \beta_{0,2}, \beta_{1,2}, \beta_{2,2}, \beta_{3,2}, \\ & \phi_{1,1}, \gamma_{1,1}, \\ & \phi_{2,2}, \gamma_{2,2}, \\ & \phi_{1,2}, \gamma_{1,2}). \end{aligned} \tag{5.6}$$

A partir de (5.5), podemos escrever a função de log-verosimilhança, \mathcal{L} , dos logaritmos dos retornos das TBMs, para indivíduos de qualquer idade e sexo, que é dada por

$$\mathcal{L}(\mathbf{p} | \mathbf{Z}(t_1), \dots, \mathbf{Z}(t_n)) = -\frac{K}{2} N \ln(2\pi) - \frac{1}{2} N \ln |\Sigma| - \frac{1}{2} \sum_{n=1}^N ((\mathbf{Z}(t_n) - \mathbf{R})' \Sigma^{-1} (\mathbf{Z}(t_n) - \mathbf{R})). \tag{5.7}$$

O estimador, $\hat{\mathbf{p}}$, de \mathbf{p} é obtido por MV. À semelhança dos modelos anteriores, as estimativas dos parâmetros foram obtidas por métodos numéricos através da minimização do simétrico da função de log-verosimilhança (neste caso, (5.7)). Normalmente tem-se um método iterativo que requer o uso de valores de arranque para os parâmetros de \mathbf{p} . Esses valores são estimados pelo método dos mínimos quadrados (para o efeito, usamos a função `nls` do R).

Para obter os intervalos de confiança para os parâmetros, assumimos que estamos em regime assintótico, considerando as propriedades de estimação por MV, e fazemos uma aproximação da matriz de informação de Fisher pela simétrica da inversa da matriz hessiana, de cuja diagonal obtemos as respetivas variâncias assintóticas (tal como descrito, nos capítulos anteriores, para os modelos unidimensionais e bidimensionais de EDEs). Para cada parâmetro, obtemos assim uma aproximação dos limites de um intervalo de confiança a $(1 - \alpha) \times 100\%$ através de

$$\hat{p} \pm z_{1-\alpha/2} (\widehat{Var}[\hat{p}])^{1/2},$$

em que $\widehat{Var}[\hat{p}]$ representa a estimativa da variância.

Se tivermos observações até um certo instante t_N e quisermos obter previsões para $t > t_N$, considerando que cada $Z_k(t)$ é um processo de Markov, temos que

$$E[\mathbf{Z}(t)|\mathbf{Z}(t_1), \dots, \mathbf{Z}(t_N)] = E[\mathbf{Z}(t)|\mathbf{Z}(t_N)] = E[\mathbf{Z}(t)] = \mathbf{R}.$$

Podemos usar para previsões a LP, para qualquer idade k ,

$$\hat{Z}_k(t) = \hat{E}[Z_k(t)|Z_1(t_N) = z_{1,N}, \dots, Z_K(t_N) = z_{K,N}] = \hat{R}_k, \quad (5.8)$$

onde \hat{E} representa o valor aproximado da esperança matemática, substituindo os valores exatos de R pelo do seu estimador de MV, \hat{R} . Se quisermos obter as previsões na escala original, vem

$$\hat{X}_k(t) = X_k(t_N) \exp\{\hat{Z}_k(t_{N+1}) + \dots + \hat{Z}_k(t)\} = X_k(t_N) \exp\{\hat{R}_k(t - t_N)\}.$$

Em termos de critérios de implementação do modelo, pretendemos um modelo com, obviamente, o menor número de parâmetros possível, para ajustar aos dados da mortalidade da população portuguesa e que forneça “boas” previsões. Neste caso, referimo-nos a um modelo que permita estimar taxas realísticas (no caso das TBMs, na escala original dos dados, estas terão que ser não negativas) e com possibilidade de fazer previsões, considerando as correlações entre diferentes idades de ambos os sexos, dentro e entre sexos.

A propósito de correlações, refira-se de novo, tal como já chamámos a atenção aquando da apresentação e implementação do modelo de Gompertz estocástico bidimensional (MGEB), que estas não representam causalidade. O que se pretende é, através de um conjunto de medições sobre os dados, verificar se efetivamente há uma relação entre as séries temporais das TBMs que os compõem e tirar partido dessa informação, incorporando-a num modelo para todas as idades e ambos os sexos, em alternativa a modelar cada série individualmente. Para o efeito, temos que usar várias dimensões e necessariamente introduzir em simultâneo todas as idades nesse modelo único e global. Desta maneira, aumentamos a dimensão e a complexidade do problema, já que é necessário incorporar no modelo uma estrutura de correlação, que, por sua vez, não é fácil de captar.

A computação do modelo não é, pelo exposto, trivial. Desde logo porque exige algoritmos adequados onde é necessário incorporar toda a informação das funções que descrevem as várias componentes matriciais do mesmo. Para clarificar estes procedimentos, na subsecção seguinte (e antes de apresentarmos os resultados), referimos alguns aspetos práticos que se prendem com a sua implementação.

5.2.3 Aspetos práticos relacionados com a implementação do modelo

Nesta fase, é necessário decidir que método numérico utilizar para maximizar a função de log-verosimilhança em (5.7), ou, equivalentemente, minimizar o seu simétrico (nossa opção) e de que forma o implementar. Como há variadas abordagens possíveis a este processo, considerando que temos um elevado número de parâmetros a estimar e muitas dimensões envolvidas, e conforme sugerido na bibliografia de referência nesta área (ver, por exemplo, [49, 67]), decidimos ser conveniente começar por usar, na otimização da log-verosimilhança, um método de busca da

direcção de descida, neste caso, do simétrico da função de log-verosimilhança, que recorra ao gradiente, $\nabla(p)$.

À partida, temos duas opções: ou construímos o nosso próprio método de otimização, ou usamos, caso exista e se adequa ao problema, uma função pré-definida de um programa de computação estatística para o efeito. Neste sentido, numa fase inicial, desenvolvemos um método iterativo, baseando-nos nos algoritmos de descida, de passo simples, com recurso ao gradiente (mostramos a sua conceção, em linhas gerais, no início do apêndice E). Uma vez que vamos utilizar um algoritmo de otimização e como tal precisamos do conhecimento de derivadas, dada a complexidade de obter as expressões explícitas das derivadas parciais, recorreremos a razões incrementais, simples de computar, como uma aproximação.

Mais uma vez, considerando o número de parâmetros a estimar (pode ser superior a 20) e o número de dimensões envolvidas (temos 200 v.a.), mas também por a própria tolerância e o passo da descida terem que ser muito pequenos (por causa das ordens de grandeza, quer dos dados, quer dos valores de arranque), o cálculo revelou-se lento e ocorreram alguns entraves à convergência. Constatámos que muitas limitações observáveis durante os testes a esse algoritmo são devidas aos erros de arredondamento inerentes ao próprio processo iterativo e à complexidade de operações no cálculo, de entre as quais destacamos o cálculo da matriz inversa de Σ e, por associação, o logaritmo do determinante desta. Não obstante não ter sido possível obter uma solução do problema por esta via, consideramos que esta etapa forneceu informações valiosas para estudar outras abordagens para o efeito.

Das experiências com este método, destaca-se o facto de ter sido possível analisar, inclusive por iterada, componentes problemáticas da função a otimizar, experimentando transformações nos parâmetros e mesmo testando até o método com outros conjuntos de dados e dados simulados. Naturalmente que, a certa altura, esse conhecimento fez-nos equacionar a necessidade de uma eventual migração para um outro método mais complexo e/ou com outra forma de implementação mais eficiente.

Explorámos então várias possibilidades disponíveis em pacotes de programas estatísticos, sendo que os resultados produzidos e apresentados foram obtidos utilizando a função `optim` do R (veja-se o código no apêndice E). Testámos resultados parciais noutros programas, como o *Mathematica*, por exemplo, mas, como eram semelhantes, optámos por manter a computação no R, pois consideramos ser suficientemente manipulável e oferecer um vasto conjunto de opções para tentar resolver o problema em causa. Em toda a computação, usamos a simplificação $t_n - t_{n-1} = 1$ (já referida nos capítulos anteriores).

Independentemente das ferramentas disponíveis, a computação de modelos estatísticos está sujeita a inúmeros problemas. Por esse motivo, vamos referir brevemente algumas das opções que tomámos para tornar possível a modelação. De seguida apresentamos uma síntese das seis etapas que consideramos necessárias à implementação do modelo. No diagrama que constitui a figura 5.3 ilustramos essas etapas, que passamos a descrever, nomeadamente através dos resultados obtidos.

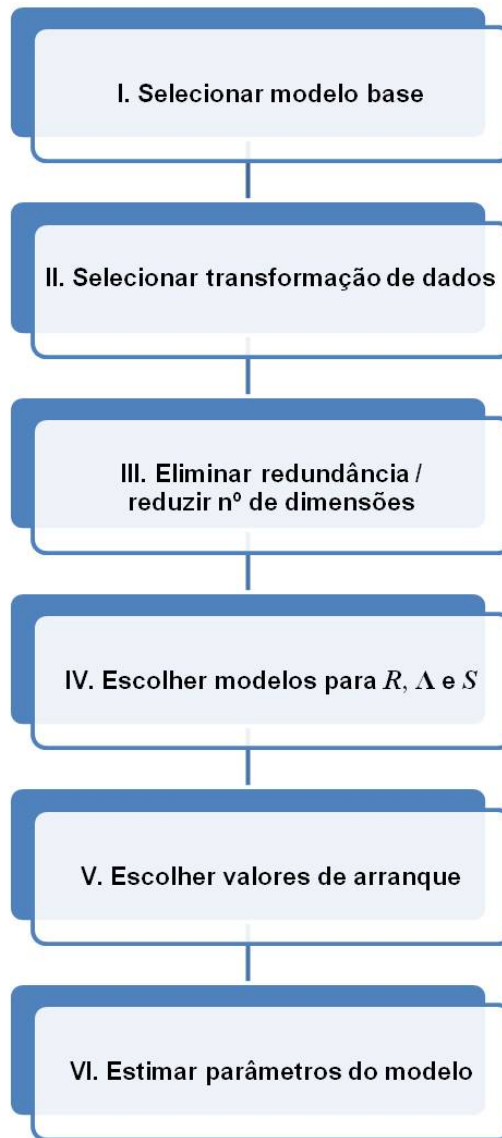


Figura 5.3: Etapas da implementação do MBGM

Em primeiro lugar, tivemos que decidir qual a base para este modelo multidimensional. A opção foi simples: baseámo-nos nos resultados dos modelos unidimensionais MBG e MGE, cuja aplicação às TBMs da população portuguesa mostrámos no capítulo 3, e escolhemos o MBG. Desde logo porque, por construção, é um modelo simples, pelo número de parâmetros e pelas propriedades da sua distribuição. Recordamos também que embora não tivéssemos constatado que um dos dois modelos apresenta vantagens significativas relativamente ao outro, tomando como critério de desempenho o EQM, e perante os resultados obtidos no comparativo dos modelos uni e bidimensionais, o MBG apresenta-se globalmente (para todas as idades) mais favorável, portanto com erros menores quer nos ajustamentos quer nas previsões, resultantes da aplicação a este tipo de dados, que o MGE.

Passamos à segunda etapa de implementação. Selecionada uma base para o modelo, decidimos modelar as TBMs em escala logarítmica (à semelhança do que fizemos nos modelos unidimensionais), pois facilita a construção e análise do modelo. No MBG, havíamos utilizado

a transformação $Y(t) = \ln(X(t)/x_0)$, com $X(t)$ a TBM dos indivíduos de uma certa idade i ($i = 1, \dots, K/2$) e sexo j ($j = 1$, para o feminino; $j = 2$, para o masculino). Agora, fazemos $Z(t) = \ln(X(t)/X(t-1))$ por comodidade de tratamento. Note-se que $X(t)$ é que segue um MBGM, sendo o uso de $Z(t)$ facilitador de cálculos.

Antes de avançarmos, vamos referir-nos à questão do elevado número de dimensões, que influencia grandemente o cálculo matricial. No primeiro teste ao modelo, através do método iterativo que desenvolvemos (e conforme já referido) em todas as etapas descritas até agora, bem como nas seguintes, foram utilizadas todas as séries das TBMs correspondentes aos grupos etários anuais de ambos os sexos (isto é, $K = 200$).

Analisando a matriz de correlações empíricas, à partida, podemos questionarmo-nos acerca de questões de multicolinearidade com correlações quase perfeitas, nomeadamente em idades próximas, o que poderia sugerir a aplicação, por exemplo, de uma técnica de análise de componentes principais com a extração de fatores, e proceder à modelação com base nesses resultados. Esta opção, embora pertinente, não é viável no contexto do problema em estudo, pois o único modo de abarcar todas as idades do arco de vida num modelo, como é nosso objetivo, é usar uma série para cada grupo de idades o que, neste caso não nos fornece dimensões extra para extrair os fatores.

Assim, para eliminar alguma redundância e, simultaneamente, reduzir o número de dimensões envolvidas no cálculo, sem perder as especificidades dos dados que vamos modelar e que ilustram o problema, optámos por usar (e obter os resultados para) as séries anuais completas (de 1940 a 2009, deixando, como habitualmente, as últimas 10 “observações” para estimar o EQM das previsões), mas agora cada idade refere-se não ao grupo etário anual, mas a grupos quinquenais. Portanto, neste capítulo, cada “idade” representa 5 anos e passamos a modelar, simultaneamente, em vez de 200, 40 séries temporais (isto é, $K = 40$), o que se traduz em enormes vantagens em termos de tempo e eficiência da computação.

As TBMs anuais para grupos de idade quinquenais foram obtidas através do *Human Mortality Database* [37]. No caso do grupo 0-4, como os dados estavam separados em dois grupos, o da idade 0 e o das idades 1 a 4, foi necessário usar os dados do número de óbitos e da população exposta ao risco de óbito (também disponíveis na mesma fonte) para obter a respetiva TBM.

Explicado que está o facto de mostrarmos os resultados para apenas 20 idades de cada sexo, está também terminada a terceira fase de implementação do modelo.

Na fase seguinte, é necessário reunir a informação que temos sobre os parâmetros a estimar, e decidir como “desenhar” os componentes vetoriais e matriciais do modelo, designadamente R , Λ e S . Entramos na quarta etapa do diagrama que constitui a figura 5.3. Esta etapa, bem como as restantes, continuarão a ser descritas, em simultâneo com a apresentação dos resultados da aplicação do modelo.

5.2.4 Resultados do ajustamento

Continuamos a descrever as etapas que constam no diagrama da figura 5.3. Refira-se, que a partir deste ponto, e de acordo com o exposto no final da subsecção anterior (quanto à necessidade de reduzir o número de dimensões envolvidas no cálculo mas continuando a abranger todas as idades do arco de vida), que todos os resultados são relativos às idades 1 a 20, que representam os grupos etários quinquenais: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94 e 95-99, de cada sexo. Nos vetores e matrizes do modelo, as idades são representadas pelo índice $k = 1, \dots, K$ quando, na mesma estrutura, guardamos consecutivamente as 20 idades do sexo feminino seguidas pelas do sexo masculino. Sempre que considerarmos operações ou relações dentro de um mesmo sexo, as idades, os 20 grupos selecionados, vêm representadas pelo índice $i = 1, \dots, K/2$. Note-se que $k = i + 20(j - 1)$, com $i = 1, \dots, 20$ e $j = 1, 2$ (1 para o sexo feminino e 2 para o masculino).

Proseguimos com a descrição da implementação do MBGM. Concretamente, e como já referimos, começamos por equacionar, de que forma desenhar R , Λ e S . Assim, com o objetivo de construir um modelo com menos parâmetros que considere o padrão de variação dos R_k e dos $\ln(\sigma_k^2)$ com a idade, vamos usar modelos polinomiais, para cada sexo, da forma descrita na subsecção 5.2.1. Analogamente, pretendemos também escrever as correlações z_ρ como função da idade.

Como já referimos, os parâmetros são estimados por MV mas como a função a otimizar, (5.7), é uma função com muitas variáveis precisamos de um método iterativo, que por sua vez necessita de valores de arranque, valores de arranque esses que vamos obter através de regras. Para o efeito, determinámos estimativas de R_k , $\ln \sigma_k^2$ e z_ρ e ajustámos as funções referidas por RNL (tal como na secção 3.4, usámos a função `nls` do R). Por sua vez, para obter essas estimativas usámos os logaritmos dos retornos dos dados, cujas médias empíricas, para cada idade k , dão os R_k , cujas variâncias empíricas, dão os σ_k^2 e cujas correlações empíricas dão os ρ_{k_a, k_b} . Daqui obtemos as estimativas de $\ln(\sigma_k^2)$ e $z_{\rho_{k_a, k_b}} = \frac{1}{2} \ln \frac{1 + \rho_{k_a, k_b}}{1 - \rho_{k_a, k_b}}$.

A técnica de regressão não linear (conforme descrito na secção 3.4), revelou-se adequada e foi natural o ajustamento de funções polinomiais às séries das médias e dos logaritmos das variâncias empíricas dos logaritmos dos retornos das TBMs de todas as idades, por sexo. Desses ajustamentos (ver figura 5.4 e caixa 5.1), resultam valores de arranque para o método iterativo.

Associado a cada gráfico da figura 5.4, mostramos a equação da curva ajustada. Na caixa 5.1, veja-se, em particular, os resultados do ajustamento para as médias estimadas para o sexo masculino, que se mostrou adequado com todos os parâmetros significativamente diferentes de zero para os níveis de significância usuais. Este procedimento foi repetido para o sexo feminino e, analogamente, para os logaritmos das variâncias de ambos os sexos com resultados também favoráveis à inclusão no modelo global (na expressão do ajustamento, r_{0M}, \dots, r_{3M} tem o mesmo significado que $\hat{\alpha}_{0,2}, \dots, \hat{\alpha}_{3,2}$ no vetor (5.6)).

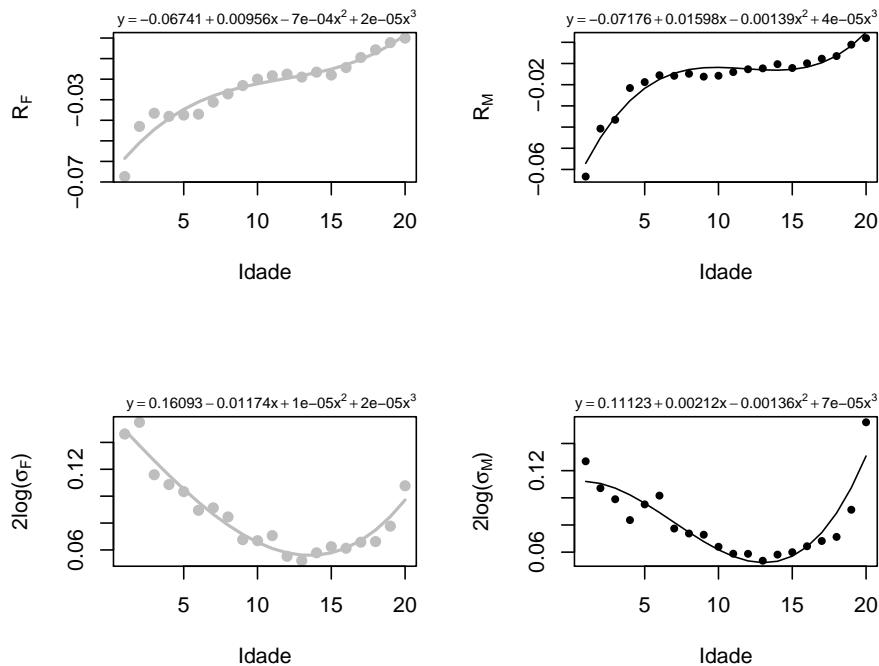


Figura 5.4: Ajustamento de funções polinomiais às médias e logaritmos das variâncias empíricas dos logaritmos dos retornos das TBMs de todas as idades e de ambos os sexos (em cima e à esquerda, as médias para o sexo feminino e, à direita, para o masculino; idem, em baixo, para os logaritmos das variâncias)

Caixa 5.1 Ajustamento de uma função polinomial às médias empíricas dos logaritmos dos retornos das TBMs de todas as idades do sexo masculino

```
>library(stats)
> ajust.Rm <- nls(Rm ~ r0M + r1M*Idade + r2M*Idade^2 + r3M*Idade^3, start=list(r0M=b[1],r1M=b[2],r2M=b[3],r3M=b[4]))
> summary(ajust.Rm)

Formula: Rm ~ r0M + r1M * Idade + r2M * Idade^2 + r3M * Idade^3
Parameters:
      Estimate Std. Error t value Pr(>|t|)
r0M -7.176e-02  3.596e-03 -19.956 9.90e-13 ***
r1M  1.598e-02  1.447e-03  11.046 6.77e-09 ***
r2M -1.390e-03  1.581e-04  -8.793 1.59e-07 ***
r3M  3.910e-05  4.956e-06   7.891 6.63e-07 ***
---Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.003292 on 16 degrees of freedom
Number of iterations to convergence: 1
Achieved convergence tolerance: 6.707e-08
```

Abordemos agora a questão da estrutura de correlação e da construção das matrizes S e Σ , respetivamente, a matriz de correlações e a matriz de variâncias-covariâncias.

Antes de avançar com um modelo complexo para z_ρ decidimos verificar se as correlações empíricas são significativamente diferentes de zero, caso contrário, faria mais sentido trabalhar com um modelo sem correlações, com menos parâmetros. Assim, foram testadas as transformadas de Fisher das correlações, $z_{\rho_{k_a, k_b}}$ (que sabemos têm distribuição aproximadamente normal), para pares de idades mais ou menos próximas, de modo a perceber se são significativamente diferentes de zero. Para o efeito, podemos usar intervalos de confiança, para um nível de significância α , $\mathcal{IC}_{(1-\alpha)\times 100\%}$, e testar na hipótese nula $z_\rho = 0$ (para simplificar, usamos, quando conveniente, z_ρ como $z_{\rho_{k_a, k_b}}$). Os limites do \mathcal{IC} são dados por

$$\mathcal{IC}_{(1-\alpha)\times 100\%}(z_\rho) = \hat{z}_\rho \pm z_{1-\alpha/2} \frac{1}{\sqrt{N-3}},$$

onde N é o número de anos utilizados.

Neste processo, fixámos várias distâncias, aplicando o teste para idades diferentes dentro do mesmo sexo e entre sexos, separadamente. Na figura 5.5, mostramos os resultados para idades contíguas, isto é, diferença de uma unidade (5 anos) entre idades e para diferenças de quatro unidades (20 anos) entre as idades, para $\alpha = 0.05$. Na maior parte das idades, as correlações são significativas. O teste foi aplicado a pares de correlações com outras diferenças fixas entre idades e os resultados são análogos, para a generalidade das combinações possíveis, o que vem validar a necessidade de um estrutura de correlação.

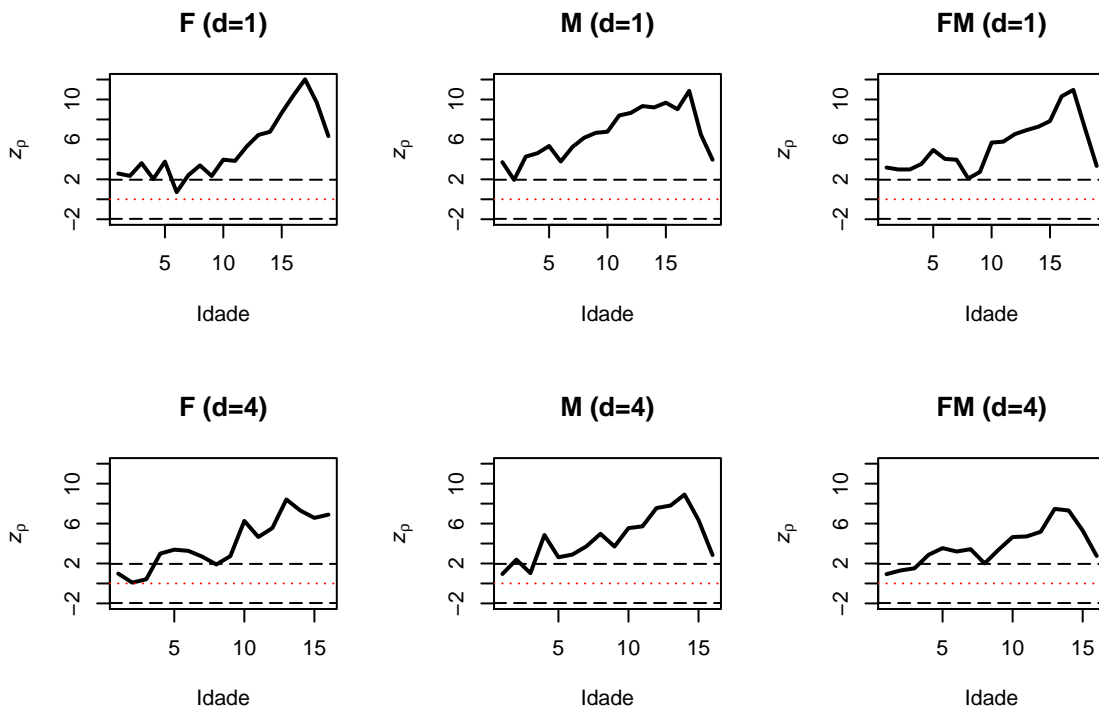


Figura 5.5: Correlações empíricas com transformação de Fisher, z_ρ , por sexo (F para feminino e M para masculino) e entre sexos (FM) (em cima, relativas a pares de idades com diferenças de 1 unidade entre si, $d=1$; em baixo, com diferenças de 4 unidades entre si, $d=4$). Indicam-se os limites aproximados da região de aceitação para z_ρ do teste da hipótese nula $\rho = 0$, baseado em z_ρ ter distribuição aproximadamente normal

Com o objetivo de estimar a matriz S (por sua vez necessária para obter Σ), começamos por observar as correlações empíricas entre as várias idades, dentro de cada sexo e entre sexos (recordamos que continuamos a trabalhar com as correlações na escala transformada, cuja distribuição é aproximadamente normal). Como à partida não temos referências para os valores dos parâmetros, como já referimos, vamos usar as correlações empíricas, para cada par de idades, e observar o seu comportamento em função da idade.

Intuitivamente, podemos supôr que há uma diminuição da correlação à medida que aumenta a diferença entre as idades ou que a correlação diminui, quando diminui uma das idades, por exemplo. Assim, desconstruímos essas matrizes de correlação e criámos um vetor com todas as correlações, ordenadas para cada par de idades, segundo os índices das colunas, para cada sexo e entre sexos (de entre as correlações para idades diferentes, mas dentro de cada sexo, excluimos as correlações para as mesmas idades e as correlações repetidas, uma vez que essas matrizes são simétricas).

Numa fase preliminar, fizemos algumas experiências com a variável z_ρ , com recurso à técnica de RNL (apenas para ter uma ideia do tipo de função que melhor se lhe ajustaria), começando por ajustar-lhe uma função média da idade e explorando graficamente diferentes escolhas de valores para os parâmetros dessa curva, o que nos dá uma medida de localização média (foram experimentadas várias versões para essa função média). Na figura 5.6, à qual está associada a caixa 5.2, representamos quatro curvas exploratórias para ver como cada uma dessas funções médias varia com diferentes valores dos parâmetros. Como observamos algum tipo de crescimento exponencial nas correlações em função da idade, usámos, por exemplo, a função $f(\text{média das idades}, (b_1, b_2, b_3)) = \exp(-b_1 \times \text{média das idades}) / (b_2 + b_3 \times \text{média das idades})$.

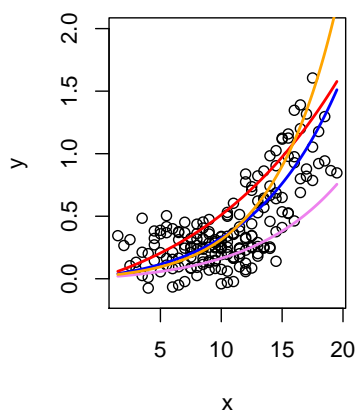


Figura 5.6: Representação simultânea da transformada de Fisher das correlações empíricas (y) e curvas do ajustamento, em função da média entre duas idades (x), consoante diferentes valores dos parâmetros, para o sexo feminino

Caixa 5.2 Ajustamento de uma função para a transformada de Fisher da correlação empírica, através de quatro curvas exploratórias de ajustamento com parâmetros diferentes, para o sexo feminino

```
library(graphics)
y <- z_ro.F
x <- Id.med #média de duas idades

ro.1 <- function(x, b1, b2, b3){exp(b1 * x) * (b2 + b3 * x)}
plot(y ~ x)
curve(ro.1(x,b1=0.1,b2=0.01,b3=0.005),add=TRUE,col="violet")
curve(ro.1(x,b1=0.1,b2=0.02,b3=0.01),add=TRUE,col="blue")
curve(ro.1(x,b1=0.05,b2=0.01,b3=0.03),add=TRUE,col="red")
curve(ro.1(x,b1=0.15,b2=0.02,b3=0.005),add=TRUE,col="orange")
```

Posteriormente, passamos ainda pela representação tridimensional das correlações empíricas, também com recurso a funções da idade (ver caixa 5.3, à qual está associada a figura 5.7, em que se representam as correlações em função do módulo da diferença entre as idades e da idade mínima), de modo a validar algumas das possíveis escolhas para incluir na estrutura de correlação do modelo.

Caixa 5.3 Representação tridimensional da transformada de Fisher das correlações empíricas, com plano de regressão, para o sexo feminino

```
x <- abs.Id #módulo da diferença entre duas idades
y <- Id.min #mínimo de duas idades
z <- z_ro.F

library(rgl)
plot3d(x, y, z, type = "s", col = "red")
ajust <- lm(z ~ x + y)
coefs <- coef(ajust)

#a, b, c = coordenadas do plano
a <- coefs["x"]
b <- coefs["y"]
c <- -1
d <- coefs["(Intercept)"]

planes3d(a, b, c, d, col = "blue", alpha = c(0.5))
```

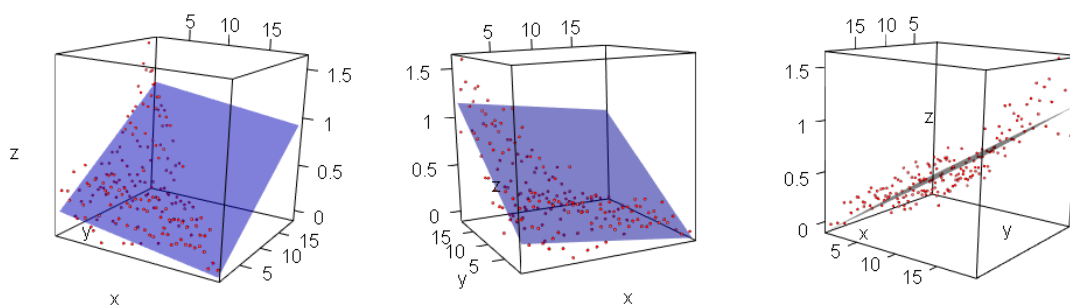


Figura 5.7: Representação tridimensional da transformada de Fisher das correlações empíricas com diferentes ângulos de rotação em torno do eixo dos z e plano de regressão, para o sexo feminino (x,y e z definidos na caixa 5.3)

Com base na informação obtida nesta fase exploratória, e de modo a encontrar uma expressão que explique as correlações, para incorporar no modelo global, representámo-las indo ao encontro da construção da matriz S : entre idades diferentes do sexo feminino, entre idades diferentes do sexo masculino e entre idades diferentes de sexos diferentes. Essa representação foi feita para as transformadas de Fisher das correlações empíricas, em função da primeira idade, i_a (que coincide

com a idade mínima entre duas idades, $\min(i_a, i_b)$, no caso das correlações entre idades diferentes de um mesmo sexo), da segunda idade, i_b (que coincide com a idade máxima entre duas idades, $\max(i_a, i_b)$, no caso das correlações entre idades diferentes de um mesmo sexo), do módulo da diferença entre duas idades, $\text{abs}(i_a - i_b)$, ou da média de duas idades, $\text{média}(i_a, i_b)$. Nesta fase, seleccionámos 12 modelos de RNL, para o efeito, que enumeramos de seguida.

Para $k_a < k_b$ temos:

$$\begin{aligned}
 \text{Modelo 1. } z_{\rho_{k_a, k_b}} &= \phi_{j_a, j_b} i_a \exp\{\gamma_{j_a, j_b} \text{média}(i_a, i_b)\} \\
 \text{Modelo 2. } z_{\rho_{k_a, k_b}} &= \phi_{j_a, j_b} i_a \exp\{\psi_{j_a, j_b} \text{abs}(i_a - i_b)\} \\
 \text{Modelo 3. } z_{\rho_{k_a, k_b}} &= \phi_{j_a, j_b} i_a^2 + \psi_{j_a, j_b} \text{abs}(i_a - i_b) \\
 \text{Modelo 4. } z_{\rho_{k_a, k_b}} &= \phi_{j_a, j_b} i_a \\
 \text{Modelo 5. } z_{\rho_{k_a, k_b}} &= \phi_{j_a, j_b} i_a + \psi_{j_a, j_b} \text{abs}(i_a - i_b) \\
 \text{Modelo 6. } z_{\rho_{k_a, k_b}} &= \eta_{j_a, j_b} \exp\{\phi_{j_a, j_b} i_a\} \\
 \text{Modelo 7. } z_{\rho_{k_a, k_b}} &= \phi_{j_a, j_b} i_a + \gamma_{j_a, j_b} \text{média}(i_a, i_b)^2 \\
 \text{Modelo 8. } z_{\rho_{k_a, k_b}} &= \exp\{\phi_{j_a, j_b} i_a + \psi_{j_a, j_b} \text{abs}(i_a - i_b)\} \\
 \text{Modelo 9. } z_{\rho_{k_a, k_b}} &= \phi_{j_a, j_b} i_a^2 \\
 \text{Modelo 10. } z_{\rho_{k_a, k_b}} &= \phi_{j_a, j_b} i_b + \gamma_{j_a, j_b} \text{média}(i_a, i_b) \\
 \text{Modelo 11. } z_{\rho_{k_a, k_b}} &= \phi_{j_a, j_b} i_a^2 + \gamma_{j_a, j_b} \text{média}(i_a, i_b) \\
 \text{Modelo 12. } z_{\rho_{k_a, k_b}} &= \phi_{j_a, j_b} i_a + \gamma_{j_a, j_b} \text{média}(i_a, i_b)
 \end{aligned}
 \tag{5.9}$$

Para $k_a > k_b$, nos modelos acima, põe-se $z_{\rho_{k_a, k_b}} = z_{\rho_{k_b, k_a}}$.

Nas figuras 5.8 a 5.10, mostramos a sobreposição entre as correlações empíricas e os valores ajustados por cada um desses modelos, para todas as combinações possíveis de pares de idades, entre os logaritmos dos retornos das TBMs por sexo e entre sexos. Em todos estes ajustamentos, as correlações foram expressas através das mesmas funções da idade, tanto para as correlações dentro de cada sexo, como entre sexos, numa tentativa de estes resultados sugerirem uma única expressão global para estimar a matriz de correlações do modelo, S .

Graficamente, é visível um ajustamento satisfatório para a generalidade dos modelos, no caso do sexo feminino, sendo que os modelos seleccionados apresentam, neste caso, o melhor desempenho. Se, para o sexo masculino, em alguns casos (como para o modelo 1 ou 7), é notório um ajustamento menos bom, em parte porque as correlações apresentam uma maior dispersão, quando observamos os ajustamentos para as correlações entre sexos, como se altera o seu padrão, são vários os modelos que se apresentam pouco flexíveis para captar a respetiva variabilidade. Ainda assim, estas são, de um conjunto alargado de opções analisadas, as que apresentam melhores resultados.

Na caixa 5.4, resume-se a informação relativa ao desempenho desses modelos, para as transformadas de Fisher das correlações empíricas, tendo em conta os critérios do EQM e o do

AIC. De todos, o modelo 2 é o melhor.

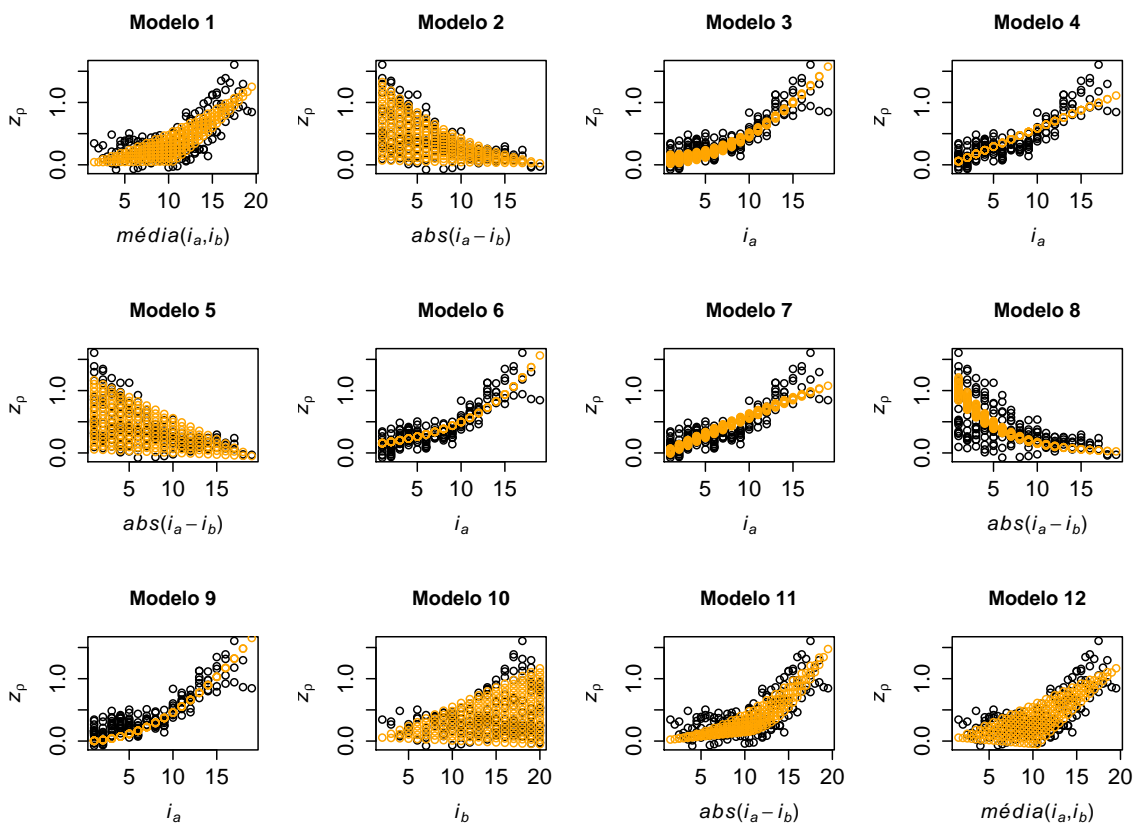


Figura 5.8: Representação conjunta da transformada de Fisher das correlações empíricas (a preto) e do ajustamento em função da idade (a amarelo), através de métodos de RNL, para o sexo feminino

Caixa 5.4 MBGM: EQM e AIC do ajustamento dos modelos por RNL para a transformada de Fisher das correlações empíricas, que ilustrámos nas figuras 5.8 a 5.10

```
> tab.resumo <- cbind(modelos,criter.F,criter.M,criter.FM)
> print(tab.resumo,quote=FALSE,digits=5)
```

Modelo	n°	EQM.F	AIC.F	EQM.M	AIC.M	EQM.FM	AIC.FM
[1,]	1	0.0265373	-144.3524	0.0488468	-28.42609	0.0973021	209.1769
[2,]	2	0.0212497	-186.5715	0.0438372	-48.98513	0.0440893	-107.4646
[3,]	3	0.0290492	-127.1687	0.0879532	83.31609	0.1452804	369.5150
[4,]	4	0.0277763	-137.6823	0.0589289	5.22617	0.1075342	247.1723
[5,]	5	0.0260544	-147.8416	0.0572630	1.77751	0.0958263	203.0635
[6,]	6	0.0241531	-162.2382	0.0622968	17.78609	0.0981202	212.5260
[7,]	7	0.0269875	-141.1558	0.0379927	-76.17202	0.0736705	97.8896
[8,]	8	0.0696020	38.8537	0.0484435	-30.00102	0.0713218	84.9298
[9,]	9	0.0334210	-102.5320	0.0949271	95.81389	0.1456645	368.5712
[10,]	10	0.0260544	-147.8416	0.0572630	1.77751	0.0733935	96.3831
[11,]	11	0.0262916	-146.1196	0.0761230	55.86967	0.0721667	89.6404
[12,]	12	0.0260544	-147.8416	0.0572630	1.77750	0.0733935	96.3831

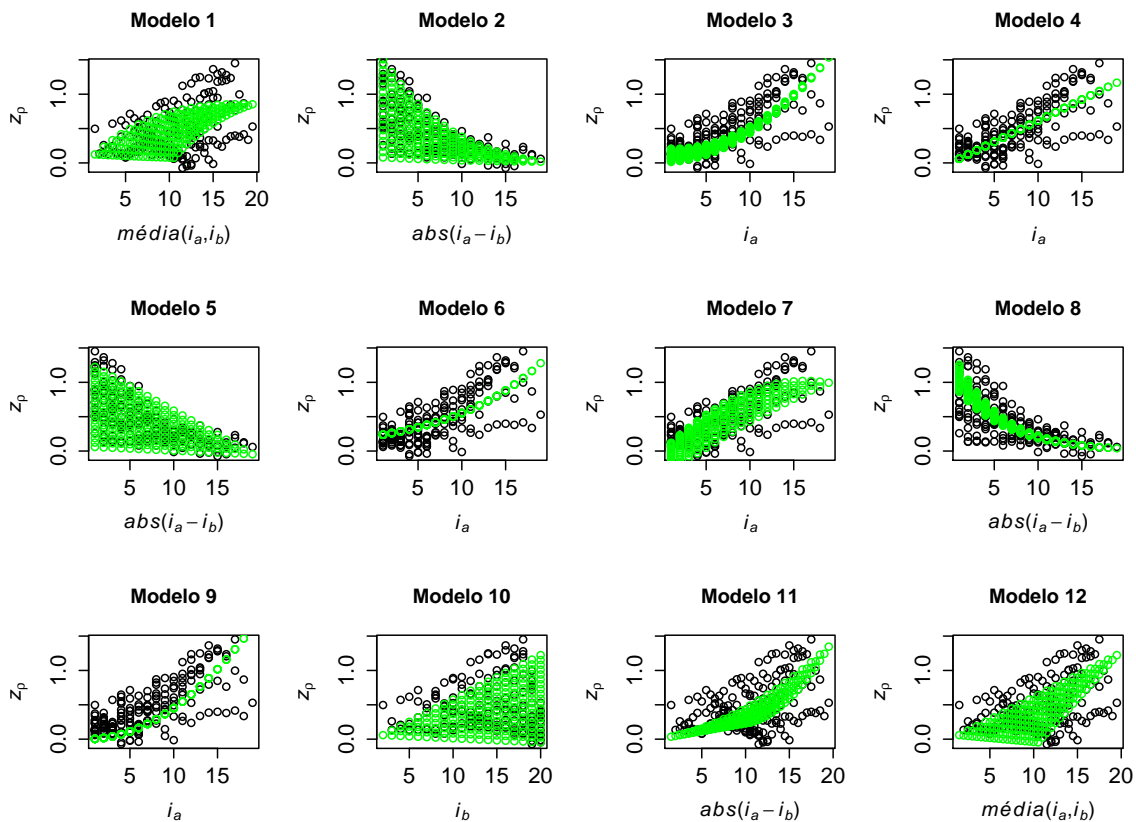


Figura 5.9: Representação conjunta da transformada de Fisher das correlações empíricas (a preto) e do ajustamento em função da idade (a verde), através de métodos de RNL, para o sexo masculino

Preparado que está o vetor dos valores iniciais (ou de arranque) da função a otimizar (5.7), encontramos-nos no último passo do diagrama da figura 5.3 e estamos em condições de aplicar o método iterativo para estimar, por MV, os parâmetros do modelo, que designámos por MBGM. No entanto, antes de apresentar os resultados finais da aplicação do mesmo, resta-nos tecer ainda algumas considerações acerca da escrita da própria função de log-verosimilhança e chamar brevemente a atenção para algumas das abordagens possíveis para o fazer, uma vez que a forma como a função é introduzida no algoritmo pode condicionar a convergência do método em alguns casos particulares (que não são pouco frequentes quando se trabalha com dados reais).

Como já referido, a estimação de parâmetros por MV de um processo que segue uma distribuição multinormal, envolve o cálculo do determinante de uma matriz, neste caso, a matriz de variâncias-covariâncias, que designámos por Σ , e da sua inversa. Esta matriz deve gozar de algumas propriedades que, pelo menos no arranque do método iterativo, devem ser asseguradas: deve ser simétrica e não singular, pois caso contrário não é possível obter a sua inversa, bem como ser definida positiva.

Na literatura sobre otimização, são frequentemente descritos problemas diversos relacionados com a computação de matrizes, uma vez que, mesmo que no início do método iterativo estejam asseguradas as condições que tornem possível o cálculo, não podemos absolutamente garantir que

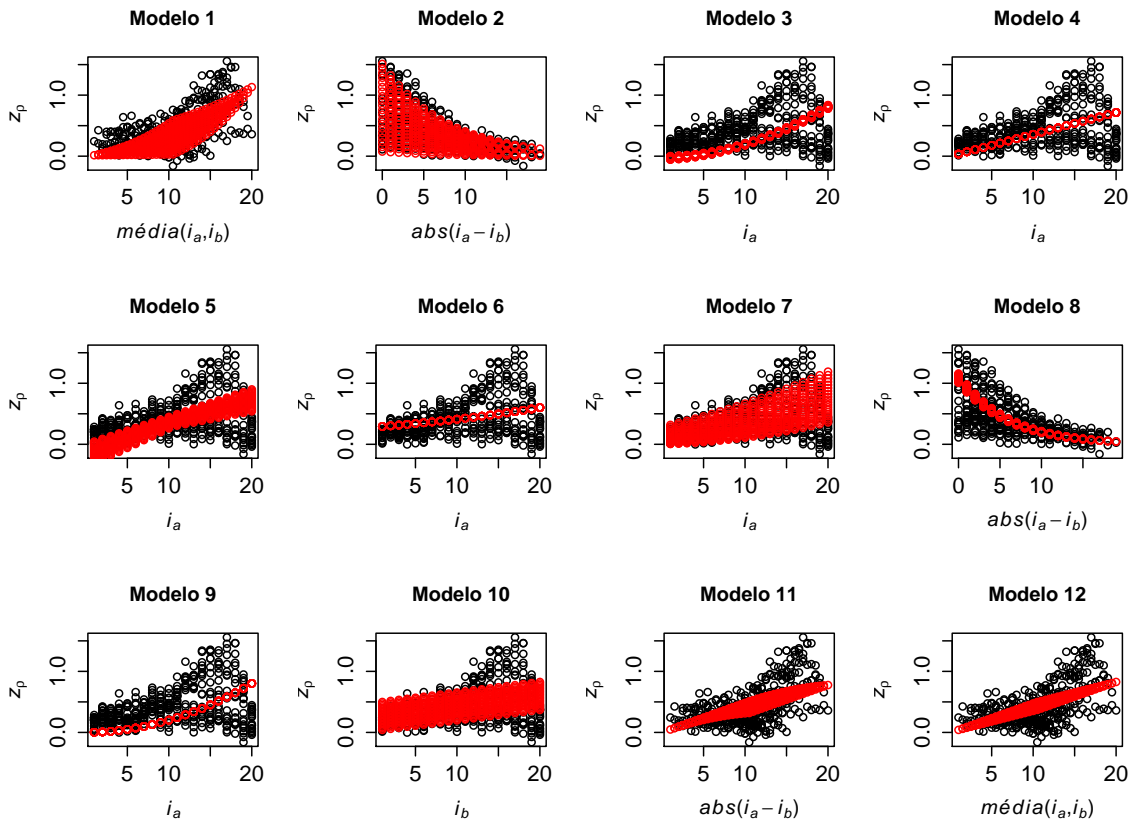


Figura 5.10: Representação conjunta da transformada de Fisher das correlações empíricas (a preto) e do ajustamento em função da idade (a vermelho), através de métodos de RNL, entre sexos

assim seja no decurso desse processo (destacamos, para além das operações com matrizes, outras, mesmo com escalares, que envolvam, por exemplo, funções exponenciais e logaritmos). Alguns destes problemas podem ter ou não a mesma causa e é necessário encontrar estratégias para os eliminar ou reduzir (quando a primeira opção não for realizável), de modo a não inviabilizarem o cálculo.

Como já referimos, testámos diferentes funções para explicar as correlações. Essas funções estão na origem das diferentes versões do modelo global, o MBGM, e são necessárias para seleccionar qual ou quais as versões que contribuem para um melhor desempenho do mesmo. Consoante a versão podem ocorrer problemas distintos na computação da matriz Σ , e, por sua vez, do determinante, relacionados sobretudo com a dimensão da matriz e a ordem de grandeza dos parâmetros a estimar. Por esse motivo, a função a otimizar foi reconstruída por várias vezes, para que se verificassem as propriedades das matrizes que garantem a não interrupção do método iterativo. Acerca de operações com matrizes no âmbito da análise numérica e otimização destacam-se [6, 31, 48].

Assim, não calculámos diretamente a inversa, nem o determinante da matriz Σ . Existem vários métodos de decomposição (conforme sugerido em [34, 74]). Usar a decomposição de Cholesky é uma forma de o fazer muito utilizada na literatura (porque à partida, se esta não existe, a matriz não

é definida positiva). Outros métodos, como os designados LU, QR, ou a decomposição através dos valores e vetores próprios, valores singulares, ou mesmo em blocos, foram experimentados, de modo a perceber como e as vantagens de, incluir a decomposição da matriz Σ no método iterativo.

Estudar os seus valores e vetores próprios é muito útil quando temos matrizes que suspeitamos são próximas de serem singulares. Pontualmente (basta que isso aconteça numa iterada), podemos ter valores próprios muito pequenos, positivos ou mesmo negativos (normalmente, é apenas um e tem a ver com o próprio algoritmo e a precisão da máquina), que podem fazer aproximar o valor do determinante a zero, o que inviabiliza o cálculo; por outro lado, se o determinante for mesmo negativo, não podemos obter o logaritmo deste (que é necessário para o cálculo da verosimilhança), o que também faz parar o processo iterativo. Assim, nestas situações em que, acompanhando o progresso das iteradas e percebendo que o método sugere um comportamento de convergência, podemos intervir diretamente nessa matriz dos valores próprios, modificando, pontualmente, apenas os valores próprios negativos, como em [78], que, das múltiplas referências a este tipo de abordagem, se revela a que provoca menores danos em termos de propagação do erro cometido na correção.

Por sua vez, decompor as matrizes em blocos (que no nosso caso faz todo o sentido pela construção de Σ) e obter a inversa de uma matriz de 2×2 blocos, como em [53], que não exige que todos os blocos sejam simultaneamente não singulares, também pode ser bastante útil. Para termos uma matriz de variâncias-covariâncias válida, alguns autores sugerem até somar um pequeno valor constante na sua diagonal (é uma forma de correção de Tikhonov do tipo $X + \lambda I$ com $\lambda \rightarrow 0$, mas é preciso atenção porque, ao introduzir esse tipo de ruído, estamos a aumentar a variância).

Em alternativa às abordagens anteriores, podemos ainda obter diretamente uma aproximação à matriz inversa e ao próprio determinante. No primeiro caso, através da matriz pseudoinversa ou de uma matriz inversa generalizada (processo muito eficiente, pois, quando a matriz a inverter não tem problemas, dá exatamente a sua inversa e, quando está próxima de ser singular, pode ser o suficiente para não interromper o processo iterativo). No caso do determinante de uma matriz, podemos aproximar o seu valor pelo do pseudodeterminante.

Neste ponto, refira-se que todos estes métodos são úteis para intervenções pontuais. Note-se, no entanto, que para cálculos com matrizes de grandes dimensões, que envolvem um número bastante elevado de iteradas, é necessário serem usados com cautela, pois, inevitavelmente, tratam-se de aproximações e comentem-se erros, que se propagam no processo iterativo e que são variáveis consoante também a própria precisão da máquina (é preciso desconstruir a matriz, aplicar os métodos de correção ou decomposição e depois reconstruir para obter uma nova matriz algebricamente mais conveniente da qual vamos obter a inversa e o determinante). Assim, se, por um lado, procuramos um processo iterativo mais eficiente e um algoritmo mais estável, por outro lado, a computação pode ser penalizada pelo acumular dos erros inerentes ao processo. Por esse motivo, e após compararmos exaustivamente resultados intermédios do processo iterativo, obtidos através de alguns destes métodos, optámos por construir a matriz de variâncias-covariâncias com recurso à matriz de correlações e à matriz diagonal das variâncias, isto é (como já referido), $\Sigma = \Lambda^{1/2} S \Lambda^{1/2}$, e usar, para a sua inversa, $\Sigma^{-1} = \Lambda^{-1/2} S^{-1} \Lambda^{-1/2}$ e, para o seu determinante, $|\Sigma| = |\Lambda^{1/2}| |S| |\Lambda^{1/2}|$.

Criada a estrutura para o modelo e encontrado um primeiro vetor de valores iniciais para os parâmetros, podemos dar início ao processo iterativo. Experimentámos várias versões do modelo global, sendo que, numa versão base, para escrever os R_k e os $\ln(\sigma_k^2)$ (que incluímos no modelo, invertendo a transformação logarítmica), usámos expressões polinomiais de grau três. Com efeito, são as que têm um melhor ajustamento em termos de julgamento empírico do coeficiente de determinação e, adicionalmente, testámos no algoritmo de otimização expressões polinomiais com grau inferior e o método não converge. Quanto às correlações, baseámo-nos nos modelos das correlações empíricas, usando um único tipo de modelo (de entre os enumerados em (5.9)) para as correlações entre idades diferentes dentro do mesmo sexo e entre sexos. Invertemos também a transformação de Fisher que aproximava as correlações à normal, tal que usámos, para cada par de idades, $\rho_{k_a, k_b} = \frac{\exp\{2z_{\rho_{k_a, k_b}}\} - 1}{\exp\{2z_{\rho_{k_a, k_b}}\} + 1}$. No apêndice E, apresentamos o código que diz respeito à construção das várias componentes do modelo global que designámos por MBGM.

Na caixa 5.5, mostramos os resultados do processo de otimização para o MBGM com as correlações descritas, como na versão 12 do modelo, enumerada em (5.9). Destacamos as estimativas de MV dos 22 parâmetros, do simétrico do valor ótimo estimado para a função de log-verosimilhança e ainda o código de convergência (0 indica que o processo iterativo foi bem sucedido).

Caixa 5.5 *Output* da aplicação do método de otimização para o MBGM (versão 12 do modelo)

```
min.LogV12 <- optim(P12, LogV12, control = list(trace = TRUE, maxit = 10000), hessian = TRUE)
>min.LogV12

$par
      r0F      r1F      r2F      r3F      r0M      r1M      r2M      r3M
-6.202490e-02  9.446018e-03 -7.708168e-04  4.400672e-05 -6.348173e-02  1.428726e-02 -1.636511e-03  8.454368e-05
      s0F      s1F      s2F      s3F      s0M      s1M      s2M      s3M
-3.604268e+00 -7.431115e-02 -1.038964e-02  1.201814e-03 -4.084286e+00  7.880474e-02 -2.476189e-02  1.659369e-03
      fiF      psiF
 6.045775e-02 -1.735374e-02
      fiM      psiM
 6.806910e-02 -2.080763e-02
      fiFM      psiFM
 1.373390e-04  2.117674e-02

$value
[1] 2479.851

$count
function gradient
      8011      NA

$convergence
[1] 0
```

Com a estimação dos parâmetros do modelo (neste caso ilustrámos com a versão 12) concluímos

a descrição das etapas da implementação do MBGM, que constam no diagrama da figura 5.3. Nesta subsecção apresentamos ainda as margens de erro que lhe estão associadas, bem como uma concretização do ajustamento, para diferentes versões do modelo.

Ao contrário de alguns programas estatísticos, no R, os desvios padrão, as estatísticas de teste e os intervalos de confiança para os parâmetros não são estimados diretamente a partir da computação da função `optim`. Assim, para o efeito, podemos obter a matriz hessiana (por exemplo, acrescentando `hessian = TRUE` como argumento da função), que invertamos e usamos como aproximação à matriz de Informação de Fisher. A partir daí, é possível obter os intervalos de confiança assintóticos (como descrito nos capítulos anteriores). Recordamos que, também neste caso, não temos de multiplicar a hessiana por -1 , porque o processo iterativo já foi feito com o simétrico da função de log-verosimilhança. Na caixa 5.6, apresentamos os valores dos parâmetros estimados bem como as margens de erro associadas aos intervalos de confiança assintóticos, neste caso, usando ainda a versão 12 do modelo global. Com exceção dos valores estimados para os parâmetros designados $r2F$, $r3F$ e $r3M$ (da mesma ordem de grandeza das estimativas dos parâmetros), as margens de erro são bastante aceitáveis.

Caixa 5.6 Estimativas dos parâmetros do MBGM (versão 12 do modelo) e margens de erro dos $\mathcal{IC}_{95\%}$ assintóticos

```
> cbind(par.name, min.LogV12$par,tol.modelo12)
      par          +/-
r0F  -6.202490e-02  6.102360e-05
r1F   9.446018e-03  1.616626e-04
r2F  -7.708168e-04  4.502343e-04
r3F   4.400672e-05  3.685700e-05
r0M  -6.348173e-02  3.426561e-05
r1M   1.428726e-02  1.504009e-04
r2M  -1.636511e-03  5.092685e-04
r3M   8.454368e-05  4.340542e-05
s0F  -3.604268e+00  2.681026e-05
s1F  -7.431115e-02  2.409874e-04
s2F  -1.038964e-02  2.403566e-03
s3F   1.201814e-03  1.081178e-06
s0M  -4.084286e+00  2.846100e-05
s1M   7.880474e-02  2.793112e-04
s2M  -2.476189e-02  2.933478e-03
s3M   1.659369e-03  7.899162e-07
fiF   6.045775e-02  1.931059e-06
psiF  -1.735374e-02  1.153927e-07
fiM   6.806910e-02  1.776918e-06
psiM  -2.080763e-02  1.210004e-07
fiFM  1.373390e-04  3.115959e-08
psiFM 2.117674e-02  3.266179e-08
```

Na figura 5.11, mostramos ainda uma concretização, neste caso ilustrada com a idade 1 de ambos os sexos, do ajustamento do modelo, através das versões 10, 12 e M. Esta última versão do

MBGM, que designamos por M (como abreviatura de modelo misto) é uma versão em que se usam regras diferentes para escrever as correlações entre idades diferentes dentro do mesmo sexo. Para as correlações entre idades diferentes dentro do sexo feminino e entre sexos, usámos o modelo 11. No caso das correlações entre idades diferentes dentro do sexo masculino, usámos o modelo 9.

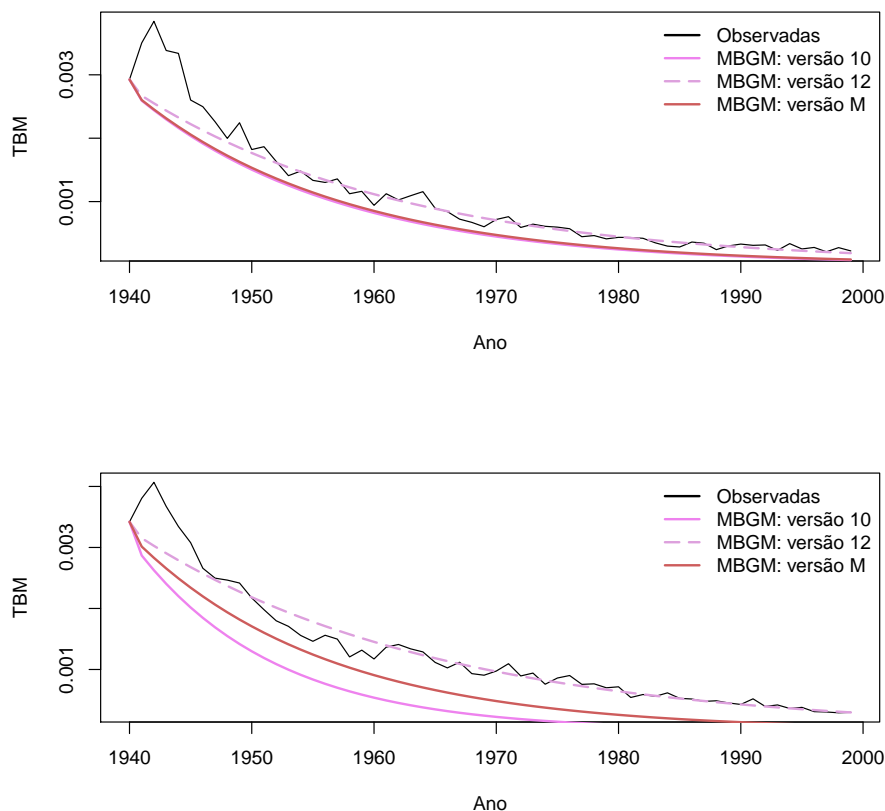


Figura 5.11: Ajustamento do MBGM (versões 10, 12 e M do modelo) para a idade 1 do sexo feminino (em cima) e masculino (em baixo)

Considerando os resultados do ajustamento do MBGM satisfatórios, vamos usá-lo para fazer previsões. Na subsecção seguinte apresentamos os resultados relativos às previsões e intervalos de confiança associados, bem como comparações entre versões do modelo global.

5.2.5 Resultados das previsões

Nas figuras 5.12 e 5.13, apresentamos as previsões a LP (para o período 2000 a 2009), no primeiro caso, para os logaritmos dos retornos das TBMs e, no segundo caso, na escala original dos dados, para as idades 1 a 10 do sexo feminino (que recordamos correspondem aos grupos etários dos 0-4 ao 45-50). Os resultados são bastante satisfatórios. Note-se que também outras versões do modelo, como a 3 ou a 9, por exemplo, apresentam resultados muito interessantes e igualmente satisfatórios (seleccionámos os da versão 12 do modelo, a título ilustrativo).

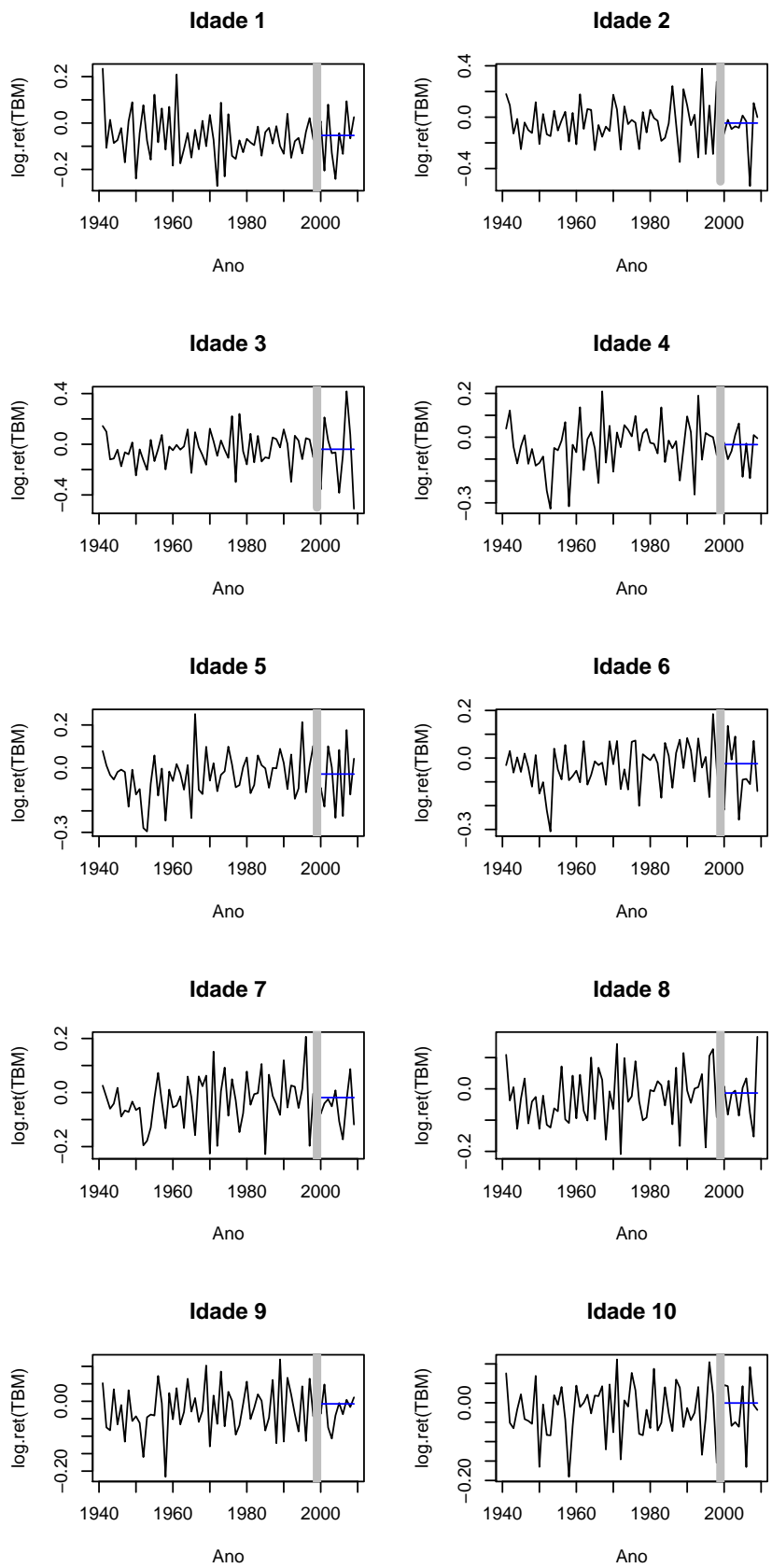


Figura 5.12: Previsões a LP (para o período 2000 a 2009, a azul) do MBGM (versão 12) dos logaritmos dos retornos das TBMs, $\log.\text{ret}(\text{TBM})$, das primeiras 10 idades do sexo feminino (que correspondem, respetivamente, aos grupos etários 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44 e 45-50)

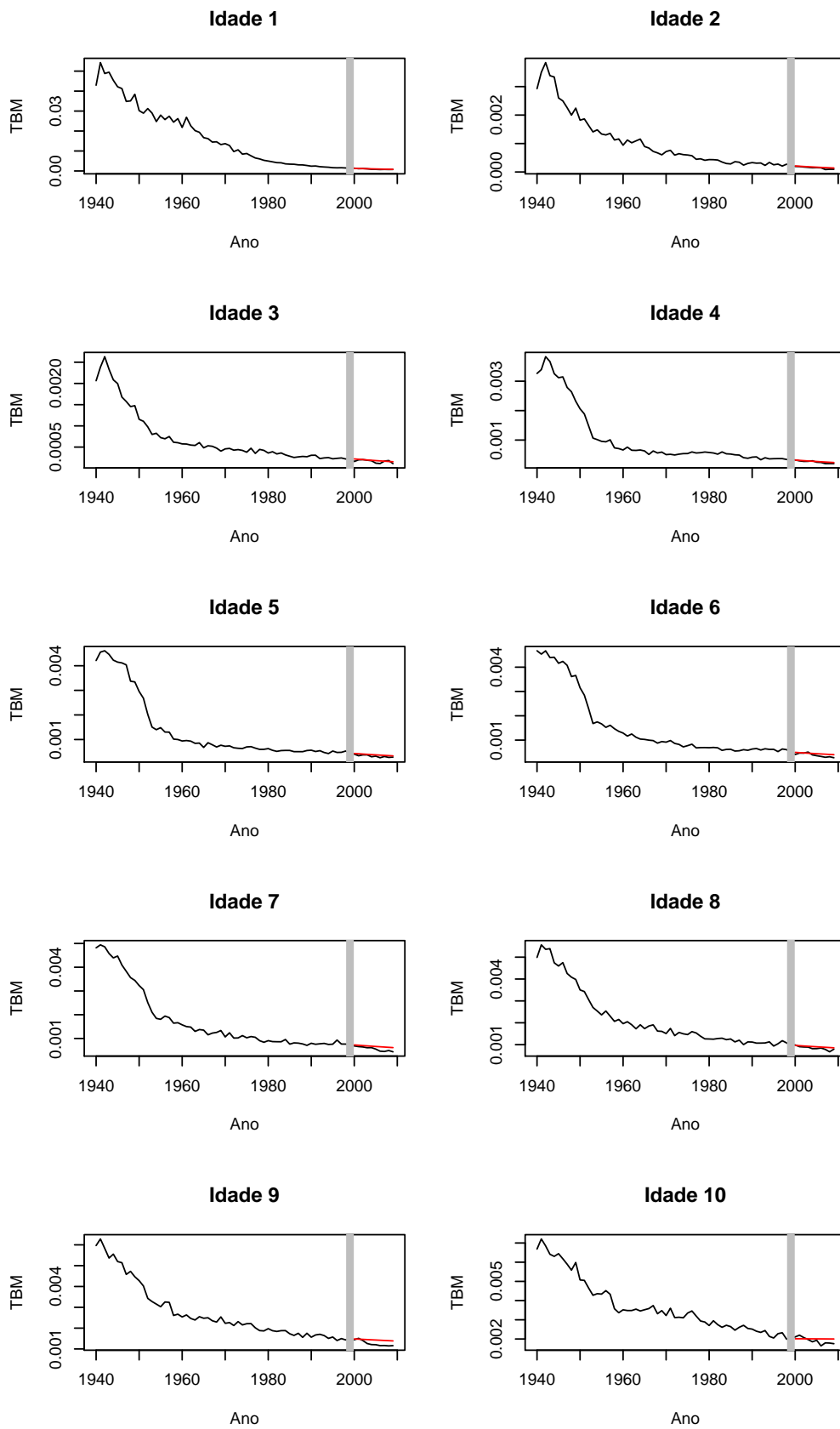


Figura 5.13: Previsões a LP (para o período 2000 a 2009, a vermelho) do MBGM (versão 12) das TBMs das primeiras 10 idades do sexo feminino (que correspondem, respetivamente, aos grupos etários 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44 e 45-50)

Os modelos de EDEs que temos vindo a aplicar, incorporam a incerteza no termo do processo de Wiener. Por isso, quando fazemos as previsões, consideramos o regime assintótico e a ausência de variabilidade ambiental tomando a esperança matemática do processo de Wiener (que é zero). Obtidas as previsões, podemos encontrar uma aproximação para os seus intervalos de confiança, estimando o que designamos por margens de variação natural do fenómeno e que refletem uma medida da influência direta dos efeitos aleatórios das variações ambientais (em sentido lato) no sistema das TBMs. Esta alternativa parece-nos preferível face à estimação dos intervalos de confiança pelo método de Delta. Com efeito, embora pudéssemos estimar a variância dos parâmetros a partir da informação da matriz empírica de Fisher (conforme ilustrado a propósito da estimação dos \mathcal{IC} assintóticos para as previsões dos modelos de EDEs unidimensionais, no capítulo 3), a aplicação do método Delta exige o cálculo de derivadas que a complexidade do modelo torna difícil obter. Outro método para obter uma aproximação dos \mathcal{IC} é através de simulação. A seguir, vamos mostrar os passos que este procedimento envolve.

Utilizando os valores dos parâmetros que foram estimados, para o período que vai do último instante observado até aos instantes que queremos prever, começamos por simular, digamos 1000 trajetórias do processo. Para cada instante t , para o qual queremos fazer a previsão, temos 1000 valores previstos, um por cada trajetória simulada. Esses valores formam uma distribuição empírica, cujas média e quantis empíricos podem ser utilizados para a previsão pontual e intervalar. Mais concretamente, tomamos a média como sendo uma aproximação ao valor previsto da TBM no instante t e os percentis 97.5% e 2.5% como extremos dos $\mathcal{IC}_{95\%}$ dessa previsão. Note-se que, em vez de simularmos diretamente as TBMs vamos antes, por comodidade, simular as trajetórias previstas para os $Z(t)$, utilizando (5.1) e, nos cálculos desta expressão, substituindo os valores dos parâmetros do modelo, pelos seus estimadores de MV.

No nosso caso, a última observação das TBMs corresponde ao ano de 1999 e queremos prever os anos 2000 a 2009 (isto é, $t = 2000, \dots, 2009$), pelo que tirámos partido de os incrementos de $\mathbf{W}(t)$ serem independentes para simular conjuntamente os anos, simulando os incrementos de 1999 para 2000, de 2000 para 2001, ..., de 2008 para 2009. Para facilitar os cálculos, simulámos $\mathbf{M}(t) = \mathbf{\Lambda}^{1/2}(\mathbf{W}(t) - \mathbf{W}(t-1))$, substituindo os parâmetros pelos seus estimadores de MV. Seja $\widetilde{\mathbf{M}}(t)$ uma dessas simulações. Os valores previstos para cada trajetória simulada são $\widetilde{\mathbf{Z}}(t) = \widehat{\mathbf{R}} + \widetilde{\mathbf{M}}(t)$. A partir desses valores podemos reconstituir a correspondente trajetória das TBMs, $\widetilde{\mathbf{X}}(t) = [\widetilde{X}_1(t), \dots, \widetilde{X}_K(t)]'$, usando iterativamente $\widetilde{X}_k(t) = \widetilde{X}_k(t-1) \exp\{\widetilde{Z}_k(t)\}$.

Coloca-se agora a questão de como simular $\widetilde{\mathbf{M}}(t)$? Partimos de um processo de Wiener multidimensional não correlacionado $\mathbf{W}^*(t)$. Para cada uma das 1000 trajetórias, simulamos o incremento $\widetilde{\mathbf{H}}(t) = \widetilde{\mathbf{W}}^*(t) - \widetilde{\mathbf{W}}^*(t-1)$, cuja matriz de variância-covariância é \mathbf{I} . Dado que \mathbf{W}^* é não correlacionado, estes incrementos são simples v.a. com distribuição $\mathcal{N}(\mathbf{0}, \mathbf{I})$ i.i.d., não oferecendo a simulação qualquer dificuldade.

Para cada trajetória, a partir destes incrementos obtemos as simulações de $\widetilde{\mathbf{M}}(t)$ através de

$$\widetilde{\mathbf{M}}(t) = \widehat{\mathbf{B}}\widetilde{\mathbf{H}}(t),$$

em que \widehat{B} se obtém substituindo em $\Lambda^{1/2}S^{1/2}$ os parâmetros pelos seus estimadores de MV. De facto, a matriz de variância-covariância de $BH(t)$ é $BB' = \Lambda^{1/2}S^{1/2}S^{1/2}\Lambda^{1/2} = \Lambda^{1/2}S\Lambda^{1/2} = \Sigma$, isto é, precisamente a matriz de variância-covariância de $M(t)$. Por outro lado, pela própria definição, os $M(t)$, para valores diferentes de t , são independentes entre si e têm média 0. Para obter $\widehat{B} = \widehat{\Lambda}^{1/2}\widehat{S}^{1/2} = \widehat{\Sigma}^{1/2}$, precisamos de obter $\widehat{\Lambda}^{1/2}$, o que é imediato, uma vez que $\widehat{\Lambda}$ é uma matriz diagonal, e precisamos também de obter $\widehat{S}^{1/2}$. Existem várias formas para obter a raiz quadrada de uma matriz mas, tratando-se de uma matriz real e simétrica com vetores próprios distintos, podemos usar a decomposição espectral, como indicado, por exemplo, em [83].

Para organizar os cálculos, procedemos da forma que ilustramos a seguir (ver código R no final do apêndice E):

- Calculamos $\widehat{R}_k = \widehat{\alpha}_{1,j}i + \widehat{\alpha}_{2,j}i^2 + \widehat{\alpha}_{3,j}i^3$ e construímos uma matriz, \widehat{R} , com 1000 colunas todas iguais em que cada coluna consta dos valores $\widehat{R}_1, \dots, \widehat{R}_K$;
- Calculamos $\ln(\widehat{\sigma}_k^2) = \widehat{\beta}_{1,j}i + \widehat{\beta}_{2,j}i^2 + \widehat{\beta}_{3,j}i^3$ e obtemos a matriz diagonal $\widehat{\Lambda}$, cujos elementos são os $\widehat{\sigma}_k^2$;
- Calculamos $z_{\rho_{k_a, k_b}}$, consoante a versão do modelo, e obtemos a estimativa da matriz de correlações \widehat{S} , usando os ρ_{k_a, k_b} ;
- Calculamos a matriz $\widehat{B} = \widehat{\Lambda}^{1/2}\widehat{S}^{1/2}$.

Para cada t ($t = 2000, \dots, 2009$), podemos obter os intervalos de confiança da previsão de $Y_k(t) = \ln(X_k(t))$ ($k = 1, \dots, K = 40$) do seguinte modo:

- Para cada k , simulamos um número suficientemente elevado de réplicas, digamos 1000, do incremento $W_k^*(t) - W_k^*(t-1) \sim \mathcal{N}(0, 1)$ i.i.d., obtendo assim uma matriz $\widetilde{H}(t)$ com K linhas (total de idades para ambos os sexos) e 1000 colunas;
- Estimamos o vetor $M(t)$ para cada uma das 1000 réplicas e construímos a matriz $\widetilde{M}(t) = \widehat{B}\widetilde{H}(t)$;
- Para cada uma das 1000 réplicas, prevemos $Z_k(t)$ pela equação (5.1) (usando os estimadores de MV dos parâmetros). Para o efeito, construímos a matriz, de dimensão $K \times 1000$, $\widetilde{Z}(t) = \widehat{R} + \widetilde{M}(t)$;
- Obtemos 1000 previsões $\widehat{Y}(t) = \bar{Y}(t_N) + \widetilde{Z}(t - t_N)$, onde $\bar{Y}(t_N)$ é uma matriz de dimensão $K \times 1000$ com colunas todas iguais, e em que cada coluna consta das últimas observações conhecidas $Y_1(t_N) = \ln(X_1(t_N)), \dots, Y_K(t_N) = \ln(X_K(t_N))$;
- Para cada idade k , isto é, em cada linha de $\widehat{Y}(t)$, ordenamos os 1000 valores dessa linha e calculamos os p -quantil e $(1-p)$ -quantil (no nosso caso usamos $p = 0.025$) que correspondem, respetivamente, aos limites inferior e superior dos IC da previsão de $Y_k(t)$;
- Em cada linha k de $\widehat{Y}(t)$, estimamos a média dos 1000 valores dessa linha, que deverá ser próxima do valor esperado da previsão de $Y_k(t)$;

- Naturalmente, os valores de X obtêm-se invertendo a transformação logarítmica.

No que se refere aos intervalos de confiança para as previsões, na figura 5.14, ilustramos (para a versão 12 do modelo global) uma aproximação aos mesmos, através de margens de variação natural para as previsões, para a idade 10 do sexo masculino (corresponde ao grupo etário 45-49). Foram usados os quantis empíricos a 2.5% e a 97.5% para estimar os limites das margens de confiança a 95%, $MC_{95\%}$, conforme descrito anteriormente. Como se pode observar, as amplitudes (na figura, em escala logarítmica) aumentam com o horizonte temporal da previsão, o que não é de todo surpreendente já que estamos a considerar a influência do ambiente em sentido lato e todas as possíveis causas suscetíveis de influenciar o fenómeno. Naturalmente, na representação na escala original dos dados, o efeito referido está ainda mais realçado (ver figura 5.15).

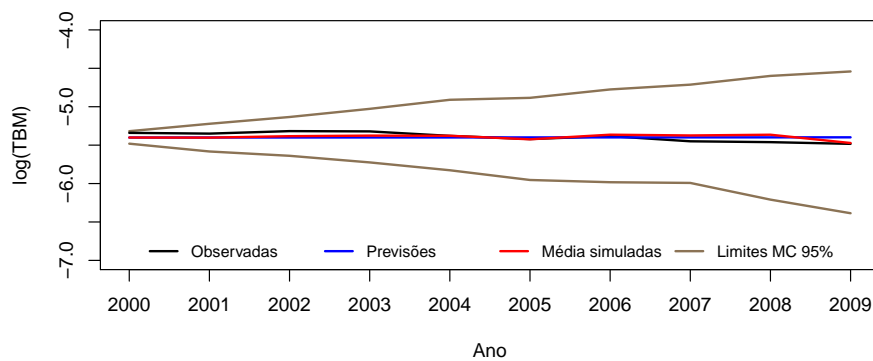


Figura 5.14: Representação conjunta das previsões a LP (em escala logarítmica), obtidas por estimação pontual (2000 a 2009) através da versão 12 do MBGM para a idade 10 (grupo etário 44-49) do sexo masculino, dos valores observados, bem como das médias das simulações e limites das margens de confiança a 95%

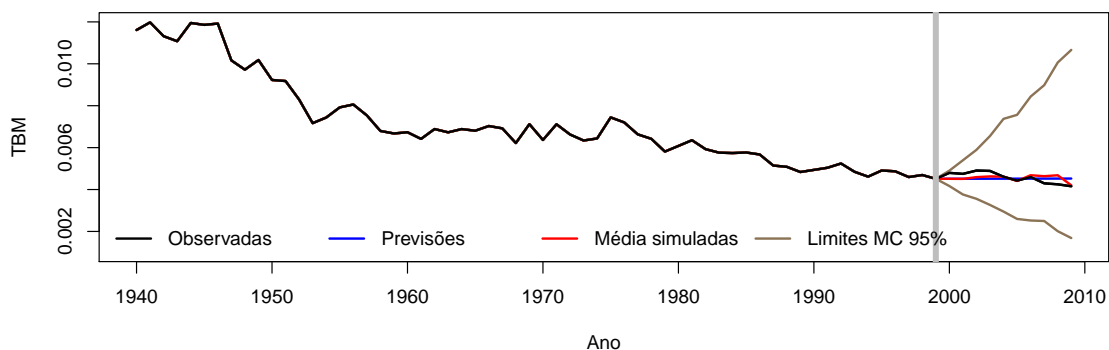


Figura 5.15: Representação conjunta das TBMs observadas, bem como das previsões a LP (de 2000 a 2009) obtidas por estimação pontual através do MBGM (versão 12) para a idade 10 (grupo etário 44-49) do sexo masculino, das médias das simulações e limites das margens de confiança a 95%

Não obstante a considerável amplitude das margens de confiança para as previsões, como expetável, em cada ano, as médias dos logaritmos das TBMs dos 1000 valores simulados são muito próximas dos valores das previsões. No exemplo da figura anterior (com exceção para o ano de 2006 e 2009), as curvas a azul e a vermelho estão sensivelmente sobrepostas (também com a das taxas observadas, o que significa um erro muito pequeno nas previsões), mas essa proximidade verifica-se para a generalidade das idades. Veja-se, a propósito, o exemplo ilustrado na figura 5.16, para o primeiro ano de previsões (que corresponde ao ano 2000), para todas as idades. Na sequência, na figura 5.17, mostramos ainda (para esse mesmo ano, mas a tendência verifica-se nos restantes anos das previsões) que as margens de confiança, para além de aumentarem com o horizonte temporal, são também mais amplas nas últimas idades (sobretudo a partir da idade 15, que corresponde ao grupo etário 75-79) em ambos os sexos.

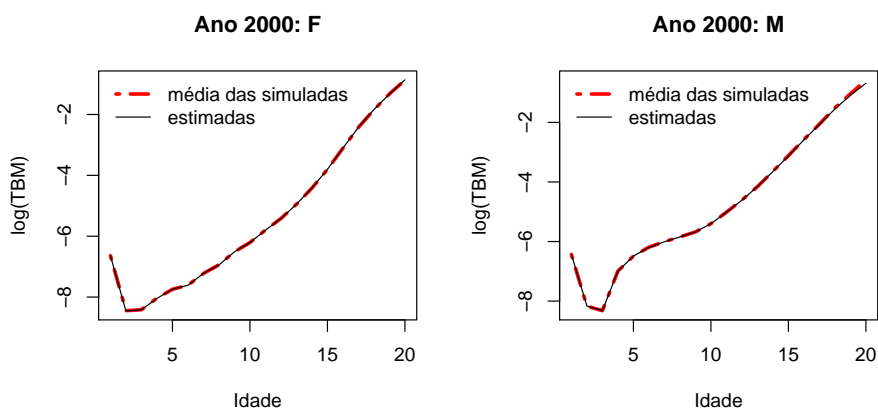


Figura 5.16: Representação conjunta das médias dos 1000 valores simulados e dos obtidos, na estimação pontual (ano 2000) para as previsões a LP, através da versão 12 do MBGM, para todas as idades e por sexo

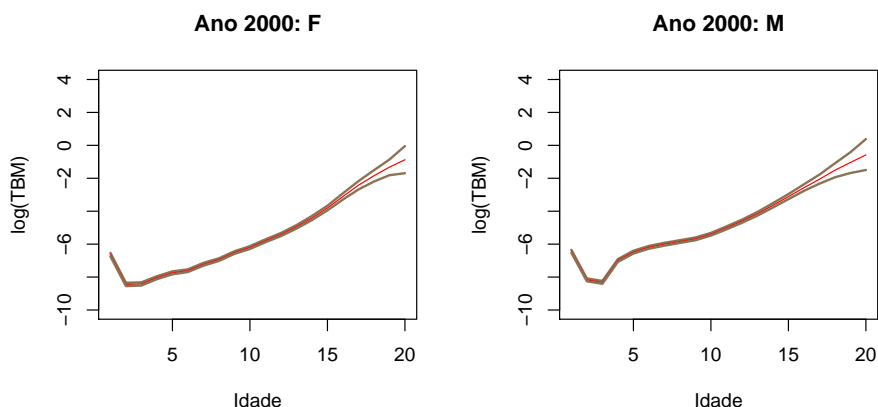


Figura 5.17: Representação conjunta das médias dos 1000 valores simulados (ano 2000) das previsões a LP do MBGM (versão 12), e dos limites das margens de confiança a 95% para todas as idades e por sexo

Para além dos resultados já apresentados, foram ajustadas outras versões do modelo global que diferem entre si no modo de escrever as correlações na matriz S . Vamos enumerar estas versões, de acordo com os modelos para as correlações, conforme indicámos em (5.9). Na caixa 5.7, mostramos os valores dos parâmetros estimados, que variam, em número, entre 19 a 22, para as versões que convergiram sem qualquer restrição no processo iterativo.

Caixa 5.7 MBGM: estimativas dos parâmetros para as diferentes versões do modelo e respetivos valores de log-verosimilhança, obtidos a partir de 5.7, e de AIC (NaN significa que o parâmetro não existe na versão do modelo em causa)

```
> print(resumoMBGM.par, quote = FALSE)
```

nomes	Modelo 3	Modelo 4	Modelo 5	Modelo 7	Modelo 9	Modelo 10	Modelo 12	Modelo M
[1,] rOF	0.0236985	-0.0464667	0.2697227	-0.0277157	-0.0952304	-0.0689216	-0.0620249	-0.0983971
[2,] r1F	0.0297488	-0.0061570	0.0131041	-0.0254013	-0.0023464	0.0072177	0.0094460	0.0229713
[3,] r2F	-0.001351	-0.0009258	-0.0037264	-0.0003264	-0.0013565	-0.0011728	-0.0007708	-0.0010695
[4,] r3F	-0.0001487	7.5938e-06	-9.1683e-05	-7.8536e-06	4.9496e-05	-0.0001661	4.4007e-05	-0.0002491
[5,] rOM	-0.0055980	-0.0472914	-0.0614742	-0.2928987	0.0275205	-0.0527798	-0.0634817	-0.11834
[6,] r1M	0.0216272	0.0093725	-0.0212830	0.0312689	-0.0134675	-0.0157667	0.0142873	0.032254
[7,] r2M	-0.0029008	-0.0018646	0.0013768	-0.0031088	-0.0017072	-0.000257	-0.0016365	-0.0022917
[8,] r3M	-7.4765e-05	5.0250e-05	1.6014e-05	-2.0529e-05	4.5905e-05	-0.000342	8.4544e-05	-2.2487e-05
[9,] sOF	-4.5799958	-3.416152	-2.856334	-3.953457	-4.100068	-4.398275	-3.604268	-3.550507
[10,] s1F	0.7567213	0.3097093	0.549167	0.5342476	0.3200653	0.578233	-0.0743112	-0.0623898
[11,] s2F	-0.0299477	0.0010773	-0.0572407	-0.0279793	-0.022682	-0.008714	-0.0103896	0.0547224
[12,] s3F	0.0001054	0.0014312	0.0021428	0.0006187	0.0022017	2.6225e-05	0.0012018	0.0007291
[13,] sOM	-4.6517804	-4.339295	-3.531858	-4.499927	-4.982052	-4.574945	-4.084286	-4.108555
[14,] s1M	0.6490056	-0.0137449	0.4516249	0.6995288	0.449083	0.7511687	0.0788047	0.162962
[15,] s2M	-0.0294651	-0.0079582	0.0013071	-0.0244491	-0.0163596	-0.0029222	-0.0247619	0.0026178
[16,] s3M	0.0008513	0.0011731	-0.0008185	0.0006143	0.0008293	-0.0004993	0.0016594	0.0003592
[17,] fi.F	0.0011640	-0.0164719	0.0038456	0.0169853	1.7552e-05	-0.1258083	0.0604578	0.0189071
[18,] psi.F	-0.6755583	NaN	-0.3227428	-0.0007605	NaN	0.2416007	-0.0173537	1.4130e-05
[19,] fi.M	0.0006326	0.0099158	-1.414464	0.0073636	1.7533e-05	-0.0905153	0.0680691	0.0546431
[20,] psi.M	-0.4011731	NaN	-0.3991361	-0.0055385	NaN	0.1880185	-0.0208076	NaN
[21,] fi.FM	0.047454	0.0027804	0.1825654	0.0095203	0.2888648	0.0122794	0.0001373	3.3431e-05
[22,] psi.FM	0.4590829	NaN	0.4876565	0.0029507	NaN	-0.0114797	0.0211767	6.0573e-05
[23,] LogV	-2716.061	-2569.801	-2410.468	-2466.576	-3070.805	-2249.849	-2479.851	-2298.985
[24,] AIC	5476.122	5177.602	4864.936	4977.152	6179.610	4543.698	5003.702	4639.970

Muito embora sugerissem um bom desempenho ao nível da estrutura de correlação, de acordo com os gráficos das figuras 5.8 a 5.10 e os resultados associados que constam na caixa 5.4, “descartámos” as versões 1, 2, 6, 8 e 11, por não se verificar a convergência do método iterativo. Recordamos que a versão do MBGM que designamos por M é uma versão em que as correlações não são obtidas à custa de uma única expressão - poder-se-ia designar por modelo misto. Neste caso, para as correlações dentro do sexo feminino e entre sexos, usámos o modelo 11 e, no caso das correlações dentro do sexo masculino, usámos o modelo 9 (esta versão tem assim 21 parâmetros). Experimentámos este tipo de construção, para testar a convergência do modelo global com uma estrutura de correlação que suspeitámos ser mais adequada por sexo, ou entre sexos, e

apresentamos apenas este exemplo a título ilustrativo dessa abordagem distinta à introdução das correlações no modelo.

Ainda na mesma caixa 5.7, apresentamos, depois dos valores dos parâmetros estimados, e para cada versão do modelo, o valor da log-verosimilhança e do AIC associado ($AIC = 2l - 2 \log V$, com l o número de parâmetros e $\log V$ o logaritmo da verosimilhança). Relativamente aos valores de log-verosimilhança obtidos, note-se que são muito próximos, sendo que as versões do modelo 10 e M apresentam os maiores valores, que por sua vez são muito próximos dos obtidos para as versões 5, 7, 12, ou mesmo para a 4. As versões 3 e 9 têm efetivamente os menores valores de log-verosimilhança e um desempenho global menos bom para a generalidade das idades de ambos os sexos. No entanto, mais à frente, veremos uma particularidade no desempenho destas versões que pode trazer vantagens na aplicação da metodologia. No que se refere ao valores obtidos para o critério AIC, estes sugerem que se tivéssemos que decidir qual a melhor versão do modelo a usar seria a versão 10 (com o menor valor de AIC) enquanto que a versão 9 (com o maior valor de AIC) seria posta de parte.

Para continuar a analisar o desempenho das diferentes versões do modelo global, estimámos o EQM das previsões a LP. Na figura 5.18, ilustramos os EQMs das previsões, de cada versão do modelo, e para cada idade.

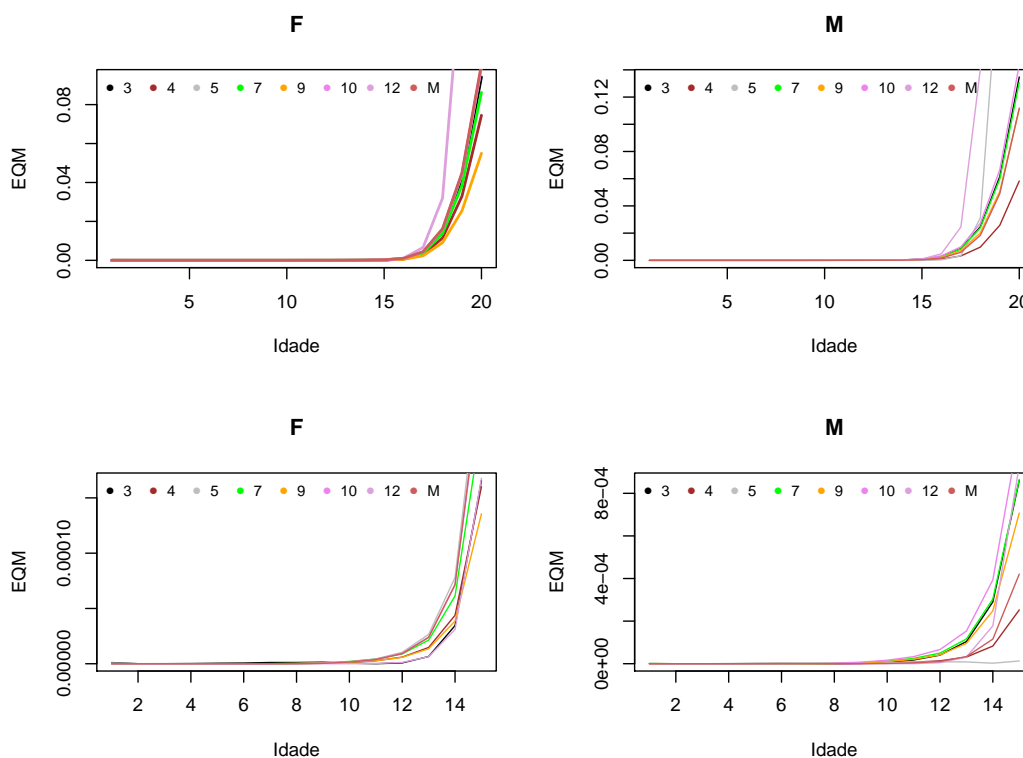


Figura 5.18: EQM das previsões a LP das TBMs, relativas ao período 2000 a 2009, pelas várias versões do MBGM, por sexo; todas as idades (que correspondem a todos os grupos etários quinquenais selecionados), em cima, e idades 1 a 15 (corresponde a excluir as idades a partir do grupo etário 70-74), em baixo

Como podemos observar, estes aumentam, em ambos os sexos, depois da idade 15, que corresponde ao grupo etário 70-74 (de resto, à semelhança de todos os modelos ajustados anteriormente, e referimo-nos, concretamente, aos modelos unidimensionais e bidimensional que apresentámos nos capítulos e nas secções anteriores). Por este motivo, e para melhor visibilidade, representámos na mesma figura os EQM apenas até à idade 15. Concluimos que, na maior parte da versões, o erro das previsões aumenta com a idade e que é superior no sexo masculino (o que também já se havia verificado como resultado da aplicação dos modelos anteriores). Na caixa 5.8, de modo a ter um indicador mais geral do desempenho do modelo, apresentamos uma estimativa da média do EQM para todas as idades e também para as idades até à idade 15, por sexo.

Caixa 5.8 Estimativas da média (para todas as idades e excluindo as últimas cinco) dos EQM das previsões a LP das TBMs, no período de 2000 a 2009, segundo as diferentes versões do MBGM, por sexo

```
> print(resumoMBGM.EQM, quote = FALSE)
      Modelo EQM.1a20.F   EQM.1a20.M   EQM.1a15.F   EQM.1a15.M
[1,] 3      0.007726530890 0.01168955976 1.429373094e-05 8.826653705e-05
[2,] 4      0.006136585897 0.00491669631 1.528051886e-05 2.611438355e-05
[3,] 5      0.008255510679 0.08105219713 3.088834018e-05 3.189856320e-06
[4,] 7      0.007168057332 0.01134282965 2.104957488e-05 9.201372565e-05
[5,] 9      0.004629126093 0.00961927345 1.323993085e-05 7.523485044e-05
[6,] 10     0.008273353459 0.01263265839 2.456319539e-05 0.000117645875
[7,] 12     0.046391211670 0.24471698336 1.378628878e-05 7.641249633e-05
[8,] M      0.008400252155 0.00936920847 2.480515235e-05 3.915070029e-05
```

Como seria expectável, os erros diminuem quando reduzimos o número de idades. Pelo exposto, não podemos afirmar que uma versão do MBGM seja francamente mais vantajosa, em termos de previsões, comparativamente às restantes. Até porque algumas das versões do modelo que não incluímos neste resumo de resultados por terem terminado o processo iterativo com alguma limitação de cálculo, apresentam mesmo assim bons resultados para algumas idades e essa informação de desempenho pode ser útil em trabalho futuro. No caso da versão 11, por exemplo (que ilustramos na figura 5.19), os resultados mostram uma inconsistência que se traduz num desempenho desigual por sexo a partir de uma certa idade (concretamente a idade 14, que corresponde ao grupo etário 65-69). No entanto, como para o sexo feminino os erros das previsões eram pequenos, comparativamente a outras versões, ocorreu-nos usar essa forma de escrever as correlações, para o sexo feminino, no modelo "misto" M.

Ainda a propósito do desempenho do modelo global, os maiores EQMs que surgem, em todas as versões, nas idades mais avançadas e traduzem um ajustamento menos bom são uma característica da generalidade dos modelos estocásticos e referenciada na literatura demográfica. Refira-se a propósito que, como sugestão de resolução deste problema, é frequentemente indicada a reunião dessas idades num único grupo dos 85 ou mais anos (que não equacionámos, embora fosse

exequível, pois teríamos que ter, nesse caso, uma amplitude do grupo etário diferente dos restantes). Como é um argumento válido o facto de qualquer acontecimento aleatório poder provocar o óbito nestas idades - e de ser quase impossível explicar estatisticamente esses acontecimentos - leva a que alguns autores sugiram ainda que simplesmente não faz sentido incluí-las conjuntamente com as restantes (sobre estes aspetos, ver, por exemplo, [47]).

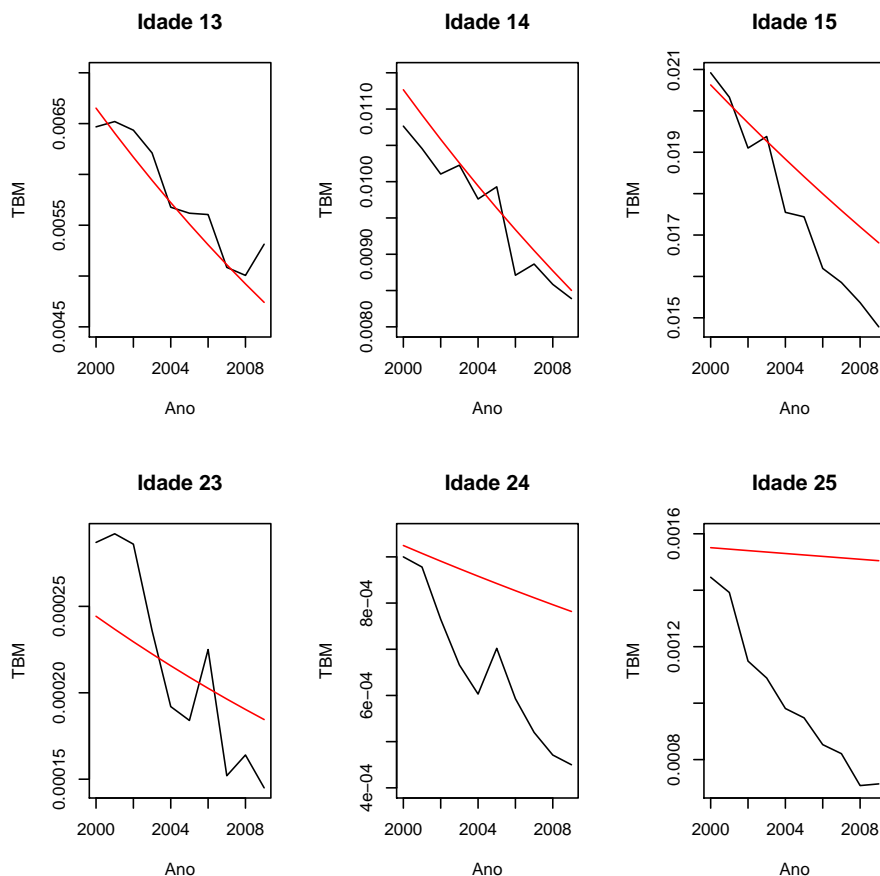


Figura 5.19: Previsões a LP, no período de 2000 a 2009 (a vermelho), segundo a versão 11 do MBGM das idades 13 a 15 (que correspondem aos grupos etários 60-64, 65-69 e 70-74) do sexo feminino (em cima) e idades correspondentes aos mesmos grupos etários do sexo masculino (em baixo)

Considerando outras possibilidades, na mesma linha de investigação que temos vindo a seguir, comparamos os valores da verosimilhança das versões do modelo global, digamos completo, com os obtidos para os modelos que só consideram $W(t)$ correlacionados por idades, mas apenas dentro de cada sexo, isto é, com correlações nulas entre idades de sexos diferentes. Estes modelos simplificados correspondem, na prática, a trabalhar separadamente o sexo feminino e o sexo masculino, sendo a sua log-verosimilhança, a soma das log-verosimilhanças correspondentes aos dois sexos. Para comparar versões do modelo, construímos um teste de razões de verosimilhanças, que facilmente se implementa, a partir dos valores da log-verosimilhança estimados. Note-se que os resultados para os casos particulares, agora referidos, são obtidos da mesma forma que para o modelo global, simplesmente o número de dimensões reduz-se para metade, isto é, $K/2$, e a matriz

de correlações, S , é constituída por um único bloco, já que apenas se consideram as correlações entre os logaritmos dos retornos das TBMs associados a idades diferentes de um mesmo sexo.

Para comparar versões do modelo com o efeito das correlações entre sexos com o modelo sem as correlações entre sexos, aplicamos então um teste de razão de verossimilhanças (salienta-se que os dois tipos de versões do modelo já consideram o efeito das correlações entre idades diferentes dentro do mesmo sexo). Para o teste, fazemos

$$H_0 : \rho_{FM} = 0 \quad \text{vs} \quad H_1 : \rho_{FM} \neq 0$$

e usamos a estatística de teste

$$D = -2 \ln \left(\frac{\text{Verossimilhança do modelo com } \rho_{FM} = 0}{\text{Verossimilhança do modelo com } \rho_{FM} \neq 0} \right),$$

que, para a hipótese nula, tem uma distribuição aproximadamente qui-quadrado com dois a quatro graus de liberdade (dependendo da versão do modelo). Na tabela 5.1, apresentamos os valores estimados da estatística de teste D para diferentes versões do modelo. Os resultados sugerem que as TBMs de ambos os sexos estão significativamente correlacionadas, não só entre idades diferentes dentro de cada sexo, mas também entre sexos.

Tabela 5.1: Resultados do teste de razão de verossimilhanças para versões do modelo sem e com correlações entre sexos (***) representa valores- $p < 0.005$)

Modelo	5	10	12	M
D	323.064***	726.442***	1151.352***	479.862***

Note-se que não mostrámos os resultados para as versões 3, 4, 7 e 9 do modelo global, porque no comparativo entre os valores das log-verossimilhanças, os resultados para os modelos que consideram apenas as correlações entre idades diferentes de um mesmo sexo foram obtidos com opções diferentes da função `optim` do R das do modelo global. Por esse motivo, consideramos que não podem ser diretamente comparáveis e por isso não os apresentámos. Não está contudo em causa a convergência e tão pouco o resultado da estatística de teste D . No caso da versão 3, por exemplo, $D = 11.956$, o que sugere que é mais vantajoso aplicar o modelo com correlações entre idades diferentes de sexos diferentes para além das correlações entre idades diferentes de um mesmo sexo, do que o modelo com apenas este último tipo de correlações.

Atendendo aos resultados da tabela acima, ilustramos, na figura 5.20, para a idade 5 (grupo etário 20-24) do sexo masculino, as previsões obtidas através da versão 10 do modelo global com e sem correlações entre idades de sexos diferentes, em que é notória a vantagem no desempenho quando se usa a versão que considera as correlações entre idades de sexos diferentes para estimar as previsões. No entanto, em algumas versões (versão 3, por exemplo), muito embora, globalmente, o resultado dos testes de razão de verossimilhanças seja favorável à versão do modelo com correlações entre idades de sexos diferentes, as previsões nas idades mais avançadas têm menores EQMs quando aquelas não são consideradas. Na figura 5.21, ilustramos os resultados obtidos pela versão 3 do modelo global para a idade 17 (grupo etário 80-84) do sexo feminino. Estes resultados, que

traduzem uma vantagem do modelo sem correlações entre idades de sexos diferentes, não podem, ainda assim, generalizar-se, nomeadamente, para todas as idades depois da idade 15 (grupo etário 70-74) das versões referidas do modelo global.

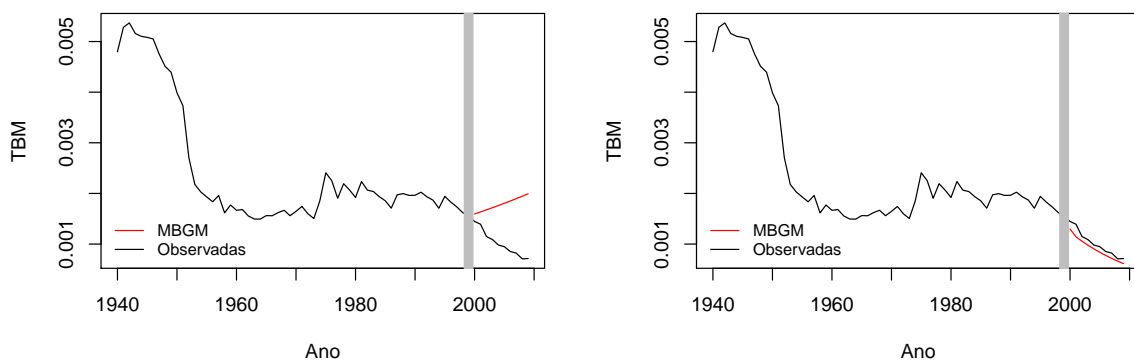


Figura 5.20: Previsões a LP (2000 a 2009) para a idade 5 (grupo etário 20-24) do sexo masculino segundo a versão 10 do MBGM, sem (à esquerda) e com (à direita) correlações entre idades de sexos diferentes

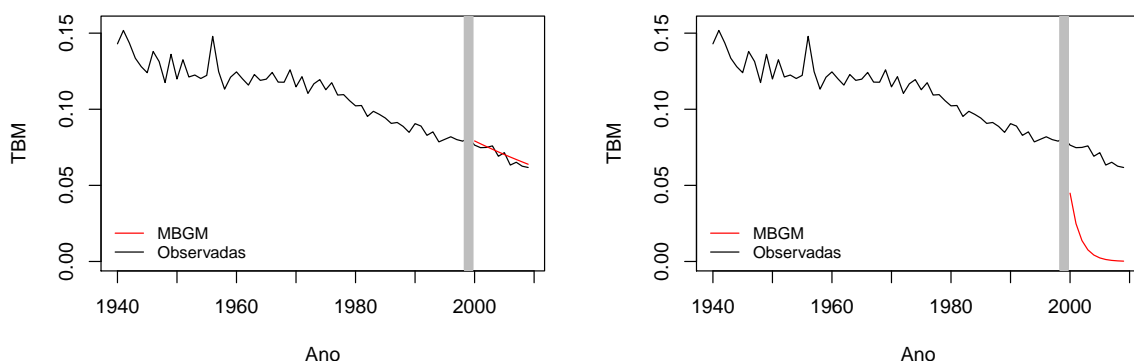


Figura 5.21: Previsões a LP (2000 a 2009) para a idade 17 (grupo etário 80-84) do sexo feminino segundo a versão 3 do MBGM, sem (à esquerda) e com (à direita) correlações entre idades de sexos diferentes

Considerando ainda as vantagens da utilização do modelo global, comparámos também as versões do modelo completo com as respetivas versões do modelo com um R_k para cada idade, isto é, em que não assumimos que os R_k são funções polinomiais da idade i , e que temos, em vez de 8 parâmetros (4 para descrever os R_k do sexo feminino e outros 4 para o sexo masculino), $K = 40$ parâmetros, os próprios R_k ($k = 1 \dots 40$).

No caso do modelo com um R_k para cada idade, não se pode aplicar o teste anterior, pois os modelos não são equivalentes (pelo motivo exposto, porque os R_k não são estimados com recurso a uma estrutura polinomial). Podemos, ainda assim, comparar os valores das log-verossimilhanças e também dos AIC obtidos. Como o número de parâmetros aumenta bastante, nestas novas versões,

é expectável que o valor da log-verosimilhança seja superior ao obtido para as versões base do modelo global. De facto, esses resultados confirmam-se. O valor da log-verosimilhança estimado, por exemplo, para a versão 10 do modelo global, é de -2249.849, enquanto que o valor correspondente a essa versão com um R_k para cada idade é de -2178.474. A partir destes valores podemos obter os valores associados de AIC para cada versão do modelo (respetivamente 4543.698, para a versão base do modelo global, com 22 parâmetros, e 4464.948, para a versão com um total de 54 parâmetros), valores esses que estão em conformidades com os de log-verosimilhança, pois sugerem um melhor desempenho da versão com um R_k para cada idade. Os valores destes critérios são, no entanto, muito próximos e muito embora o processo iterativo destas novas versões tenha sido concluído sem erros, o esforço computacional envolvido é bastante mais elevado que o necessário às versões base (que já não é reduzido), o que nos faz questionar acerca das efetivas vantagens da sua aplicação. Adicionalmente, a análise pormenorizada das iteradas (observando as fases de expansão e contração) sugere possíveis problemas de convergência, que carecem de análise complementar e possivelmente de uma nova abordagem de programação mais sensível ao elevado número de parâmetros (mais de 50) e dimensões envolvidas, como é sugerido, por exemplo, em [32].

Por último, podemos ainda comparar as log-verosimilhanças das versões do MBGM com a soma das log-verosimilhanças do MBG. Ou seja, vamos comparar os resultados do movimento browniano geométrico multidimensional, com uma estrutura de correlação para os processos de Wiener associados às EDEs que modelam as TBMs de idades diferentes dentro de um mesmo sexo e de sexos diferentes, com os resultados da aplicação do movimento browniano geométrico às TBMs de cada idade e sexo, isto é, sem qualquer estrutura de correlação entre os processos de Wiener de cada EDE. A soma das log-verosimilhanças, resultantes da aplicação do MBG unidimensional a cada um dos 20 grupos etários quinquenais de cada sexo, é -2657.674, a que corresponde um $AIC=5475.348$. Mais uma vez, não podemos fazer o teste da razão de verosimilhanças descrito anteriormente pois os modelos não são diretamente comparáveis: nos modelos unidimensionais não escrevemos os R_k e os σ_k^2 como funções polinomiais.

Se considerarmos os resultados da caixa 5.7, relativos às diferentes versões do MBGM (cujo número de parâmetros varia entre 19 a 22), constatamos que o valor da soma das log-verosimilhanças resultantes da aplicação do modelo unidimensional (com um total de 80 parâmetros, uma vez que temos um R e um σ^2 diferentes para cada idade e por sexo) é menor que os obtidos para a maior parte das versões do modelo multidimensional global (exceção para as versões 3 e 9). Este comparativo traduz assim uma vantagem na aplicação do modelo multidimensional, confirmada também se compararmos os valores dos AIC obtidos (menores nas versões do modelo multidimensional, exceto no caso das versões 3 e 9).

Pelo exposto, o MBGM parece ser adequado para este tipo de dados, com previsões muito satisfatórias, sobretudo até às idades que correspondem ao grupo etário 45-49. De resto, ao longo desta secção, mostrámos os resultados obtidos a partir de diferentes versões do modelo global, em vez de indicar qual a “melhor versão”. Se atendermos aos resultados conjuntos dos valores da log-verosimilhança e do AIC, que apresentamos na caixa 5.7, seria a versão 10. Se atendermos

a outros critérios, e considerarmos, por exemplo, os EQMs das previsões, que apresentamos na caixa 5.8, seria a versão 4, se considerássemos todas as idades de ambos os sexos, ou a 5, apenas para as idades do sexo masculino até ao grupo etário 70-74 (idade 15). Note-se, no entanto, que os valores dos EQMs apresentados, em termos de síntese, na caixa referida são as médias dos EQMs para todas as idades e que valores mais elevados em algumas versões (designadamente na versão 10) podem traduzir apenas casos excecionais de previsões menos boas (em que o EQM pode “disparar”) em algumas idades, o que vem prejudicar o desempenho global do modelo, mas não deve traduzir imediatamente a sua não adequação aos dados. Em suma, decidimos não fixar uma forma estanque para o modelo, para que se possa continuar a trabalhar na metodologia e, por esta via, continuar a apostar na melhoria do seu desempenho.

Para terminar, considerando as vantagens da utilização do MBGM, ou pelo menos para enquadrar os resultados, na figura 5.22, fazemos a representação conjunta das previsões a LP (previsões a 25 anos para o período de 2000 a 2024) obtidas para a idade 17 (grupo etário 80-84), de acordo com a versão 3 do modelo global, sem correlações entre sexos, e, para a idade 80 (que corresponde ao grupo etário anual 80), através do modelo unidimensional MBG.

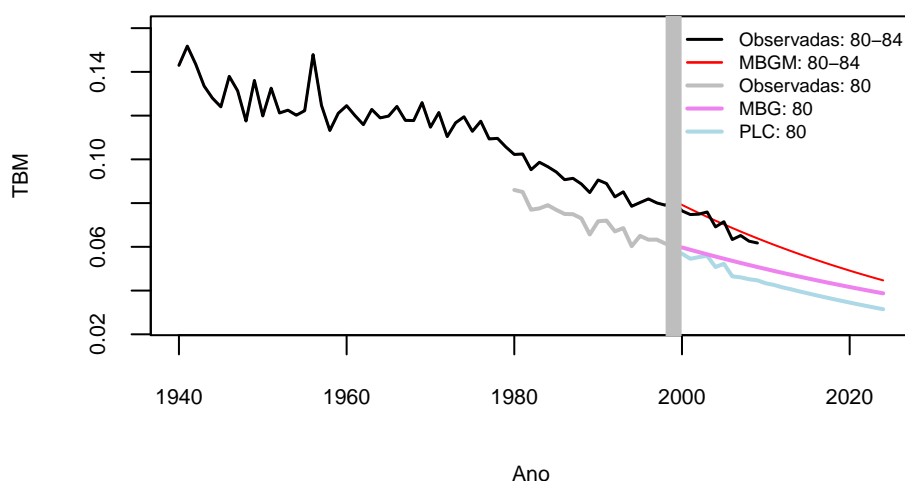


Figura 5.22: Representação conjunta das previsões a LP (25 anos: de 2000 a 2024) para o grupo etário 80-84 (idade 17) do sexo feminino, segundo a versão 3 do MBGM sem correlações entre sexos; previsões a 25 anos do MBG para a idade 80 do sexo feminino e as projeções obtidas a partir do modelo de Poisson-Lee-Carter (PLC), para o mesmo período das anteriores e também para a idade 80 do sexo feminino (fonte: figura adaptada de [21],pág.10 - os valores para o período referido foram gentilmente cedidos pelo autor)

Notamos que os resultados representados, que seleccionámos a título ilustrativo, não podem ser diretamente comparados, pois, efetivamente, as idades representam, no caso do modelo multivariado, um grupo de cinco anos, enquanto que, no modelo unidimensional, um ano (e também

porque a TBM é superior no caso do grupo etário 80-84 comparativamente à idade 80). Podemos, no entanto, observar uma tendência semelhante, que, no caso da previsão para o grupo quinquenal, capta melhor a variabilidade da série e, no caso da previsão unidimensional, sobrestima. Refira-se ainda que, para enquadramento dos resultados, ilustramos, nesta representação conjunta, as projeções, para a idade 80, para a componente mortalidade no âmbito do exercício de 2008 de projeções para a população portuguesa e cuja metodologia, que passa pela aplicação do modelo de Poisson-Lee-Carter, é descrita em [21].

6

Conclusões e trabalho futuro

Como se explica a mortalidade? Em geral, se falarmos de um indivíduo, podemos por vezes, em condições muito especiais, ter uma explicação objetiva e afirmar que, naquele momento, aquele indivíduo morreu por uma razão que conhecemos e até antecipar, numa certa medida, a sua morte. Já no contexto de uma população, podemos, quando muito, apontar algumas causas mais ou menos gerais que influenciam, em algum momento de tempo, a probabilidade de cada indivíduo ou de grupos de indivíduos morrerem. Em suma, uma das únicas certezas da vida é que, para cada indivíduo, num certo tempo, ocorre o óbito. Explicar, através de um modelo estatístico, a tendência evolutiva das taxas de mortalidade de uma população ao longo do tempo e fazer previsões, por idade e sexo, com margens de erro associadas, parece-nos pertinente, sobretudo na atual conjuntura sócio-económica e considerando o fenómeno geral de envelhecimento populacional.

É sabido que não há consenso quanto à questão sobre qual é o melhor modelo para explicar a mortalidade. Existem, sim, vantagens na utilização das diferentes abordagens, consoante a finalidade a que se destinam. Neste trabalho, propusemo-nos fazer uma análise transversal da mortalidade ao longo do tempo e modelar as taxas brutas de mortalidade (TBMs) associadas a cada idade e por sexo, através de equações diferenciais estocásticas (EDEs). Mais, pretendemos como objetivo final explicar, num único modelo, a mortalidade em todo o arco de vida (e para ambos os sexos) e fazer previsões.

Os modelos propostos e aplicados não consideram explicitamente causas “observáveis”, como uma epidemia de gripe, um fenómeno meteorológico extremo (ou outro tipo de catástrofe), as acessibilidades aos cuidados de saúde ou as mudanças na legislação ou políticas de saúde, entre outras. No entanto, são capazes de as incorporar. Ao considerarmos a mortalidade um processo estocástico e modelarmos as TBMs através de EDEs, incorporamos uma componente aleatória nos modelos, que traduz a influência do ambiente (em sentido lato) sobre a mortalidade, representando o conjunto de todas as causas suscetíveis de fazer variar a mortalidade. Simultaneamente, enveredámos por uma análise transversal ao longo do tempo, que tem a vantagem de não se colocarem questões de *timing* relativamente ao óbito, pois, efetivamente, estudam-se as séries temporais disponíveis, que apresentam, quando organizadas deste modo, curvas mais simples de ajustamento.

Para além de aplicarmos modelos já muito difundidos de EDEs (no entanto, para este tipo de dados, com escassas aplicações), como é o caso do modelo de Gompertz estocástico (MGE) ou o do movimento browniano geométrico (MBG), propusemos ainda três novos modelos multivariados. Nos três casos, introduzimos uma estrutura de correlação associada aos processos de Wiener que representam a componente estocástica em cada modelo. Nos modelos bidimensionais, considerámos as correlações entre os processos de Wiener associados às EDEs entre sexos, para cada idade, fixando a idade. Já, no modelo multidimensional, considerámos as correlações entre os processos de Wiener associados às EDEs de diferentes idades dentro de cada sexo e de sexos diferentes. No seu conjunto, consideramos a nossa abordagem inovadora e não encontramos referências semelhantes na literatura.

Assim, para perceber a tendência evolutiva, ao longo do tempo, da variável demográfica TBM,

aplicámos, enquanto abordagem inicial, os modelos unidimensionais de EDEs, MBG e MGE às séries temporais das TBMs da população portuguesa, disponíveis no período de 1940 a 2009. Para estes modelos, fizemos um estudo analítico, estimámos os parâmetros para cada idade e sexo e fizemos previsões passo a passo e a médio/longo prazo, para as quais foi ainda possível estimar os intervalos de confiança associados. Consideramos os resultados obtidos bastante satisfatórios, com erros que, em geral, aumentam com a idade e são maiores no sexo masculino (em parte, devido ao padrão de mortalidade diferente entre as idades jovens adultas e até sensivelmente à idade 40). A análise dos resultados sugere que, na maior parte das idades, o MBG tem em ambos os sexos, um desempenho melhor que o MGE, tanto ao nível dos ajustamentos como das previsões (exceção para algumas idades, sobretudo depois da idade 85, em que esta tendência se inverte).

Comparámos ainda os resultados obtidos através da aplicação dos modelos de EDEs univariados com os obtidos por outros métodos de análise transversal de dados (caso dos modelos de regressão não linear (RNL) e dos, também estocásticos, modelos autorregressivos integrados e de médias móveis (ARIMA)), no sentido de justificar a consistência da metodologia e validar os resultados da aplicação dos modelos, comparando, nomeadamente, a capacidade preditiva. Dos modelos alternativos aos de EDEs, os modelos ARIMA, com uma construção muito simples (sobretudo se compararmos com os sofisticados modelos longitudinais), revelaram-se muito adequados para descrever o comportamento das séries e para efetuar previsões a curto/médio prazo, por idade e sexo. Também os modelos de RNL têm um melhor desempenho no ajustamento aos dados (como era exetável, uma vez que estima por mínimos quadrados) que os modelos de EDEs. Apenas nas previsões para vários grupos de idades, quer os modelos ARIMA, quer os modelos de RNL têm um desempenho menos satisfatório (se atendermos, por exemplo, às diferenças entre os respetivos EQMs) que os modelos de EDEs.

Toda esta primeira fase de aplicação da metodologia foi preparatória à elaboração dos modelos multivariados, nos quais, como já referimos, criámos estruturas de correlação associadas aos processos de Wiener independentes para cada grupo etário, que designámos por idade. Refira-se ainda que, seja no movimento browniano geométrico bidimensional (MBGB), no modelo de Gompertz estocástico bidimensional (MGEB), ou no movimento browniano geométrico multidimensional (MBGM), para além da abordagem analítica referida para os modelos unidimensionais, desenhámos os respetivos métodos de estimação de parâmetros. Os resultados para os modelos bivariados revelam, comparativamente com os correspondentes modelos univariados, vantagens na modelação para as primeiras idades e também em quase todas as idades depois da idade 50, no caso do modelo de Gompertz, vantagens essas que também ocorrem no caso do movimento browniano geométrico. No comparativo entre os dois, o movimento browniano geométrico bidimensional (MBGB) apresenta um melhor desempenho global. Refira-se a propósito que, para estes modelos (e à semelhança do que fizemos para os casos unidimensionais), comparámos ainda os resultados com um modelo do tipo vetores autorregressivos e de médias móveis (VARMA), o que permite perceber como é que as variáveis se relacionam e com que desfasamento de tempo.

Os modelos do tipo VARMA apresentam, no entanto, algumas limitações (por exemplo, porque

se o desfasamento for grande, necessitam de um elevado número de parâmetros, o que não é desejável). Ainda assim, consideramos que são uma ferramenta muito interessante para explorar as relações dinâmicas entre este tipo de variáveis demográficas e a sua aplicação pode fornecer pistas para a construção de modelos dinâmicos afins.

No caso do MBGM, a maior parte das versões do modelo global mostra-se mais vantajosa até à idade 55-59, sendo que algumas versões conseguem captar a variabilidade do fenómeno, de modo satisfatório, até ao grupo etário 74-79. Para ambos os modelos, os erros das previsões aumentam com a idade, em ambos os sexos, e são superiores no sexo masculino, o que vai ao encontro dos resultados obtidos através dos modelos unidimensionais.

Como em todos os estudos, contudo, existiram limitações. Era nosso objetivo inicial aplicar o MBGM a todas as idades do arco de vida, não por grupo etário quinquenal (muito embora seja a representação usual em demografia e sirva à nossa finalidade), mas anual. Tentámos resolver o problema computacional que o impediu e que acreditamos se deveu ao elevado número de dimensões envolvidas e à natureza do cálculo, mas tal não foi possível. Trata-se de um desafio para ser analisado, eventualmente, através de outras técnicas de análise numérica ou de outros recursos informáticos, que consideramos não cabem neste formato. Seria também interessante verificar qual a tendência de mortalidade para depois de 2009 e comparar essa informação com as previsões, quer dos modelos de EDEs unidimensionais, quer dos multivariados, mas os dados de 2010 a 2012, embora estivessem disponíveis no *site* do Instituto Nacional de Estatística, não estavam ainda no *site* do *Human Mortality Database* (HMD) e não podemos usar os primeiros, porque a estimativa da população média anual, por idade e por sexo, não corresponde à utilizada pelo HMD (sofre um ligeiro alisamento e, como não está claro através das notas explicatórias qual foi o critério, optámos por não os usar).

Pelo exposto, consideramos que os resultados apresentados respondem à problemática inicial que nos propusemos estudar, isto é, os modelos de EDEs conseguiram captar a variabilidade do fenómeno da mortalidade, numa perspetiva transversal ao longo do tempo, e fazer previsões. Mais, parece-nos que a metodologia é bastante promissora para este tipo de dados. Relativamente ao modelo multivariado proposto, não nos preocupámos em usar uma expressão fechada, como no caso unidimensional, escolhendo a melhor versão do modelo global para todas as idades e ambos os sexos, pois, efetivamente, isso seria limitador em termos de trabalho futuro e reduziria o potencial de melhoria do desempenho do modelo. Uma opção de melhoria do modelo global agora proposto, pode passar por usar diferentes estruturas de correlação para diferentes grupos etários, isto é, em que a estimação dos parâmetros varia consoante as várias fases do arco de vida (fazendo sentido, provavelmente, um modelo trifásico na idade, marcado pelas diferenças no regime da mortalidade entre as primeiras idades, a maior parte das idades jovens-adultas e as idades mais avançadas). Aplicar também o MBGM com versões diferentes para as correlações, para grandes grupos de idades de sexos diferentes e construir, por exemplo, um modelo por ramos também é uma possibilidade, em particular para trabalhar com os grupos de idades anuais e mesmo para melhorar as previsões nas idades mais avançadas.

Por último, e em termos de trabalho futuro, consideramos que, naturalmente, o passo a seguir poderia passar por aplicar os modelos a dados da população de outros países com horizontes temporais distintos e analisar o seu desempenho. Tal exercício poderia também incluir variações nos modelos agora propostos e aplicados, como a introdução de mudanças de regime ou saltos (para melhor captar valores extremos de mortalidade). Outro aspeto que consideramos de interesse seria estudar o fenómeno na perspectiva de grandezas derivadas ou complementares da mortalidade, como sejam a esperança de vida, por exemplo, ou estudar o tempo que uma taxa de mortalidade demora a atingir um dado valor de referência.

Adicionalmente, podemos também considerar incluir preditores externos ao sistema das taxas de mortalidade nos modelos estudados. Porque era um dos objetivos secundários do plano inicial desta tese, este é já um trabalho em curso. Fizemos um estudo exploratório com variáveis económicas e climáticas. Das primeiras estudámos, entre outras, a relação entre a TBMs e a taxa de crescimento do produto interno bruto (PIB) *per capita* a preços constantes ou com o rendimento nacional bruto em % de PIB. No caso do clima, tratámos os dados disponíveis nas várias estações meteorológicas do país, coincidentes com o período em análise para as TBMs, através da técnica de análise de componentes principais e, a partir daí, estimámos, por exemplo, as taxas de variação anual das médias diárias das temperaturas máximas, das mínimas ou das amplitudes (fizemos esse exercício também apenas para os meses de verão e inverno). Em todo o caso, a curto prazo, não se encontra relação com significância estatística que justifique a inclusão destas variáveis no modelo.

Em termos de outras potencialidades, é ainda possível a produção de interfaces de suporte para a computação, designadamente, um pacote estatístico para o R, e acreditamos que o tipo de resultados produzidos na aplicação da metodologia possa ser útil na análise conjunta com os produzidos noutras áreas do conhecimento, a que a mortalidade possa interessar, que sejam igualmente obtidos através de uma análise transversal ao longo do tempo e que incorporem também algum tipo de incerteza inerente a uma variabilidade ambiental em sentido lato.

Era nosso objetivo inicial explicar a tendência evolutiva da mortalidade da população portuguesa. Como é esse o pressuposto dos modelos aplicados e consideramos os resultados da aplicação da metodologia satisfatórios, neste ponto aceitamos a hipótese de que existe uma ou mais variáveis (não sabemos quais, porque podem ou não ser observáveis) que são suscetíveis de afetar a probabilidade de ocorrer o óbito de um grupo de indivíduos (da mesma idade ou de idades diferentes e do mesmo sexo ou de sexos diferentes) num certo período de tempo. Acreditamos que o aperfeiçoamento deste tipo de modelos passa por extraírmos ainda mais informação dos dados das populações em estudo, tornando a estimação dos parâmetros mais flexível e assim melhorando o seu desempenho em geral.



Código do R: Aspectos gerais

```

1 ##PREPARAÇÃO DOS DADOS##
2
3 #Ler a base de dados
4 dados <- read.table("D:\TBM1x1.csv",sep=";")
5 attach(dados)
6
7 #Organizar os dados
8 TP <- dados[,1] #Tempo: 1940–2009
9 TaxasP <- dados[,2:dim(dados)[2]] #TBM por idade: colunas 2–101 sexo feminino; 102–201 masculino
10 IP <- dim(TaxasP)[1]
11 cP <- dim(TaxasP)[2]
12
13 #Definir o horizonte temporal para as previsões: 10 e 25 anos
14 nPREV <- 10 #corresponde ao período 2000 a 2009
15 nPREV2 <- 25 #corresponde ao período 2000 a 2024
16 Tfinal <- c(T,c(2000:2024)) #Criar vetor para as previsões a 25 anos
17 tf <- length(Tfinal)
18
19 #Selecionar os dados para o ajustamento (deixamos 10 anos de fora para estimar EQM)
20 T <- dados[1:(IP-nPREV),1]
21 Taxas <- dados[1:(IP-nPREV),2:201]
22 l <- dim(Taxas)[1]
23 C <- dim(Taxas)[2]
24 n <- l-1
25
26 #Identificar os 10 anos de dados não considerados no ajustamento
27 Tprev <- TP[1:(l+nPREV)]
28 taxas.prev <- TaxasP[1:(l+nPREV),]

```

NOTAS SOBRE O CÓDIGO IMPORTADO DA CONSOLA DO R:

- o conteúdo da caixa ##PREPARAÇÃO DOS DADOS##, acima, vai remeter para todas os outros apêndices;
- existem atribuições nas caixas dos apêndices relativos aos modelos multidimensionais, MGEB e MBGM (respetivamente, nos apêndices D e E), que remetem para as dos modelos unidimensionais base MBG e MGE (respetivamente, nos apêndices B e C);
- consoante a família de modelos base seja MBG ou MGE existem atribuições com a mesma designação que têm significados diferentes (por exemplo, Y ou YP, para as matrizes dos dados transformados), donde, por esse motivo, com exceção do referido no item anterior, cada apêndice deve ser considerado isoladamente para que não existam ambiguidades.

B

Código do R: MBG

```

1 ##MGB: ESTIMATIVAS DOS PARÂMETROS DO MODELO UNIVARIADO, POR IDADE, POR SEXO##
2
3 #Aplicar transformação, usando a escala logarítmica, às TBM#
4 YP <- matrix(nrow = (l + 10), ncol = C)
5 for (p in 1:(l + 10)){
6     for (q in 1:C){
7         YP[p,q] <- log(TaxasP[p,q] / TaxasP[1,q])
8     }
9 }
10
11 #Selecionar os dados transformados para ajustar o modelo
12 Y <- YP[1:(l+10)-nPREV,]
13
14 #R = alpha - (sigma^2) / 2
15 R <- vector(length=C)
16 for (i in 1:C){
17     R[i] <- Y[l,i] / l
18 }
19
20 #V = sigma^2
21 V <- vector(length = C)
22 for (i in 1:C){
23     V[i] <- (1 / n) * sum((Y[(2:l),i] - Y[(1:(l-1)),i] - R[i])^2)
24 }
25
26 #Estimar o parâmetro alpha do MGB
27 alpha <- vector(length = C)
28 for (i in 1:C){
29     alpha[i] <- R[i] + V[i] / 2
30 }
31
32 #Estimar o parâmetro sigma do MGB
33 sig <- vector(length = C)
34 for (i in 1:C){
35     sig[i] <- sqrt(V[i])
36 }
37
38 #Criar tabela com estimativas dos parâmetros
39 p.MBG <- cbind(R, V, alpha, sig)

```

```

1 ##MGB: AJUSTAMENTO##
2
3 #Ajustamento a LP: estimar as TBM em escala logarítmica
4 Y.ajust.MBG <- matrix(nrow = l, ncol = C)
5 for (p in 1:l){
6     for (q in 1:C){
7         Y.ajust.MBG[1,q] <- Y[1,q]
8     }
9 }
10 for (p in 2:l){
11     for (q in 1:C){
12         Y.ajust.MBG[p,q] <- Y[1,q] + R[q] * (T[p] - T[1])
13     }
14 }
15
16 #Inverter a transformação Y para os valores ajustados a LP
17 Tx.ajust.MBG <- matrix(nrow = l, ncol = C)
18 for (p in 1:l){
19     for (q in 1:C){
20         Tx.ajust.MBG[p,q] <- Taxas[1,q] * exp(Y.ajust.MBG[p,q])
21     }
22 }
23
24 # Estimar o EQM do ajustamento a LP
25 EQM.ajust.MBG <- vector(length = l)
26 erro.ajust.MBG <- matrix(nrow = (l + 1), ncol = C)
27 for (p in 1:l) {
28     for (q in 1:C){
29         erro.ajust.MBG[p,q] <- ((Tx.ajust.MBG[p,q] - Taxas[p,q])^2)
30     }
31 }
32 for (p in 1:l) {
33     for (q in 1:C){
34         erro.ajust.MBG[(l+1),q] <- sum(erro.ajust.MBG[(1:p),q]) / l
35     }
36 }
37 EQM.ajust.MBG <- erro.ajust.MBG[(l+1),]
38
39 #Ajustamento PP: estimar as TBM em escala logarítmica
40 Y.ajust.MBG.PP <- matrix(nrow = l, ncol = C)
41 for (p in 1:l){
42     for (q in 1:C){

```

```

43     Y.ajust.MBG.PP[1,q] <- Y[1,q]
44   }
45 }
46 for (p in 2:l){
47   for (q in 1:C){
48     Y.ajust.MBG.PP[p,q] <- Y[p-1,q] + R[q] * (T[p] - T[p-1])
49   }
50 }
51
52 #Inverter a transformação Y para os valores ajustados PP
53 Tx.ajust.MBG.PP <- matrix(nrow = l, ncol = C)
54 for (p in 1:l){
55   for (q in 1:C){
56     Tx.ajust.MBG.PP[p,q] <- Taxas[1,q] * exp(Y.ajust.MBG.PP[p,q])
57   }
58 }
59
60 # Estimar o EQM do ajustamento PP
61 EQM.ajust.MBG.PP <- vector(length = l)
62 erro.ajust.MBG.PP <- matrix(nrow = (l + 1), ncol = C)
63 for (p in 1:l) {
64   for (q in 1:C){
65     erro.ajust.MBG.PP[p,q] <- ((Tx.ajust.MBG.PP[p,q] - Taxas[p,q])^2)
66   }
67 }
68 for (p in 1:l) {
69   for (q in 1:C){
70     erro.ajust.MBG.PP[(l+1),q] <- sum(erro.ajust.MBG.PP[(1:p),q]) / l
71   }
72 }
73 EQM.ajust.MBG.PP <- erro.ajust.MBG.PP[(l+1),]

```

```

1 ##MGB: PREVISÕES##
2
3 #Previsões a LP: 10 anos
4 Y.prev.LP.MBG <- matrix(nrow = (nPREV + 1), ncol = C)
5 for (p in 1:(nPREV+1)){
6     for (q in 1:C){
7         Y.prev.LP.MBG[1,q] <- YP[l,q]
8     }
9 }
10 for (p in 2:(nPREV+1)){
11     for (q in 1:C){
12         Y.prev.LP.MBG[p,q] <- YP[l,q] + R[q] * (TP[(l + p - 1)] - TP[l])
13     }
14 }
15
16 #Inverter a transformação Y para as previsões a LP: 10 anos
17 Tx.prev.LP.MBG <- matrix(nrow = (nPREV + 1), ncol = C)
18 for (p in 1:(nPREV+1)){
19     for (q in 1:C){
20         Tx.prev.LP.MBG[p,q] <- TaxasP[1,q] * exp(Y.prev.LP.MBG[p,q])
21     }
22 }
23
24 #Estimar o EQM das previsões a LP: 10 anos
25 EQM.prev.LP.MBG <- vector(length = (nPREV + 1))
26 erro.prev.LP.MBG <- matrix(nrow = (nPREV + 2), ncol = C)
27 for (p in 1:(nPREV+1)) {
28     for (q in 1:C){
29         erro.prev.LP.MBG[p,q] <- ((Tx.prev.LP.MBG[p,q] - taxas.prev[p,q])^2)
30     }
31 }
32 for (p in 1:(nPREV+1)) {
33     for (q in 1:C){
34         erro.prev.LP.MBG[(nPREV+2),q] <- sum(erro.prev.LP.MBG[(1:p),q]) / (nPREV + 1)
35     }
36 }
37 EQM.prev.LP.MBG <- erro.prev.LP.MBG[(nPREV + 2),]
38
39 #Previsões a LP: 25 anos
40 Y.prev.LP25.MBG <- matrix(nrow=(nPREV2+1),ncol = C)
41 for (p in 1:1){
42     for (q in 1:C){

```

```

43     Y.prev.LP25.MBG[1,q] <- YP[l,q]
44   }
45 }
46 for (p in 2:(nPREV2+1)){
47   for (q in 1:C){
48     Y.prev.LP25.MBG[p,q] <- YP[l,q]+R[q]*(Tfinal[(l+p-1)]-Tfinal[l])
49   }
50 }
51
52 #Inverter a transformação Y para as previsões a LP: 25 anos
53 Tx.prev.LP25.MBG <- matrix(nrow = (nPREV2 + 1), ncol = C)
54 for (p in 1:(nPREV2+1)){
55   for (q in 1:C){
56     Tx.prev.LP25.MBG[p,q] <- TaxasP[1,q] * exp(Y.prev.LP25.MBG[p,q])
57   }
58 }
59
60 #Previsões PP: 10 anos
61 IP.n <- (dim(TaxasP)[1]) - 1
62 k <- IP.n - nPREV
63
64 IP.ZP <- YP
65
66 IP.R <- matrix(nrow = nPREV, ncol = C)
67 IP.V <- matrix(nrow = nPREV, ncol = C)
68 IP.alpha <- matrix(nrow = nPREV, ncol = C)
69 IP.sig <- matrix(nrow = nPREV, ncol = C)
70
71 Y.prev.1p.MBG <- matrix(nrow = (nPREV + 1), ncol = C)
72 for (p in 1:(nPREV+1)){
73   for (q in 1:C){
74     Y.prev.1p.MBG[1,q] <- YP[(k+1),q]
75   }
76 }
77
78 for (i in 1:C){
79   #Atualizar anualmente as previsões
80   for (j in 1:(nPREV)){
81     IP.R[j,i] <- IP.YP[(k+j),i] / (k+j)
82     IP.V[j,i] <- (1 / (k + j)) * sum((IP.YP[(2:(k+j)),i] - IP.YP[(1:(k+j-1)),i] - IP.R[j,i])^2)
83     IP.alpha[j,i] <- IP.R[j,i] + IP.V[j,i] / 2
84     IP.sig[j,i] <- sqrt(IP.V[j,i])

```



```

85
86     #Previsões PP
87     Y.prev.1p.MBG[(j+1),i] <- YP[(k+j),i] + IP.R[j,i]
88   }
89 }
90
91 #Inverter a transformação Y para as previsões PP: 10 anos
92 Tx.prev.1p.MBG <- matrix(nrow = (nPREV+1),ncol = C)
93 for (p in 1:(nPREV+1)){
94   for (q in 1:C){
95     Tx.prev.1p.MBG[p,q] <- TaxasP[1,q] * exp(Y.prev.1p.MBG[p,q])
96   }
97 }
98
99 #Estimar o EQM das previsões PP: 10 anos
100 EQM.prev.1p.MBG <- vector(length = (nPREV + 1))
101 erro.prev.1p.MBG <- matrix(nrow = (nPREV + 2), ncol = C)
102 for (p in 1:(nPREV+1)) {
103   for (q in 1:C){
104     erro.prev.1p.MBG[p,q] <- ((Tx.prev.1p.MBG[p,q] - taxas.prev[p,q])^2)
105   }
106 }
107 for (p in 1:(nPREV+1)) {
108   for (q in 1:c){
109     erro.prev.1p.MBG[(nPREV+2),q] <- sum(erro.prev.1p.MBG[(1:p),q]) / (nPREV + 1)
110   }
111 }
112 EQM.prev.1p.MBG <- erro.prev.1p.MBG[(nPREV+2),]

```

```

1 ## TRANSFORMAR PREVISÕES (10 anos) PARA COORTE ##
2 #Previsões, na escala original, obtidas a partir do MBG
3 #Coorte dos indivíduos que nasceram no ano de 1999
4
5 #Organizar previsões por coorte, para o sexo feminino
6 prev.COORTE.01F <- vector(length=nPREV)
7 for (i in 1:nPREV){
8     prev.COORTE.01F[i] <- Tx.prev.LP.MBG[i+1,i]
9 }
10
11 #Organizar previsões por coorte, para o sexo masculino
12 prev.COORTE.01M <- vector(length=nPREV)
13 for (i in 1:nPREV){
14     prev.COORTE.01M[i] <- Tx.prev.LP.MBG[i+1,i+(C/2)]
15 }

```

C

Código do R: MGE

```

1 ##MGE: ESTIMATIVAS DOS PARÂMETROS DO MODELO UNIVARIADO, POR IDADE, POR SEXO##
2
3 #Aplicar transformação logarítmica às TBM#
4 YP <- log(TaxasP)
5
6 #Selecionar os dados transformados para ajustar o modelo
7 Y <- YP[1:(IP-nPREV),]
8
9 #Criar matriz com desfasamento de 1 ano
10 Yt <- matrix(nrow = n, ncol = C)
11 for (p in 1:n){
12     for (q in 1:C){
13         Yt[p,q] <- Y[(p+1),q]
14     }
15 }
16 Yt_1 <- matrix(nrow = n, ncol = C)
17 for (p in 1:n){
18     for (q in 1:C){
19         Yt_1[p,q] <- Y[p,q]
20     }
21 }
22
23 #Estimar parâmetros por MV: A = Ln(a), b e sigma
24 parametros.MGE <- vector(length = C)
25 library(stats)
26 for (i in 1:C){
27     f.A <- function(beta){
28         ((n * (1 - exp(-beta)))^(-1)) * (sum(Yt[,i] - Yt_1[,i] * exp(-beta)))
29     }
30     f.sigma <- function(Alpha,beta){
31         (((2*beta) * ((n * (1 - exp(-2 * beta)))^(-1))) *
32         sum((Yt[,i] - Alpha - (Yt_1[,i] - Alpha) * exp(-beta))^2))^(1/2)
33     }
34     f.LogV <- function(beta){
35         ((n / 2) * log(2 * pi)) + ((n / 2) * log(((f.sigma(f.A(beta), beta))^2) / (2 * beta))) +
36         ((n / 2) * log(1 - exp(-2 * beta))) + ((beta * (((f.sigma(f.A(beta), beta))^2) *
37         (1 - exp(-2 * beta)))^(-1))) * (sum((Yt[,i] - f.A(beta) - (Yt_1[,i] - f.A(beta)) * exp(-beta))^2)))
38     }
39     parametros.MGE[i] <- optimize(f.LogV, int = c(0.0001, 10))$min
40 }
41
42 A <- vector(length = C)

```

```
43 b <- vector(length = C)
44 sigma <- vector(length = C)
45 a <- vector(length = C)
46
47 for (i in 1:C){
48     b[i] <- parametros.MGB[i]
49     A[i] <- f.A(b[i])
50     sigma[i] <- f.sigma(f.A(b[i]), b[i])
51     a[i] <- exp(A[i])
52 }
53
54 #Criar tabela com estimativas dos parâmetros
55 p.MGB <- cbind(A, b, sigma, a)
```

```

1 ##MGE: AJUSTAMENTO##
2
3 #Ajustamento a LP: estimar os logaritmos das TBM
4 Y.ajust.MGE <- matrix(nrow = l, ncol = C)
5 for (p in 1:l){
6     for (q in 1:C){
7         Y.ajust.MGE[1,q] <- Y[1,q]
8     }
9 }
10 for (p in 2:l){
11     for (q in 1:C){
12         Y.ajust.MGE[p,q] <- A[q] + (Y[1,q] - A[q]) * exp(-b[q] * (T[p] - T[1]))
13     }
14 }
15
16 #Inverter a transformação Y para os valores ajustados a LP
17 Tx.ajust.MGE <- matrix(nrow = l, ncol = C)
18 for (p in 1:l){
19     for (q in 1:C){
20         Tx.ajust.MGE[p,q] <- exp(Y.ajust.MGE[p,q])
21     }
22 }
23
24 #Estimar o EQM do ajustamento a LP
25 EQM.ajust.MGE <- vector(length = l)
26 erro.ajust.MGE <- matrix(nrow = (l + 1), ncol = C)
27 for (p in 1:l) {
28     for (q in 1:C){
29         erro.ajust.MGE[p,q] <- ((Tx.ajust.MGE[p,q] - Taxas[p,q])^2)
30     }
31 }
32 for (p in 1:l) {
33     for (q in 1:C){
34         erro.ajust.MGE[(l+1),q] <- sum(erro.ajust.MGE[(1:p),q]) / l
35     }
36 }
37 EQM.ajust.MGE <- erro.ajust.MGE[(l+1),]
38
39 #Ajustamento PP: estimar os logaritmos das TBM
40 Y.ajust.MGE.PP <- matrix(nrow = l, ncol = C)
41 for (p in 1:l){
42     for (q in 1:C){

```

```

43     Y.ajust.MGE.PP[1,q] <- Y[1,q]
44   }
45 }
46 for (p in 2:l){
47   for (q in 1:C){
48     Y.ajust.MGE.PP[p,q] <- A[q] + (Y[p-1,q] - A[q]) * exp(-b[q] * (T[p] - T[p-1]))
49   }
50 }
51
52 #Inverter a transformação Y para os valores ajustados PP
53 Tx.ajust.MGE.PP <- matrix(nrow = l, ncol = C)
54 for (p in 1:l){
55   for (q in 1:C){
56     Tx.ajust.MGE.PP[p,q] <- exp(Y.ajust.MGE.PP[p,q])
57   }
58 }
59
60 #Estimar o EQM do ajustamento PP
61 EQM.ajust.MGE.PP <- vector(length = l)
62 erro.ajust.MGE.PP <- matrix(nrow = (l + 1), ncol = C)
63 for (p in 1:l) {
64   for (q in 1:C){
65     erro.ajust.MGE.PP[p,q] <- ((Tx.ajust.MGE.PP[p,q] - Taxas[p,q])^2)
66   }
67 }
68 for (p in 1:l) {
69   for (q in 1:C){
70     erro.ajust.MGE.PP[(l+1),q] <- sum(erro.ajust.MGE.PP[(1:p),q]) / l
71   }
72 }
73 EQM.ajust.MGE.PP <- erro.ajust.MGE.PP[(l+1),]

```

```

1 ##MGE: PREVISÕES##
2
3 #Previsões a LP: 10 anos
4 Y.prev.LP.MGE <- matrix(nrow = (nPREV + 1), ncol = C)
5 for (p in 1:(nPREV+1)){
6     for (q in 1:C){
7         Y.prev.LP.MGE[1,q] <- YP[l,q]
8     }
9 }
10 for (p in 2:(nPREV+1)){
11     for (q in 1:C){
12         Y.prev.LP.MGE[p,q] <- A[q] + (YP[l,q] - A[q]) * exp(-b[q] * (TP[(l + p - 1)] - TP[l]))
13     }
14 }
15
16 #Inverter a transformação Y para as previsões a LP: 10 anos
17 Tx.prev.LP.MGE <- matrix(nrow = (nPREV+1), ncol = C)
18 for (p in 1:(nPREV+1)){
19     for (q in 1:C){
20         Tx.prev.LP.MGE[p,q] <- exp(Y.prev.LP.MGE[p,q])
21     }
22 }
23
24 #Estimar o EQM das previsões a LP: 10 anos
25 EQM.prev.LP.MGE <- vector(length = (nPREV+1))
26 erro.prev.LP.MGE <- matrix(nrow = (nPREV + 2), ncol = C)
27 for (p in 1:(nPREV+1)) {
28     for (q in 1:C){
29         erro.prev.LP.MGE[p,q] <- ((Tx.prev.LP.MGE[p,q] - taxas.prev[p,q])^2)
30     }
31 }
32 for (p in 1:(nPREV+1)) {
33     for (q in 1:C){
34         erro.prev.LP.MGE[(nPREV+2),q] <- sum(erro.prev.LP.MGE[(1:p),q]) / (nPREV)
35     }
36 }
37 EQM.prev.LP.MGE <- erro.prev.LP.MGE[(nPREV+2),]
38
39 #Previsões a LP: 25 anos
40 Y.prev.LP25.MGE <- matrix(nrow = (nPREV2+1), ncol = C)
41 for (p in 1:1){
42     for (q in 1:C){

```



```

43     Y.prev.LP25.MGE[1,q] <- YP[l,q]
44   }
45 }
46 for (p in 2:(nPREV2+1)){
47   for (q in 1:C){
48     Y.prev.LP25.MGE[p,q] <- A[q] + (YP[l,q] - A[q]) * exp(-b[q] * (Tfinal[(l + p - 1)] - Tfinal[l]))
49   }
50 }
51
52 #Inverter a transformação Y para as previsões a LP: 25 anos
53 Tx.prev.LP25.MGE <- matrix(nrow = (nPREV2 + 1), ncol = C)
54 for (p in 1:(nPREV2+1)){
55   for (q in 1:C){
56     Tx.prev.LP25.MGE[p,q] <- exp(Y.prev.LP25.MGE[p,q])
57   }
58 }
59
60 #Previsões PP: 10 anos
61 IP.n <- (dim(YP)[1]) - 1
62
63 IP.Yt <- matrix(nrow = IP.n, ncol = C)
64 for (p in 1:IP.n){
65   for (q in 1:C){
66     IP.Yt[p,q] <- YP[(p+1),q]
67   }
68 }
69 IP.Yt_1 <- matrix(nrow = IP.n, ncol = C)
70 for (p in 1:IP.n){
71   for (q in 1:C){
72     IP.Yt_1[p,q] <- YP[p,q]
73   }
74 }
75
76 k <- IP.n - nPREV
77
78 IP.b <- matrix(nrow = nPREV, ncol = C)
79 IP.A <- matrix(nrow = nPREV, ncol = C)
80 IP.sigma <- matrix(nrow = nPREV, ncol = C)
81 IP.a <- matrix(nrow = nPREV, ncol = C)
82
83 IP.parametro.MGE <- matrix(nrow = nPREV, ncol = C)
84

```

```

85 IP.Y.prev.1p.MGE <- matrix(nrow = (nPREV + 1), ncol = C)
86 for (p in 1:(nPREV+1)){
87     for (q in 1:C){
88         IP.Y.prev.1p.MGE[1,q] <- YP[k+1,q]
89     }
90 }
91
92 for (i in 1:C){
93     #Atualizar anualmente as previsões
94     for (j in 1:(nPREV)){
95
96         IP.f.A <- function(IP.beta){
97             (((k + j) * (1 - exp(-IP.beta)))^(-1)) * (sum(IP.Yt[1:(k+j),i] -
98                 IP.Yt_1[1:(k+j),i] * exp(-IP.beta)))
99         }
100         IP.f.sigma <- function(IP.Alpha,IP.beta){
101             ((2*IP.beta) * (((k + j) * (1 - exp(-2 * IP.beta)))^(-1))) *
102             sum((IP.Yt[1:(k+j),i] - IP.Alpha - (IP.Yt_1[1:(k+j),i] -
103                 IP.Alpha) * exp(-IP.beta))^2)^(1/2)
104         }
105         IP.f.LogV <- function(IP.beta){
106             (((k + j) / 2) * log(2 * pi)) + (((k + j) / 2) *
107             log(((IP.f.sigma(IP.f.A(IP.beta), IP.beta))^2) / (2 * IP.beta))) +
108             (((k + j) / 2) * log(1 - exp(-2 * IP.beta))) +
109             ((IP.beta * (((IP.f.sigma(IP.f.A(IP.beta), IP.beta))^2) *
110                 (1 - exp(-2 * IP.beta)))^(-1))) * (sum((IP.Yt[1:(k+j),i] -
111                 IP.f.A(IP.beta) - (IP.Yt_1[1:(k+j),i] - IP.f.A(IP.beta)) * exp(-IP.beta))^2)))
112         }
113         IP.parametro.MGE[j,i] <- optimize(IP.f.LogV, int=c(0.0001,10))$min
114
115         #Estimativas dos parâmetros: A=Ln(a), b e sigma, para c=0 e Y=Ln(X)
116         IP.b[j,i] <- IP.parametro.MGE[j,i]
117         IP.A[j,i] <- IP.f.A(IP.b[j,i])
118         IP.sigma[j,i] <- IP.f.sigma(IP.f.A(IP.b[j,i]), IP.b[j,i])
119         IP.a[j,i] <- exp(IP.A[j,i])
120
121
122         #Previsões PP
123         IP.Y.prev.1p.MGE[(j+1),i] <- IP.A[j,i] + (YP[(k+j),i] - IP.A[j,i]) * exp(-IP.b[j,i])
124
125     }
126 }

```

```

127
128 #Inverter a transformação Y para as previsões PP: 10 anos
129 Tx.prev.1p.MGE <- matrix(nrow = (nPREV+1), ncol = C)
130 for (p in 1:(nPREV+1)){
131     for (q in 1:C){
132         Tx.prev.1p.MGE[p,q] <- exp(IP.Y.prev.1p.MGE[p,q])
133     }
134 }
135
136 #Estimar o EQM das previsões PP: 10 anos
137 EQM.prev.1p.MGE <- vector(length = (nPREV + 1))
138 erro.prev.1p.MGE <- matrix(nrow = (nPREV+2), ncol = C)
139 for (p in 1:(nPREV+1)) {
140     for (q in 1:C){
141         erro.prev.1p.MGE[p,q] <- ((Tx.prev.1p.MGE[p,q] - taxas.prev[p,q])^2)
142     }
143 }
144 for (p in 1:(nPREV+1)) {
145     for (q in 1:C){
146         erro.prev.1p.MGE[(nPREV+2),q] <- sum(erro.prev.1p.MGE[(1:p),q]) / (nPREV)
147     }
148 }
149 EQM.prev.1p.MGE <- erro.prev.1p.MGE[(nPREV+2),]

```


D

Código do R: MGEB

```

1 ##MGEb: ESTIMATIVAS DOS PARÂMETROS DO MODELO BIDIMENSIONAL, POR IDADE##
2
3 #Dividir a matriz de dados e de defasamentos anuais por sexo: Y1 feminino, Y2 masculino
4 Y1.t <- Yt[,1:(C/2)]
5 Y2.t <- Yt[,((C/2)+1):C]
6 Y1.t.1 <- Yt.1[,1:(C/2)]
7 Y2.t.1 <- Yt.1[,((C/2)+1):C]
8
9 #Criar o vetor de valores de arranque, por sexo
10 #(usar as estimativas dos parâmetros A, b, sigma do MGE para valores iniciais)
11 Abs1 <- p.MGE[1:(C/2),]
12 Abs2 <- p.MGE[((C/2)+1):C,]
13
14 #Criar sequência para valores iniciais do parâmetro de correlação, r
15 rI <- seq(-0.9, 0.9, 0.01)
16
17 #Estimar os parâmetros: A1 = Ln(a1), b1, sigma1, A2 = Ln(a2), b2, sigma2, ro.Biv = (ro.Wien *
18 (1 - exp(-b1 - b2)) * (((1 - exp(-2 * b1)) * (1 - exp(-2 * b2)))^(-1/2)) * 2 * (((b1 * b2)^(1/2)) / (b1 + b2)))
19
20 A1 <- vector(length = (C/2))
21 a1 <- vector(length = (C/2))
22 b1 <- vector(length = (C/2))
23 sigma1 <- vector(length = (C/2))
24 ro.Biv <- vector(length = (C/2))
25 ro.Wien <- vector(length = (C/2))
26 A2 <- vector(length = (C/2))
27 a2 <- vector(length = (C/2))
28 b2 <- vector(length = (C/2))
29 sigma2 <- vector(length = (C/2))
30 valorV.ro <- vector(length = (C/2))
31 conv.ro <- vector(length = (C/2))
32 m1 <- vector(length = (C/2))
33 m2 <- vector(length = (C/2))
34 M.CoVar <- list(length=(C/2))
35
36 #Fazer mudança de variável para simplificar operações com logaritmos
37 m1 <- vector(length = (C/2))
38 m2 <- vector(length = (C/2))
39 for (i in 1:(C/2)){
40     m1[i] <- ((1 - exp(-2 * Abs1[i,2])) * ((Abs1[i,3]^2)) / (2 * Abs1[i,2]))
41     m2[i] <- ((1 - exp(-2 * Abs2[i,2])) * ((Abs2[i,3]^2)) / (2 * Abs2[i,2]))
42 }

```

```

43
44 library(stats)
45 library(numDeriv)
46 for (i in 1:(C/2)){
47     j <- 1
48     parametros.MGEB <- function(w){ #w[1]=A1;w[2]=b1;w[3]=m1;w[4]=ro.Biv;w[5]=A2;w[6]=b2;w[7]=m2
49         n * log(2 * pi) + (n/2) * log((1 - (w[4]^2)) * w[3] * w[7]) + ((2 * (1 - (w[4]^2)))^(-1)) *
50         ((1/w[3]) * sum((Y1.f[,i] - w[1] - (Y1.t.1[,i] - w[1]) * exp(-w[2]))^2) + (1 / w[7]) *
51         sum((Y2.f[,i] - w[5] - (Y2.t.1[,i] - w[5]) * exp(-w[6]))^2) - 2 * w[4] * ((w[3] * w[7])^(-1/2)) *
52         sum((Y1.f[,i] - w[1] - (Y1.t.1[,i] - w[1]) * exp(-w[2])) * (Y2.f[,i] - w[5] - (Y2.t.1[,i] - w[5]) *
53         exp(-w[6]))))
54     }
55     f.LogV.Biv <- nlm(b(c(Abs1[i,1], Abs1[i,2], m1[i], r1[j], Abs2[i,1], Abs2[i,2], m2[i]), parametros.MGEB)
56     A.1[i] <- minbiv$par[1]
57     b.1[i] <- minbiv$par[2]
58     m1[i] <- minbiv$par[3]
59     ro.Biv[i] <- minbiv$par[4]
60     A.2[i] <- minbiv$par[5]
61     b.2[i] <- minbiv$par[6]
62     m2[i] <- minbiv$par[7]
63     valorV.ro[i] <- minbiv$objective
64     conv.ro[i] <- minbiv$convergence
65     a.1[i] <- exp(A.1[i])
66     sigma.1[i] <- ((2 * b.1[i] * m1[i]) / (1 - exp(-2 * b.1[i])))^(1/2)
67     a.2[i] <- exp(A.2[i])
68     sigma.2[i] <- ((2 * b.2[i] * m2[i]) / (1 - exp(-2 * b.2[i])))^(1/2)
69     ro.Wien[i] <- (ro.Biv[i]) / ((1 - exp(-b.1[i] - b.2[i])) * (((1 - exp(-2 * b.1[i])) *
70     (1 - exp(-2 * b.2[i])))^(-1/2)) * 2 * (((b.1[i] * b.2[i])^(1/2)) / (b.1[i] + b.2[i])))
71     M.CoVar[[i]] <- solve(hessian(parametros.MGEB, c(A.1[i], b.1[i], sigma.r1[i], ro.Wien[i],
72     A.2[i], b.2[i], sigma.2[i])))
73     #Criar condição que obriga a experimentar valores iniciais de r para haver convergência
74     while (valor.V[i] == "NaN" | conv[i] != 0) {
75         j <- (j + 1)
76         parametros.MGEB <- function(w){
77             n * log(2 * pi) + (n / 2) * log((1 - (w[4]^2)) * w[3] * w[7]) + ((2 * (1 - (w[4]^2)))^(-1)) *
78             ((1 / w[3]) * sum((Y1.f[,i] - w[1] - (Y1.t.1[,i] - w[1]) * exp(-w[2]))^2) +
79             (1 / w[7]) * sum((Y2.f[,i] - w[5] - (Y2.t.1[,i] - w[5]) * exp(-w[6]))^2) -
80             2 * w[4] * ((w[3] * w[7])^(-1/2)) * sum((Y1.f[,i] - w[1] - (Y1.t.1[,i] - w[1]) *
81             exp(-w[2])) * (Y2.f[,i] - w[5] - (Y2.t.1[,i] - w[5]) * exp(-w[6]))))
82         }
83         f.LogV.Biv <- nlm(b(c(Abs1[i,1], Abs1[i,2], im1[i], r1[j], Abs2[i,1], Abs2[i,2], im2[i]),
84         parametros.MGEB)

```

```

85     A.1[i] <- minbiv$par[1]
86     b.1[i] <- minbiv$par[2]
87     m1[i] <- minbiv$par[3]
88     ro.Biv[i] <- minbiv$par[4]
89     A.2[i] <- minbiv$par[5]
90     b.2[i] <- minbiv$par[6]
91     m2[i] <- minbiv$par[7]
92     valorV.ro[i] <- minbiv$objective
93     conv.ro[i] <- minbiv$convergence
94     a.1[i] <- exp(A.1[i])
95     sigma.1[i] <- ((2 * b.1[i] * m1[i]) / (1 - exp(-2 * b.1[i])))^(1/2)
96     a.2[i] <- exp(A.2[i])
97     sigma.2[i] <- ((2 * b.2[i] * m2[i]) / (1 - exp(-2 * b.2[i])))^(1/2)
98     ro.Wien[i] <- (ro.Biv[i]) / ((1 - exp(-b.1[i] - b.2[i])) * (((1 - exp(-2 * b.1[i])) *
99     (1 - exp(-2 * b.2[i]))^(-1/2)) * 2 * ((b.1[i] * b.2[i])^(1/2)) / (b.1[i] + b.2[i])))
100     M.CoVar[[i]] <- solve(hessian(parametros.MGEB, c(A.1[i], b.1[i], sigma.r1[i], ro.Wien[i],
101     A.2[i], b.2[i], sigma.2[i])))
102   }
103 }
104
105 p.MGEB <- cbind(A.1, b.1, sigma.1, a.1, ro.Biv, ro.Wien, A.2, b.2, sigma.2, a.2)
106
107 A.Biv <- c(p.MGEB[,1], p.MGEB[,7])
108 b.Biv <- c(p.MGEB[,2], p.MGEB[,8])
109 sigma.Biv <- c(p.MGEB[,3], p.MGEB[,9])
110 a.Biv <- c(p.MGEB[,4], p.MGEB[,10])

```



```

1
2 ##TESTE RAZÃO DE VEROSIMILHANÇAS (MGE vs MGEb)##
3
4 G <- vector(length=(C/2))
5 for (i in 1:(C/2)){
6     G[i] <- -2 * (-valorV[i] + (valorV.ro[i]))
7 }
8
9 #Teste por idade (n=1): Qui-Quadrado(1;0.95) = 3.841
10 valor.p <- 3.841
11
12 #Resultado do teste
13 G.teste <- vector(length=(C/2))
14 for (i in 1:(C/2)){
15     if (G[i] > valor.p) {G.teste[i]<-"OK! MGEb é melhor!"} else {G.teste[i] <- "MGEb não é melhor!"}
16 }
17
18 #Teste conjunto para todas as idades (n=100): Qui-Quadrado(100;0.95) = 124.342
19
20 sum(G) #[1] 321.3642, MGEb é melhor globalmente
21 #[1] 1820.015

```


E

Código do R: MBGM

```

1 ##MÉTODO GRADIENTE DESCENDENTE##
2
3 #Função auxiliar para estimar o vetor R
4 get.R <- function(P,I,C=200){
5     R <- vector(length=C)
6     for (i in 1:C){
7         if (i <= (C/2)) {
8             R[i] <- P[1] + P[2]*I[i] + P[3]*I[i]^2 + P[4]*I[i]^3
9         } else {
10            R[i] <- P[5] + P[6]*I[(i-(C/2))] + P[7]*I[(i-(C/2))]^2 + P[8]*I[(i-(C/2))]^3
11        }
12    }
13    return(R)
14 }
15
16 #Função auxiliar para estimar a matriz Sigma (inverter transformações de sigmas e correlações)
17 #Exemplo para versão 9 do MBGM: 19 parâmetros
18 get.SIGMA <- function(P,I,C=200){
19     SIGMA.FF <- matrix(nrow=C/2,ncol=C/2)
20     #bloco 1: F*F (submatriz simétrica com diagonal=sigmaF^2)
21     for (p in 1:(C/2)){
22         for (q in 1:(C/2)){
23             if (p == q){
24                 SIGMA.FF[p,q] <- (exp((P[9] + P[10]*I[p] + P[11]*I[p]^2 + P[12]*I[p]^3)/2))^2
25             } else if (p < q){
26                 SIGMA.FF[p,q] <- (exp((P[9] + P[10]*I[p] + P[11]*I[p]^2 + P[12]*I[p]^3)/2))*
27                     (exp((P[9] + P[10]*I[q] + P[11]*I[q]^2 + P[12]*I[q]^3)/2))*
28                     ((exp(2*(P[17]*min(p,q)))-1)/(exp(2*(P[17]*min(p,q)))+1))
29             } else {
30                 SIGMA.FF[p,q] <- (exp((P[9] + P[10]*I[q] + P[11]*I[q]^2 + P[12]*I[q]^3)/2))*
31                     (exp((P[9] + P[10]*I[p] + P[11]*I[p]^2 + P[12]*I[p]^3)/2))*
32                     ((exp(2*(P[17]*min(p,q)))-1)/(exp(2*(P[17]*min(p,q)))+1))
33             }
34         }
35     }
36     #bloco 2: F*M (submatriz não simétrica com diagonal=sigmaF*sigmaM*róFM)
37     SIGMA.FM <- matrix(nrow=C/2,ncol=C/2)
38     for (p in 1:(C/2)){
39         for (q in 1:(C/2)){
40             SIGMA.FM[p,q] <- (exp((P[9] + P[10]*I[p] + P[11]*I[p]^2 + P[12]*I[p]^3)/2))*
41                 (exp((P[13] + P[14]*I[p] + P[15]*I[p]^2 + P[16]*I[p]^3)/2))*
42                 ((exp(2*(P[19]*q))-1)/(exp(2*(P[19]*q))+1))

```

```

43     }
44 }
45 #bloco 3: M*F (submatriz não simétrica com diagonal=sigmaM*sigmaF*róMF)
46 SIGMA.MF <- t(SIGMA.FM)
47
48 #bloco 4: M*M (submatriz simétrica com diagonal=sigmaM^2)
49 SIGMA.MM <- matrix(nrow=C/2, ncol=C/2)
50 for (p in 1:(C/2)){
51     for (q in 1:(C/2)){
52         if (p == q){
53             SIGMA.MM[p,q] <- (exp((P[13] + P[14]*f[p] + P[15]*f[p]^2 + P[16]*f[p]^3)/2))^2
54         } else if (p < q){
55             SIGMA.MM[p,q] <- (exp((P[13] + P[14]*f[p] + P[15]*f[p]^2 + P[16]*f[p]^3)/2))*
56                 (exp((P[13] + P[14]*f[q] + P[15]*f[q]^2 + P[16]*f[q]^3)/2))*
57                 ((exp(2*(P[18]*min(p,q)))-1)/(exp(2*(P[18]*min(p,q)))+1))
58         } else {
59             SIGMA.MM[p,q] <- (exp((P[13] + P[14]*f[q] + P[15]*f[q]^2 + P[16]*f[q]^3)/2))*
60                 (exp((P[13] + P[14]*f[p] + P[15]*f[p]^2 + P[16]*f[p]^3)/2))*
61                 ((exp(2*(P[18]*min(p,q)))-1)/(exp(2*(P[18]*min(p,q)))+1))
62         }
63     }
64 }
65 SIGMA <- rbind(cbind(SIGMA.FF,SIGMA.FM), cbind(SIGMA.MF,SIGMA.MM))
66 return(SIGMA)
67 }
68
69 #Para a matriz inversa generalizada (aproximação à inversa)
70 #library(MASS)
71 #ginv(SIGMA)
72
73 #Função para eliminar, pontualmente, valor próprio negativo
74 del.neg.VP <- function(M){ #M=matriz retangular
75     Q <- nrow(M)
76     E <- eigen(M)$values
77     U <- eigen(M)$vectors
78     v <- as.numeric(E < 0)
79     m <- sum(v) # nº de VP negativos
80     if(m > 0){
81         G <- sum(v*E)*2
82         W <- (G*G*100)+1
83         P <- E[Q - m] # o mais pequeno valor positivo
84         h <- Q - m + 1

```

```

85     for(i in h:Q){
86         B <- E[i]
87         E[i] <- P * (G-B)*(G-B)/W
88     }
89     HC <- U %*% diag(E) %*% t(U)
90 } else {
91     HC <- M
92 }
93 return(HC)
94 }
95
96 #Função que estima o vetor gradiente
97 #(usa razões incrementais como aproximação às derivadas parciais)
98 gradiente <- function(P,L,D,l) { #P=vetor de valores iniciais para os parâmetros; L=incremento;
99                                     #D=matriz dos dados; l=vetor das idades
100     N <- dim(D)[1]    #Definir dimensões dos dados
101     C <- dim(D)[2]
102
103     R.i <- get.R(P,age)    #Obter o vetor R inicial
104     SIGMA.i <- get.SIGMA(P,l)    #Obter a matriz Sigma inicial
105     InvSIGMA.i <- solve(SIGMA.i)    #Obter a inversa da matriz SIGMA.i
106
107     #Estimar o simétrico da log-verossimilhança a partir do vetor dos valores iniciais
108     SOMA.i <- 0
109     for (i in 1:N){
110         SOMA.i <- SOMA.i + sum((t(D[i,]-R.i))%*%InvSIGMA.i%*%(D[i,]-R.i))
111     }
112     LogV.i <- (C/2)*N*log(2*pi) + (1/2)*N*log(det(SIGMA.i)) + (1/2)*SOMA.i
113
114     inc <- L    #Definir o incremento
115     IP <- length(P)
116     GRAD <- vector(length=IP)
117     incP <- vector(length=IP)
118
119     LogV <- vector(length=IP)
120     for (i in 1:IP){
121         incP[i] <- P[i] + inc*P[i] #Implementar o incremento
122         newP <- replace(P, i, incP[i]) #Atualizar vetor dos parâmetros
123
124         R <- get.R(newP,l)
125         SIGMA <- get.SIGMA(newP,l)
126         InvSIGMA <- solve(SIGMA)

```

```

127     #Atualizar LogV (efeito do incremento)
128     SOMA <- 0
129     for (j in 1:N){
130         SOMA <- SOMA + sum((t(D[j,]-R))%*%InvSIGMA%*%(D[j,]-R))
131     }
132     LogV[i] <- (C/2)*N*log(2*pi) + (1/2)*N*log(det(SIGMA)) + (1/2)*SOMA
133     #Estimar razões incrementais como aproximação às derivadas parciais
134     GRAD[i] <- (LogV[i]-LogV.i)/(incP[i]-P[i])
135 }
136 return(GRAD)
137 }
138
139 #Função que define método iterativo, com gradiente e de passo simples descendente
140 #eps=critério de convergência: 10e-10; zeta = tamanho passo da descida: 10e-08
141 grad.descend <- function(P,L,D,I,itmax,zeta,eps){
142     N <- dim(D)[1]
143     C <- dim(D)[2]
144     #Estimar log-Veros para vetor de valores inicial
145     R.i <- get.R(P,I)
146     SIGMA.i <- get.SIGMA(P,I)
147     InvSIGMA.i <- solve(SIGMA.i)
148     SOMA.i <- 0
149     for (i in 1:N){
150         SOMA.i <- SOMA.i + sum(t(D[i,]-R.i)%*%InvSIGMA.i%*%(D[i,]-R.i))
151     }
152     LogV.i <- (C/2)*N*log(2*pi) + (1/2)*N*log(det(SIGMA.i)) + (1/2)*SOMA.i
153     #Criar vetor para guardar valor da log-Veros entre iteradas
154     LogV <- numeric()
155     #Atribuir o 1º valor das estimativas da log-Veros ao vetor inicial P
156     LogV[1] <- LogV.i
157     #Iniciar o processo iterativo
158     for (i in 2:itmax){
159         #Definir processo iterativo do método que atualiza vetor dos parâmetros
160         P <- P - zeta * gradiente(P,L,D,I)
161         #Atualizar log-Veros para novo vetor de parâmetros
162         R <- get.R(P,I)
163         SIGMA <- get.SIGMA(P,I)
164         InvSIGMA <- solve(SIGMA)
165         SOMA <- 0
166         for (j in 1:N){
167             SOMA <- SOMA + sum(t(D[j,]-R)%*%InvSIGMA%*%(D[j,]-R))
168         }

```

```

169     LogV[i] <- (C/2)*N*log(2*pi) + (1/2)*N*log(det(SIGMA)) + (1/2)*SOMA
170     #Definir critério de paragem
171     if (abs((LogV[i] - LogV[i-1])/LogV[i-1]) < eps) { break }
172     #Criar vetor para guardar valores estimados
173     X <- c(i,P,LogV[i])
174     print(X)
175 }
176 #devolve vetor com nº iterada, parâmetros e valor da log-verosimilhança
177 return(X)
178 }

```



```

1 ##OTIMIZAÇÃO##
2
3 #Ler a base de dados para grupos idades quinquenais
4 dados.5 <- read.table("D:\TBM5x1.csv",sep=";")
5 attach(dados.5)
6
7 #Organizar os dados
8 TP <- dados[,1] #Tempo: 1940–2009
9 TaxasP <- dados.5[,2:41] #TBM por idade: colunas 2–21 sexo feminino; 22–41 masculino
10
11 #Selecionar os dados e o horizonte temporal para o ajustamento/previsões
12 #(todas as atribuições são feitas como no apêndice A)
13 IP <- 70
14 CP <- 40
15 nPREV <- 10
16 T <- c(1940:1999)
17 Taxas <- TaxasP[1:60,]
18 Tprev <- c(61:70)
19 taxas.prev <- TaxasP[61:70,]
20
21 #Obter os logaritmos dos retornos ( $Z = \ln(X_t/X_{t-1})$ )
22 Zp <- matrix(nrow=np,ncol=Cp)
23 for (p in 1:np){
24     for (q in 1:Cp){
25         Zp[p,q] <- log(Taxas[(p+1),q]/Taxas[p,q])
26     }
27 }
28 #Selecionar os dados para o ajustamento do modelo
29 Z <- Zp[1:(np-nPREV),]
30 l <- dim(Z)[1]
31 C <- dim(Z)[2]
32 n <- l
33
34 #Criar matrizes com logaritmos dos retornos por sexo
35 Z.F <- Z[,1:(C/2)]#Feminino
36 Z.M <- Z[,((C/2)+1):C]#Masculino
37
38 #-----
39
40 #Estimar matrizes de correlação por sexo e entre sexos
41 cor.logr.Fi <- cor(Z.F)
42 cor.logr.Mi <- cor(Z.M)

```

```

43 cor.logr.FMi <- cor(Z.F,Z.M)
44
45 #Transformar correlações p/a Normal (transformação de Fisher)
46 cor.logr.FN <- matrix(nrow=(C/2),ncol=(C/2))
47 for (p in 1:(C/2)){
48     for (q in 1:(C/2)){
49         cor.logr.FN[p,q] <- (1/2)*log((1+cor.logr.Fi[p,q])/(1-cor.logr.Fi[p,q]))
50     }
51 }
52 cor.logr.MN <- matrix(nrow=(C/2),ncol=(C/2))
53 for (p in 1:(C/2)){
54     for (q in 1:(C/2)){
55         cor.logr.MN[p,q] <- (1/2)*log((1+cor.logr.Mi[p,q])/(1-cor.logr.Mi[p,q]))
56     }
57 }
58 cor.logr.FMN <- matrix(nrow=(C/2),ncol=(C/2))
59 for (p in 1:(C/2)){
60     for (q in 1:(C/2)){
61         cor.logr.FMN[p,q] <- (1/2)*log((1+cor.logr.FMi[p,q])/(1-cor.logr.FMi[p,q]))
62     }
63 }
64 #Atualizar correlações com transformação p/a Normal
65 cor.logr.F <- cor.logr.FN
66 cor.logr.M <- cor.logr.MN
67 cor.logr.FM <- cor.logr.FMN
68
69 #Marcar valores repetidos e da diagonal (unitária) nas matrizes de correlações de F e M (como são
70 #simétricas, vamos modelar apenas 1 matriz triangular inferior e os restantes elementos são transpostos)
71 #Sexo feminino
72 cor.logr.F2 <- matrix(nrow = dim(cor.logr.F)[1], ncol = dim(cor.logr.F)[2])
73 for (p in 1:(dim(cor.logr.F)[1])){
74     for (q in 1:(dim(cor.logr.F)[2])){
75         if (p <= q) {
76             cor.logr.F2[p,q] <- NA
77         } else {cor.logr.F2[p,q] <- cor.logr.F[p,q]}
78     }
79 }
80 #Sexo masculino
81 cor.logr.M2 <- matrix(nrow = dim(cor.logr.M)[1], ncol = dim(cor.logr.M)[2])
82 for (p in 1:(dim(cor.logr.M)[1])){
83     for (q in 1:(dim(cor.logr.M)[2])){
84         if (p <= q) {

```

```

85         cor.logr.M2[p,q] <- NA
86     } else { cor.logr.M2[p,q] <- cor.logr.M[p,q] }
87 }
88 }
89
90 #-----
91
92 #Preparar as variáveis para os modelos de RNL (valores arranque correlação)
93 #Criar matriz com C/2 colunas e com o mesmo valor por linha
94 mat.idades.col <- matrix(nrow=(C/2),ncol=(C/2))
95 for (p in 1:(C/2)){
96     for (q in 1:(C/2)){
97         mat.idades.col[p,q] <- q
98     }
99 }
100 #Criar matriz com C/2 linhas e com o mesmo valor por coluna
101 mat.idades.lin <- matrix(nrow=(C/2),ncol=(C/2))
102 for (p in 1:(C/2)){
103     for (q in 1:(C/2)){
104         mat.idades.lin[p,q] <- p
105     }
106 }
107
108 v.idades <- (C/2)*(C/2)
109
110 #Criar vetor com funções da idade e correlações, por sexo
111 Dist.corr <- matrix(nrow=v.idades,ncol=5)
112 for (p in 1:v.idades){
113     Dist.corr[1] <- c(mat.idades.col) #criar vetor a partir das colunas da matriz
114 }
115 for (p in 1:v.idades){
116     Dist.corr[2] <- c(mat.idades.lin)
117 }
118 for (p in 1:v.idades){
119     Dist.corr[3] <- abs(Dist.corr[1]-Dist.corr[2])
120 }
121 for (p in 1:v.idades){
122     Dist.corr[4] <- c(cor.logr.F2) #ler a matriz de correlações, em coluna, e transformá-la num vetor
123 }
124 for (p in 1:v.idades){
125     Dist.corr[5] <- c(cor.logr.M2)
126 }

```

```

127 colnames(Dist.corr) <- c("idade.i","idade.j","abs(i-j)","ro.F","ro.M")
128
129 #Eliminar valores repetidos de correlações e tb \rho=1
130 Dist.corr <- Dist.corr[apply(Dist.corr, 1, function(x) !any(is.na(x))),]
131
132 #Criar variáveis explicativas das correlações, por sexo
133 ld1 <- Dist.corr[,1] #corresponde à ld.min (mínimo entre duas idades)
134 ld2 <- Dist.corr[,2] #corresponde à ld.max (máximo entre duas idades)
135
136 abs.ld <- Dist.corr[,3] #distância entre duas idades
137
138 ld.min <- vector(length=dim(Dist.corr)[1])
139 for (i in 1:(dim(Dist.corr)[1])){
140     ld.min[i] <- min(Dist.corr[i,1],Dist.corr[i,2])
141 }
142
143 ld.max <- vector(length=dim(Dist.corr)[1])
144 for (i in 1:(dim(Dist.corr)[1])){
145     ld.max[i] <- max(Dist.corr[i,1],Dist.corr[i,2])
146 }
147
148 ld.med <- vector(length=dim(Dist.corr)[1]) #média entre duas idades
149 for (i in 1:(dim(Dist.corr)[1])){
150     ld.med[i] <- (Dist.corr[i,1] + Dist.corr[i,2])/2
151 }
152
153 ro.F <- Dist.corr[,4]
154 ro.M <- Dist.corr[,5]
155
156 #-----
157
158 #Criar vetor com funções da idade e correlações, entre sexos
159 #(matriz .FM não é simétrica, logo não tem valores repetidos, e diagonal não é 1)
160 Dist.corrFM <- matrix(nrow=v.idades,ncol=4)
161 for (p in 1:v.idades){
162     Dist.corrFM[,1] <- c(mat.idades.col)
163 }
164 for (p in 1:v.idades){
165     Dist.corrFM[,2] <- c(mat.idades.lin)
166 }
167 for (p in 1:v.idades){
168     Dist.corrFM[,3] <- abs(Dist.corrFM[,1]-Dist.corrFM[,2])

```

```

169 }
170 for (p in 1:v.idades){
171     Dist.corrFM[4] <- c(cor.logr.FM) #ler a matrix de correlações, em coluna, e transformá-la num vetor
172 }
173 colnames(Dist.corrFM) <- c("idade.i","idade.j","abs(i-j)","ro.FM")
174
175 #Criar variáveis explicativas das correlações, entre sexos
176 ld1FM <- Dist.corr[,1] #!= da ld.minFM
177 ld2FM <- Dist.corr[,2]
178
179 abs.ldFM <- Dist.corr[,3]
180
181 ld.minFM <- vector(length=dim(Dist.corr)[1])
182 for (i in 1:(dim(Dist.corr)[1])){
183     ld.minFM[i] <- min(Dist.corr[i,1],Dist.corr[i,2])
184 }
185
186 ld.maxFM <- vector(length=dim(Dist.corr)[1])
187 for (i in 1:(dim(Dist.corr)[1])){
188     ld.maxFM[i] <- max(Dist.corr[i,1],Dist.corr[i,2])
189 }
190
191 ld.medFM <- vector(length=dim(Dist.corr)[1])
192 for (i in 1:(dim(Dist.corr)[1])){
193     ld.medFM[i] <- (Dist.corr[i,1] + Dist.corr[i,2])/2
194 }
195
196 ro.FM <- Dist.corr[,4]
197
198 #-----
199
200 library(stats)
201 #Aplicar modelo de RNL às correlações (exemplo para versão 12 do MBGM)
202 #Sexo feminino
203 corr.FITf12 <- nls(ro.F = b*ld1 + c*ld.med, start = list(b = 1, c = 1))
204 summary(corr.FITf12)
205 coef(corr.FITf12)
206 AIC(corr.FITf12)
207 FITf12 <- fitted.values(corr.FITf12)
208 #Estimar o EQM associado ao ajustamento
209 erro.ajust.f12 <- vector(length=length(ro.F))
210 for (i in 1:length(ro.F)) {

```

```

211     erro.ajust.f12[i] <- ((FITf12[i]-ro.F[i])^2)
212 }
213 erro.ajust.f12 <- sum(erro.ajust.fm2)/length(ro.F)
214
215 #Sexo masculino
216 corr.FITm12 <- nls(ro.M = b*ld1 + c*ld.med, start = list(b = 1, c = 1))
217 summary(corr.FITm12)
218 coef(corr.FITm12)
219 AIC(corr.FITm12)
220 FITm12 <- fitted.values(corr.FITm12)
221 #Estimar o EQM associado ao ajustamento
222 erro.ajust.m12 <- vector(length=length(ro.M))
223 for (i in 1:length(ro.M)) {
224     erro.ajust.m12[i] <- ((FITm12[i]-ro.M[i])^2)
225 }
226 erro.ajust.m12 <- sum(erro.ajust.m12)/length(ro.M)
227
228 #Entre sexos
229 corr.FITfm12 <- nls(ro.FM = b*ld1FM + c*ld.medFM, start = list(b = 0.5, c = 1))
230 summary(corr.FITfm12)
231 coef(corr.FITfm12)
232 AIC(corr.FITfm12)
233 FITfm12 <- fitted.values(corr.FITfm12)
234 #Estimar o EQM associado ao ajustamento
235 erro.ajust.fm12 <- vector(length=length(ro.FM))
236 for (i in 1:length(ro.FM)) {
237     erro.ajust.fm12[i] <- ((FITfm12[i]-ro.FM[i])^2)
238 }
239 erro.ajust.fm12 <- sum(erro.ajust.fm12)/length(ro.FM)
240
241 #-----
242
243 #Obter valores iniciais (ou valores de arranque) para método iterativo de otimização por MV
244 ldade <- c(1:(C/2))
245
246 #Importar valores dos parâmetros iniciais do MGB unidimensional (foram ajustados polinómios
247 #de grau 3 aos R e log(sigma^2), em função da idade)
248
249 #Para R, por sexo
250 rF.MBG <- c(-0.0674,0.0096,-0.0007,2e-05) #vetor, por ordem crescente do grau dos coeficientes
251 rM.MBG <- c(-0.0718,0.016,-0.0014,4e-05)
252 #Para 2*ln(sigma), por sexo

```

```

253 sF.MBG <- c(-3.6993,-0.1059,-0.0163,0.001)
254 sM.MBG <- c(-4.4258,0.0709,-0.0335,0.0015)
255
256 #Alternativa para vetor arranque: estimar os R e os sigma a partir dos dados, por sexo
257 RR <- vector(length=C)
258 for (i in 1:C){
259     RR[i] <- mean(Z[,i])
260 }
261 r.iF <- RR[1:(C/2)]
262 r.iM <- RR[((C/2)+1):C]
263
264 SS <- vector(length=C)
265 for (i in 1:C){
266     SS[i] <- sd(Z[,i])
267 }
268 s.iF <- SS[1:(C/2)]
269 s.iM <- SS[((C/2)+1):C]
270
271 #Estimar valores iniciais associados aos Rs (obtidos dos dados) em função da idade, por sexo
272 nls.r.F <- nls(r.iF = r0F + r1F*Idade + r2F*Idade^2 + r3F*Idade^3,
273             start = list(r0F = rF.MBG[1], r1F = rF.MBG[2], r2F = rF.MBG[3], r3F = rF.MBG[4]))
274 rF <- coef(nls.r.F)
275 summary(nls.r.F) #resumo da estimação do modelo de RNL com testes significância
276
277 nls.r.M <- nls(r.iM = r0M + r1M*Idade + r2M*Idade^2 + r3M*Idade^3,
278             start = list(r0M = rM.MBG[1], r1M = rM.MBG[2], r2M = rM.MBG[3], r3M = rM.MBG[4]))
279 rM <- coef(nls.r.M)
280 summary(nls.r.M)
281
282 #Estimar valores iniciais associados aos sigmas (obtidos dos dados) em função da idade, por sexo
283 #(usamos: LogS2iF = 2*log(s.iF) e LogS2iM = 2*log(s.iM))
284 nls.s.F <- nls(LogS2iF = s0F + s1F*Idade + s2F*Idade^2 + s3F*Idade^3,
285             start = list(s0F = sF.MBG[1], s1F = sF.MBG[2], s2F = sF.MBG[3], s3F = sF.MBG[4]))
286 sF <- coef(nls.s.F)
287
288 nls.s.M <- nls(LogS2iM = s0M + s1M*Idade + s2M*Idade^2 + s3M*Idade^3,
289             start = list(s0M = sM.MBG[1], s1M = sM.MBG[2], s2M = sM.MBG[3], s3M = sM.MBG[4]))
290 sM <- coef(nls.s.M)
291
292 #Selecionar expressão para correlações (a partir deste ponto, ilustramos os passos que levam
293 #ao método iterativo e à estimação das previsões com a versão 12 do MBGM)
294 roF.i12 <- coef(corr.FITf12)

```

```

295 roM.i12 <- coef(corr.FITm12)
296 roFM.i12 <- coef(corr.FITfm12)
297
298 #Criar vetor de arranque do método iterativo de otimização por MV
299 #Via valores obtidos no MBG unidimensional
300 P12.MBG <- c(rF.MBG,rM.MBG,sF.MBG,sM.MBG,rF.i12,roM.i12,roFM.i12)
301 #Através dos dados
302 P12 <- c(c(rF,rM),c(sF,sM),roF.i12,roM.i12,roFM.i12)
303
304 #-----
305
306 #Obter função de log-Verosimilhança
307 LogV12 <- function(P){ #P=vetor de valores iniciais do modelo (na versão 12 do MBGM=22)
308     N <- dim(D)[1]
309     C <- dim(D)[2]
310     I <- c(1:(C/2))
311     #Estimar vetor R
312     R <- vector(length=C)
313     for (i in 1:C){
314         if (i <= (C/2)) {
315             R[i] <- P[1] + P[2]*I[i] + P[3]*I[i]^2 + P[4]*I[i]^3
316         } else {
317             R[i] <- P[5] + P[6]*I[(i-(C/2))] + P[7]*I[(i-(C/2))]^2 + P[8]*I[(i-(C/2))]^3
318         }
319     }
320     #Estimar matriz de correlações, por blocos (inverter ro.Z: r=((exp(2*(ro.Z))-1)/(exp(2*(ro.Z))+1)))
321     CORR.FF <- matrix(nrow=(C/2),ncol=(C/2))
322     for (p in 1:(C/2)){
323         for (q in 1:(C/2)){
324             if (p == q){
325                 CORR.FF[p,q] <- 1
326             } else {
327                 CORR.FF[p,q] <- (exp(2*(P[17]*min(p,q)+P[18]*(p+q)/2))-1)/
328                     (exp(2*(P[17]*min(p,q)+P[18]*(p+q)/2))+1)
329             }
330         }
331     }
332     CORR.FM <- matrix(nrow=(C/2),ncol=(C/2))
333     for (p in 1:(C/2)){
334         for (q in 1:(C/2)){
335             CORR.FM[p,q] <- (exp(2*(P[21]*q+P[22]*(p+q)/2))-1)/
336                 (exp(2*(P[21]*q+P[22]*(p+q)/2))+1)

```



```

337     }
338 }
339 CORR.MF <- t(CORR.FM)
340 CORR.MM <- matrix(nrow=(C/2),ncol=(C/2))
341 for (p in 1:(C/2)){
342     for (q in 1:(C/2)){
343         if (p == q){
344             CORR.MM[p,q] <- 1
345         } else {
346             CORR.MM[p,q] <- (exp(2*(P[19]*min(p,q)+P[20]*(p+q)/2))-1)/
347                 (exp(2*(P[19]*min(p,q)+P[20]*(p+q)/2))+1)
348         }
349     }
350 }
351 #Obter matriz de correlações
352 CORR <- rbind(cbind(CORR.FF,CORR.FM),cbind(CORR.MF,CORR.MM))
353 #Estimar matriz lambda
354 S <- vector(length=C) #obter desvios padrão (inverter transformação Z.sigma: sigma=exp(Z.sigma/2))
355 for (i in 1:C){
356     if (i <= (C/2)) {
357         S[i] <- exp((P[9] + P[10]*I[i] + P[11]*I[i]^2 + P[12]*I[i]^3)/2)
358     } else {
359         S[i] <- exp((P[13] + P[14]*I[(i-(C/2))] + P[15]*I[(i-(C/2))]^2 +
360             P[16]*I[(i-(C/2))]^3)/2)
361     }
362 }
363 LAMBDA <- diag(S) #SIGMA=sqrt(LAMBDA)%*%CORR%*%sqrt(LAMBDA)
364 #Estimar simétrico da função de Log-Verosimilhança
365 logC <- (C/2)*N*log(2*pi) + (1/2)*N*log(det(LAMBDA)*det(CORR)*det(LAMBDA))
366 SOMA.t <- 0
367 for (j in 1:N){
368     SOMA.t <- SOMA.t+sum(t(Z[j,]-R)%*%(diag(1/diag(LAMBDA))%*%
369         solve(CORR,(diag(1/diag(LAMBDA))))%*%(Z[j,]-R))
370 }
371 MlogV <- logC + (1/2)*SOMA.t
372 return(MlogV)
373 }
374
375 #Aplicar método iterativo (optim necessita de: library(stats), já declarada acima)
376 #Por defeito: método de Nelder-Mead; dimensão do passo da descida de 1e-06; tolerância relativa de 1e-08
377 min.LogV12 <- optim(P12, LogV12, control = list(trace = TRUE, maxit = 10000), hessian = TRUE)
378

```

```

379 #-----
380
381 #Obter parâmetros do modelo (versão 12 do MBGM)
382 par12 <- min.LogV12$par
383
384 #Obter aproximação à matriz de covariâncias
385 VAR12 <- solve(min.LogV12$hessian)
386
387 #Estimar margens de erro dos ICs dos parâmetros
388 p1.12tol <- 1.96*sqrt(VAR12[1,1])
389 p2.12tol <- 1.96*sqrt(VAR12[2,2])
390 ...
391 p21.12tol <- 1.96*sqrt(VAR12[21,21])
392 p22.12tol <- 1.96*sqrt(VAR12[22,22])
393
394 #Obter o vetor R
395 R.par12 <- vector(length=C)
396 for (i in 1:C){
397     if (i <= (C/2)) {
398         R.par12[i] <- par12[1] + par12[2]*Idade[i] + par12[3]*Idade[i]^2 + par12[4]*Idade[i]^3
399     } else {
400         R.par12[i] <- par12[5] + par12[6]*Idade[(i-(C/2))] + par12[7]*Idade[(i-(C/2))]^2 +
401             par12[8]*Idade[(i-(C/2))]^3
402     }
403 }
404
405 #Obter o vetor S (diagonal da matriz LAMBDA)
406 S.par12 <- vector(length=C)
407 for (i in 1:C){
408     if (i <= (C/2)) {
409         S.par12[i] <- par12[9] + par12[10]*Idade[i] + par12[11]*Idade[i]^2 + par12[12]*Idade[i]^3
410     } else {
411         S.par12[i] <- par12[13] + par12[14]*Idade[(i-(C/2))] + par12[15]*Idade[(i-(C/2))]^2 +
412             par12[16]*Idade[(i-(C/2))]^3
413     }
414 }
415
416 #Obter ajustamentos a LP na escala original dos dados
417 ajust12.TBM <- matrix(nrow=(l+1),ncol=cp)
418 for (p in 1:(l+1)){
419     for (q in 1:cp){
420         ajust12.TBM[p,q] <- Taxas[1,q]*exp(p*R.par12[q])

```

```

421     }
422 }
423
424 #Estimar o EQM associado aos ajustamentos a LP
425 EQM.ajust12 <- vector(length=(l+1))
426 erro.ajust12 <- matrix(nrow=(l+1),ncol=c) #+ 1 linha para calcular erro
427 for (p in 1:l) {
428     for (q in 1:c){
429         erro.ajust12[p,q] <- ((ajust12.TBM[p,q]-Taxas[p,q])^2)
430     }
431 }
432 for (p in 1:l) {
433     for (q in 1:c){
434         erro.ajust12[(l+1),q] <- sum(erro.ajust12[(1:p),q])/l
435     }
436 }
437 EQM.ajust12 <- erro.ajust12[(l+1),]
438
439 #Inverter previsões a LP: 10 anos dos logaritmos dos retornos (Z = ln(Xt/Xt-1)) para escala original dos dados
440 prev12.TBM <- matrix(nrow=nPREV,ncol=cp)
441 for (p in 1:nPREV){
442     for (q in 1:cp){
443         prev12.TBM[p,q] <- Taxas[60,q]*exp(p*R.par12[q])
444     }
445 }
446
447 #Estimar o EQM associado às previsões a LP: 10 anos
448 EQM.prev12 <- vector(length=(nPREV+1))
449 erro.prev12 <- matrix(nrow=(nPREV+2),ncol=C) #+ 1 linha para calcular erro
450 for (p in 1:(nPREV+1)) {
451     for (q in 1:C){
452         erro.prev12[p,q] <- ((prev12.TBM[p,q]-taxas.prev[p+1,q])^2)
453     }
454 }
455 for (p in 1:(nPREV+1)) {
456     for (q in 1:C){
457         erro.prev12[(nPREV+2),q] <- sum(erro.prev12[(1:p),q])/(nPREV)
458     }
459 }
460 EQM.prev12 <- erro.prev12[(nPREV+2),]

```

```

1 ## IC PARA PREVISÕES a LP## (ilustramos com versão 12 do MBGM)
2
3 #Repetir o procedimento a seguir para cada ano das previsões, neste caso, 10 anos (a forma mais
4 #eficiente seria criar um ciclo com o nº de iteradas = nº previsões, mas teríamos que guardar as matrizes
5 #e os vetores em listas, cuja indexação não é muito intuitiva, pelo que decidimos aqui ilustrar desta forma)
6 #Todos os objetos designados a1. ... referem-se ao 1º ano de previsões (a2. ..., para o 2º ...)
7 #Nos objetos a1.logTBM e a1.logPrev, onde está t, usar o período de tempo a que se refere a previsão
8 #(ano 1: t=1; ano 2: t=2...)
9
10 #-----
11 #Previsões: ANO 2000 (t=1)
12 #-----
13 #Estimar uma matriz de nºs aleatórios com distribuição Normal(0,1), por linha (as linhas representam as idades)
14 H1 <- matrix(rnorm(40*1000, mean = 0, sd = 1), 40, 1000, byrow = TRUE)
15
16 #Criar matriz com 1000 repetições do vetor R
17 matriz.R.par12 <- matrix(nrow=c, ncol=1000)
18 for (p in 1:c){
19     for (q in 1:1000){
20         matriz.R.par12[p,q] <- R.par12[p]
21     }
22 }
23
24 #Estimar a matriz B=(SIGMA.12)^(1/2)
25 SIGMA.12.eig <- eigen(SIGMA.12)
26 B <- SIGMA.12.eig$vectors %*% diag(sqrt(SIGMA.12.eig$values)) %*% solve(SIGMA.12.eig$vectors)
27
28 #Obter estrutura de correlação
29 M1 <- B %*% H1
30
31 #Estimar valores dos logaritmos dos retornos das TBM para cada perturbação do ambiente
32 Z1 <- matriz.R.par12 + M1
33
34 #Inverter a transformação dos logaritmos dos retornos das TBM para os log(TBM), que têm distribuição normal
35 Y1.logTBM <- matrix(nrow = C, ncol = 1000)
36 for (p in 1:C){
37     for (q in 1:1000){
38         Y1.logTBM[p,q] <- log(Taxas[60,p]) + Z1[p,q]*t
39     }
40 }
41
42 #Estimar as médias dos log(TBM) para cada idade

```

```

43 Y1.medias <- vector(length= C)
44 for (i in 1:C){
45     Y1.medias[i] <- mean(Y1.logTBM[i,])
46 }
47
48 #Ordenar, para cada idade, as réplicas dos log(TBM)
49 Y1.ord <- t(apply(Y1.logTBM,1,sort))
50
51 #Obter quartis 25% e 75%, para cada idade
52 Y1.Q25 <- vector(length=C)
53 for (i in 1:C){
54     Y1.Q25[i] <- quantile(Y1.ord[i,],name=FALSE)[2]
55 }
56 #
57 Y1.Q75 <- vector(length=C)
58 for (i in 1:C){
59     Y1.Q75[i] <- quantile(Y1.ord[i,],name=FALSE)[4]
60 }
61
62 #Representar as previsões com IC, para uma certa idade, em escala logarítmica
63 #ex: idade 30 que corresponde ao grupo etário 45–49 do sexo masculino
64 par(mfrow=c(1,1))
65 plot(c(Y1.Q75[30],Y2.Q75[30],Y3.Q75[30],Y4.Q75[30],Y5.Q75[30],Y6.Q75[30],Y7.Q75[30],Y8.Q75[30],
66     Y9.Q75[30],Y10.Q75[30]),type="l", main="Prev. c/ IC Grupo etário 45-49 do sexo masculino",
67     xlab="Ano", col="brown", xaxt="n", ylab="log(TBM) ", ylim=c(-7,-4))
68 axis(1, at=1:10, labels=c(2000:2009))
69 lines(log(taxas.prev[2:11,30]),col="black")
70 lines(c(Y1.logPrev[30],Y2.logPrev[30],Y3.logPrev[30],Y4.logPrev[30],Y5.logPrev[30],Y6.logPrev[30],
71     Y7.logPrev[30],Y8.logPrev[30],Y9.logPrev[30],Y10.logPrev[30]), col="blue")
72 lines(c(Y1.medias[30],Y2.medias[30],Y3.medias[30],Y4.medias[30],Y5.medias[30],Y6.medias[30],Y7.medias[30],
73     Y8.medias[30],Y9.medias[30],Y10.medias[30]), col="red")
74 lines(c(Y1.Q25[30],Y2.Q25[30],Y3.Q25[30],Y4.Q25[30],Y5.Q25[30],Y6.Q25[30],Y7.Q25[30],Y8.Q25[30],
75     Y9.Q25[30],Y10.Q25[30]), col="brown")
76 legend("bottom", horiz=TRUE, legend=c("Observadas","Previsões","Média simuladas","Limites MC 95%"),
77     lty=c(1,1,1,1), col=c("black","blue","red","brown"),bty="n")

```



**Modelos ARIMA ajustados aos dados
da mortalidade da população
portuguesa**

Tabela F.1: Modelos ARIMA ajustados às series temporais das TBMs da população portuguesa para o sexo feminino e por grupo etário

Feminino		
0-1	$(1+0.522B)(1-B)(Y_t+0.056)=\varepsilon_t$	$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2=0.687 E^{-2}), Y_t = \ln(X_t)$
1-4	$(1+0.588B)(1-B)(Y_t+0.069)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.151 E^{-1}), Y_t = \ln(X_t)$
5-9	$(1+0.472B)(1-B)(Y_t+0.046)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.187 E^{-1}), Y_t = \ln(X_t)$
10-14	$(1+0.433B)(1-B)(Y_t+0.042)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.125 E^{-1}), Y_t = \ln(X_t)$
15-19	$(1+0.944B)(1-B)(Y_t+0.038)=(1+0.861B)\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.108 E^{-1}), Y_t = \ln(X_t)$
20-24	$(1-B)(Y_t+0.039)=(1+0.336B^5)\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.956 E^{-2}), Y_t = \ln(X_t)$
25-29	$(1-B)(Y_t+0.041)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.840 E^{-2}), Y_t = \ln(X_t)$
30-34	$(1+0.315B)(1-B)(Y_t+0.033)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.741 E^{-2}), Y_t = \ln(X_t)$
35-39	$(1+0.413B)(1-B)(Y_t+0.028)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.592 E^{-2}), Y_t = \ln(X_t)$
40-44	$(1+0.440B)(1-B)(Y_t+0.024)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.364 E^{-2}), Y_t = \ln(X_t)$
45-49	$(1-B)(Y_t+0.020)=(1-0.463B)\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.395 E^{-2}), Y_t = \ln(X_t)$
50-54	$(1-B)(Y_t+0.019)=(1-0.654B)\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.384 E^{-2}), Y_t = \ln(X_t)$
55-59	$(1+0.455B)(1-B)(Y_t+0.018)=(1-0.361B^3)\varepsilon_t$	$\varepsilon_t \sim NWN(0, \sigma_\varepsilon^2=0.222 E^{-2}), Y_t = \ln(X_t)$
60-64	$(1-B)(Y_t+0.019)=(1-0.528B)\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.212 E^{-2}), Y_t = \ln(X_t)$
65-69	$(1-B)(Y_t+0.018)=(1-0.548B)\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.256 E^{-2}), Y_t = \ln(X_t)$
70-74	$(1-B)(Y_t+0.018)=(1-0.475B)\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.317 E^{-2}), Y_t = \ln(X_t)$
75-79	$(1-B)(Y_t+0.016)=(1-0.425B)\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.306 E^{-2}), Y_t = \ln(X_t)$
80-84	$(-0.733B+0.218B^5)(Y_t)/(1-B)(Y_t+0.01)=\varepsilon_t$	$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2=0.262 E^{-2}), Y_t = \ln(X_t)$
85-89	$(1+0.540B+0.523B^2-0.296B^5-0.358B^7)(1-B)(Y_t+0.007)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.218 E^{-2}), Y_t = \ln(X_t)$
90-94	$(1+0.511B-0.245B^5)(1-B)(Y_t)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.387 E^{-2}), Y_t = \ln(X_t)$
95-99	$(1+0.506B)(1-B)(Y_t)=\varepsilon_t$	$\varepsilon_t \sim NWN(0, \sigma_\varepsilon^2=0.855 E^{-2}), Y_t = \ln(X_t)$

Tabela F.2: Modelos ARIMA ajustados às series temporais das TBMs da população portuguesa para o sexo masculino e por grupo etário

Masculino		
0-1	$(1+0.435B)(1-B)(Y_t+0.054)=\varepsilon_t$	$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2=0.612 E^{-2}), Y_t = \ln(X_t)$
1-4	$(1+0.485B)(1-B)(Y_t+0.065)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.129 E^{-1}), Y_t = \ln(X_t)$
5-9	$(1+0.468B)(1-B)(Y_t+0.044)=\varepsilon_t$	$\varepsilon_t \sim WMN(0, \sigma_\varepsilon^2=0.937 E^{-2}), Y_t = \ln(X_t)$
10-14	$(1+0.399B)(1-B)(Y_t+0.036)=\varepsilon_t$	$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2=0.853 E^{-2}), Y_t = \ln(X_t)$
15-19	$(1-B)(Y_t+0.022)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.688 E^{-2}), Y_t = \ln(X_t)$
20-24	$(1-0.813B)(1-B)(Y_t)=(1-0.602B)\varepsilon_t$	$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2=0.494 E^{-7}), Y_t = X_t$
25-29	$(1+0.277B)(1-B)(Y_t)=\varepsilon_t$	$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2=0.939 E^{-2}), Y_t = \ln(X_t)$
30-34	$(1+0.998B)(1-B)(Y_t)=(1+0.984B)\varepsilon_t$	$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2=0.567 E^{-7}), Y_t = X_t$
35-39	$(1+0.999B)(1-B)(Y_t)=(1+0.993B)\varepsilon_t$	$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2=0.779 E^{-7}), Y_t = X_t$
40-44	$(1+0.253B)(1-B)(Y_t+0.017)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.522 E^{-2}), Y_t = \ln(X_t)$
45-49	$(0.358B^5 Y_t)(1-B)(Y_t+0.015)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.218 E^{-6}), Y_t = X_t$
50-54	$(0.334B^4 Y_t)(1-B)(Y_t+0.015)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.304 E^{-2}), Y_t = \ln(X_t)$
55-59	$(1+0.395B)(1-B)(Y_t+0.013)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.2869 E^{-2}), Y_t = \ln(X_t)$
60-64	$(0.370B+0.272B^2)(Y_t)(1-B)(Y_t+0.013)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.2448 E^{-2}), Y_t = \ln(X_t)$
65-69	$(1+0.530B)(1-B)(Y_t+0.011)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.258 E^{-2}), Y_t = \ln(X_t)$
70-74	$(-0.263B^7 Y_t)(1-B)(Y_t+0.014)=(1-0.485)\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.266 E^{-2}), Y_t = \ln(X_t)$
75-79	$(1+0.396B-0.396B^5)(1-B)(Y_t)=\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.280 E^{-2}), Y_t = \ln(X_t)$
80-84	$(1-B)(Y_t+0.008)=(1-0.713B)\varepsilon_t$	$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2=0.290 E^{-2}), Y_t = \ln(X_t)$
85-89	$(1-B)(Y_t+0.007)=(1-0.761B)\varepsilon_t$	$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2=0.308 E^{-2}), Y_t = \ln(X_t)$
90-94	$(1-B)(Y_t)=(1-0.628B)\varepsilon_t$	$\varepsilon_t \sim Niid(0, \sigma_\varepsilon^2=0.609 E^{-2}), Y_t = \ln(X_t)$
95-99	$(1-B)(Y_t)=(1-0.799B)\varepsilon_t$	$\varepsilon_t \sim NWN(0, \sigma_\varepsilon^2=0.1278 E^{-1}), Y_t = \ln(X_t)$

Referências bibliográficas

- [1] J. Albert e M. Rizzo. *R by Example*. Springer, Nova York, 2012.
- [2] J. Alho e B. Spencer. *Statistical Demography and Forecasting* (Springer Series in Statistics). Springer, Nova York, 2005.
- [3] L. Arnold. *Stochastic differential equations: Theory and Applications*. Krieger Publishing Company, Florida, 1992.
- [4] D. Bates e D. Watts. *Nonlinear Analysis and Its Applications* (Wiley Series in Probability and Statistics). John Wiley & Sons, Nova York, 1988.
- [5] H. Booth e L. Tickle. Mortality modelling and forecasting: A review of methods. *Working Paper 3, The Australian Demographic & Social Research Institute*, 2008.
- [6] R. Bowen e C. Wang. *Introduction to Vectors and Tensors*, 2ª Edição. Dover Publications, Nova York, 2008.
- [7] P. Box e G. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, São Francisco, 1976.
- [8] C.A. Braumann. Growth and extinction in randomly varying populations. *Computers and Mathematics with Applications*, 56(3), 631-644, 2008.
- [9] C.A. Braumann. *Introdução às Equações Diferenciais Estocásticas e Aplicações*. Edições SPE, Lisboa, 2005.
- [10] C.A. Braumann. Estimação dos parâmetros para uma ou várias trajectórias do movimento browniano geométrico. *Afirmar a estatística: um desafio para o século XXI, Actas do VI Congresso Anual da Sociedade Portuguesa de Estatística*, 149-156, 1999.
- [11] J. Braun e D. Murdoch. *A First Course in Statistical Programming with R*. Cambridge University Press, Nova York, 2007.
- [12] J. Bravo. *Tábuas de mortalidade contemporâneas e prospectivas: modelos estocásticos, aplicações actuariais e cobertura do risco de longevidade*. Tese de Doutoramento, Universidade de Évora, Évora, 2007.

- [13] J.M. Bravo e C.A. Braumann. The value of a random life: modelling survival probabilities in a stochastic environment. *Bulletin of International Statistical Institute*, LXII, 5741-5744, 2007.
- [14] N. Brites. *Modelos estocásticos de crescimento individual e desenvolvimento de software de estimação e previsão*. Dissertação de Mestrado, Universidade de Évora, Évora, 2010.
- [15] A. Bryk e S. Raudenbush. Heterogeneity of Variance in Experimental Studies: A Challenge to Conventional Interpretations. *Psychological Bulletin*, 104(3), 396-404, 1988.
- [16] P. Brockwell e R. Davis. *Time Series: Theory and Methods*. 2ª Edição, Springer-Verlag, Nova York, 1991.
- [17] P. Brockwell e R. Davis. *Introduction to Time Series and Forecasting*. 2ª Edição, Springer-Verlag, Nova York, 2002.
- [18] A. Cairns, D. Blake e K. Dowd. Modelling and management of the mortality risk: a review. *Scandinavian Actuarial Journal*, 108(2-3), 79-113, 2008.
- [19] A. Cairns, B. Blake, K. Dowd, G. Coughlan e D. Epstein. Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, 48, 355-367, 2011.
- [20] G. Casella e R. Berger. *Statistical Inference*. 2ª Edição, Duxbury, Nova Deli, 2002.
- [21] E. Coelho, M. Magalhães e J. Bravo. Mortality projections in Portugal. *Conference of European Statisticians*. UNECE-Eurostat, 2010.
- [22] A. Debón, F. Montes e F. Puig. Modelling and Forecasting mortality in Spain. *European Journal of Operational Research*, 189, 624-637, 2008.
- [23] K. Dowd, A. Cairns, B. Blake, G. Coughlan, D. Epstein e M. Khalaf-Allah. Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics*, 47, 255-265, 2010.
- [24] J. Faraway. *Linear Models with R* (Texts in Statistical Science). Taylor & Francis, Florida, 2005.
- [25] P.A. Filipe. *Equações diferenciais estocásticas na modelação do crescimento individual em ambiente aleatório*. Tese de Doutoramento, Universidade de Évora, Évora, 2011.
- [26] P.A. Filipe, C.A. Braumann e C.J. Roquete. Crescimento individual em ambiente aleatório: Várias trajectórias. *Actas do XV Congresso Anual da SPE*. 259-268, 2008.
- [27] P.A. Filipe, C.A. Braumann e C.J. Roquete. Multiphasic Individual Growth Models in Random Environments. *Methodology and Computing in Applied Probability*, 14(1), 49-56, 2012.
- [28] P.A. Filipe, C.A. Braumann, N. Brites e C.J. Roquete. *Prediction for Individual Growth in a Random Environment*. Recent Developments in Modeling and Applications in Statistics, Series: Studies in Theoretical and Applied Statistics. Subseries: Selected Papers of the Statistical Societies (P. Oliveira, M. Temido, C. Henriques e V. Maurizio (Editores)), 193-201, 2013.

- [29] J. Fox. *An R companion to applied regression*, 2ª Edição. Sanford Weisberg, Califórnia, 2011.
- [30] J. Franco. *Maximum likelihood estimation of mean reverting processes*. Real Options Practice - Onward Inc., 2003.
- [31] F. Gantmacher. *Applications of the Theory of Matrices*. Dover Publications, Nova York, 2005.
- [32] F. Gao e L. Han. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications*, 51(1), 259-277, 2012.
- [33] M. George, S. Smith, D. Swanson e J. Tayman. Population Projections. Em S. Jacob e D. Swanson, editores, *The Methods and Materials of Demography*, 2ª Edição, 561-601. Elsevier Academic Press, 2003.
- [34] G.H. Golub e C.F. Van Loan. *Matrix Computations*, 3ª Edição. Johns Hopkins University Press, Baltimore, 1996.
- [35] D. Gomes. *Sebenta de Séries Temporais*. Universidade de Évora, Évora, 2005.
- [36] S. Hang Li. *Stochastic Mortality Models with Applications in Financial Risk Management*. Tese de Doutoramento, Universidade de Waterloo, Canada, 2007.
- [37] Human Mortality Database. University of California e Max Planck Institute for Demographic Research [extraído em 4-10-2011 de <http://www.mortality.org>].
- [38] R.J. Hyndman e H. Booth. Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, 24, 323-342, 2008.
- [39] R.J. Hyndman e M.S. Ullah. Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics and Data Analysis*, 51(10), 4942-4956, 2007.
- [40] Instituto Nacional de Estatística (2005). Projecções de População Residente, Portugal e Nuts III, 2000-2050. Instituto Nacional de Estatística [extraído em 4-3-2009 de <http://www.ine.pt>].
- [41] Instituto Nacional de Estatística (2008). Estimativas da População Residente, de 2007. Instituto Nacional de Estatística [extraído em 4-3-2009 de <http://www.ine.pt>].
- [42] Instituto Nacional de Estatística (2012). Censos 2011 Resultados definitivos - Portugal: Autor. Recuperado em 8 de Maio de 2013 de <http://www.ine.pt>.
- [43] G. Janacek. *Practical Time Series* (Arnold Texts in Statistics). Arnold, Londres, 2001.
- [44] P. Jevtic, E. Luciano e E. Vigna. Mortality surface by means of continuous time cohort models. *Insurance: Mathematics and Economics*, 53(1), 122-133, 2013.
- [45] N. Keilman. *Erroneous Population Forecasts, Perspectives on Mortality Forecasting. II. Probabilistic Models*. Swedish Social Insurance Agency, 2005.

- [46] N. Keyfitz e H. Caswell. *Applied Mathematical Demography* (Statistics for biology and health), 3ª Edição. Springer, Nova York, 2005.
- [47] H. Kintner. The Life Table. Em S. Jacob e D. Swanson, editores, *The Methods and Materials of Demography*, 2ª Edição, 301-340. Elsevier Academic Press, 2003.
- [48] P. Lancaster e M. Tismenetsky. *The Theory of Matrices, Second Edition: With Applications* (Series in Computer Science and Scientific Computing). Academic Press, San Diego, 1985.
- [49] K. Lange. *Optimization* (Springer Series in Statistics). Springer, Nova York, 2013.
- [50] R. Lee. The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications. *North American Actuarial Journal*, 4(1), 80-93, 2000.
- [51] R. Lee e L. Carter. Modelling and forecasting the time series of US mortality. *Journal of the American Statistical Association*, 87, 659-671, 1992.
- [52] Life Office Mortality Committee (2007). Stochastic projection methodologies: Lee-Carter model features, example results and implications. Continuous Mortality Investigation - Artigo em preparação 25: 1-30 [extraído em 10-01-2009 de <http://www.actuaries.org.uk/data/assets/pdf/0011/38846/wp25revised.pdf>].
- [53] T. Lu e S. Shiou. Inverses of 2 x 2 Block Matrices. *Computers and Mathematics with Applications*, 43, 119-129, 2000.
- [54] H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlim, 2005.
- [55] H. Maier e G. Dandy. *Application of Multivariate Time Series Modelling to the Prediction of Salinity*. Relatório de Investigação 129, Departamento de Engenharia Civil e Ambiental, Universidade de Adelaide, 1995.
- [56] J. Maindonald e W.J. Braun. *Data Analysis and Graphics Using R - an Example-Based Approach*, 3ª Edição, Cambridge University Press, Nova York, 2010.
- [57] M. Mcgehee. Mortality. Em S. Jacob e D. Swanson, editores, *The Methods and Materials of Demography*, 2ª Edição, 265-300. Elsevier Academic Press, 2003.
- [58] M. Mendes. *Mortalidade: Tábua de Mortalidade*. Universidade de Évora, Évora, 2004.
- [59] T. Mexia e P. Corte-Real. Tabelas de mortalidade auto-correctivas: o caso português. *Boletim do Instituto de Actuários Portugueses*, 35, 1-120, 1995.
- [60] T. Mishra. Stochastic Demographic Dynamics and Economic Growth: An Application and Insights from the World Data. *Historical Social Research*, 33(4), 9-187, 2008.
- [61] M. Morais. *Causas de Morte no Século XX. Transição e Estruturas da Mortalidade em Portugal Continental*. Edições Colibri e CIDEHUS-UE, Lisboa, 2002.

- [62] D. Müller. *Processos Estocásticos e Aplicações*. Edições Almedina, Coimbra, 2007.
- [63] P. Murrell. *R Graphics (The R Series)*. 2ª Edição, Chapman & Hall/CRC, Florida, 2011.
- [64] B. Murteira, D. Müller e K. Turkman. *Análise de sucessões cronológicas*. McGraw-Hill, Lisboa, 1993.
- [65] K. Namboodiri e C.M. Suchindran. *Life table techniques and their applications (Studies in Population)*. Academic Press, Orlando, 1987.
- [66] J. Nicolau. *Modelação e Estimação de Séries Financeiras através de Equações Diferenciais Estocásticas*. Tese de Doutoramento, Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa, 2001.
- [67] J. Nocedal e S. Wright. *Numerical Optimization*. 2ª Edição, Springer, Nova York, 2006.
- [68] B. Øksendal. *Stochastic Differential Equations. An Introduction with Applications*. 6ª Edição, Springer-Verlag, Berlim, 2003.
- [69] H.S. Park. The survival probability of mortality intensity with jump-diffusion. *Journal of the Korean Statistical Society*, 37, 355-363, 2008.
- [70] D. Pestana e S. Velosa. *Introdução à Probabilidade e à Estatística*. Volume I, Fundação Calouste Gulbenkian, Lisboa, 2002.
- [71] B. Pfaff. *Analysis of Integrated and Cointegrated Time Series with R*. 2ª Edição. Springer, Nova York, 2008.
- [72] B. Pfaff. VAR, SVAR and SVEC Models: Implementation Within R Package vars. *Journal of Statistical Software*, 27(4), 2008, [extraído em 17-4-2012 de <http://www.jstatsoft.org/v27/i04/paper>].
- [73] R. Prado e M. West. *Time Series: Modeling, Computation, and Inference (Texts in Statistical Science)*. Chapman & Hall/CRC, Florida, 2010.
- [74] W. Press, S. Teukolsky, W. Vetterling e B. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, Nova York, 2007.
- [75] S. Preston, P. Heuveline e M. Guillot. *Demography: measuring and modeling population processes*. Blackwell, Oxford, 2004.
- [76] G. Reinsel. *Elements of Multivariate Time Series Analysis*. 2ª Edição, Springer-Verlag, Nova York, 1997.
- [77] C. Ritz e J. Streibig. *Nonlinear Regression with R*. Springer, Nova York, 2008.
- [78] L.R. Schaeffer. Modification of negative eigenvalues to create positive definite matrices and approximation of standard errors of correlation estimates. University of Guelph, Canadá, 2010, [extraído em 14-3-2011 de <http://www.aps.uoguelph.ca/~lrs/LRSsite/PDforce.pdf>].

- [79] R. Schumway e D. Stoffer. *Time Series Analysis and Its Applications* (Springer Texts in Statistics). 2ª Edição, Springer, Nova York, 2006.
- [80] C. Sims. Macroeconomics and reality. *Econometrica*, 48(1), 1-48, 1980.
- [81] C. Skiadas. Exact Solutions of Stochastic Differential Equations: Gompertz, Generalized Logistic and Revised Exponential. *Methodology and Computing in Applied Probability*, 12, 261-270, 2010.
- [82] R. Sokal e F. Rohlf. *Biometry*. 3ª Edição, Freeman and Company, Nova York, 1998.
- [83] N.H. Timm *Applied Multivariate Analysis* (Springer texts in statistics). Springer-Verlag, Nova York, 2002.
- [84] Y.L. Tong. *Multivariate Normal Distribution* (Springer Series in Statistics). Springer-Verlag, Nova York, 1989.
- [85] T. Wilson e P. Rees. Recent Developments in Population Projection Methodology: A Review, *Population. Space and Place*, 11, 337-360, 2005.
- [86] R. Yafee. *Time Series Analysis and Forecasting - with applications of SAS and SPSS*. Academic Press, EUA, 2000.
- [87] A.I. Yashin, k.G. Arbeev, I. Akushevich, A. Kulminski, L.Akushevich, e S.V. Ukraintseva. Stochastic model for analysis of longitudinal data on aging and mortality. *Mathematical Biosciences*, 208(2), 538-551, 2007.
- [88] A. Zeileis, C. Kleiber, W. Krämer e K. Hornik. Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis*, 44, 109-123, 2003.
- [89] A. Zuur, E. Ieno e E. Meesters. *A Beginner's Guide to R*. Springer, Nova York, 2009.



Contactos:

Universidade de Évora

Instituto de Investigação e Formação Avançada - IIFA

Palácio do Vimioso — Largo Marquês de Marialva, Apart. 94

7002-554 Évora — Portugal

Tel: (+351) 266 706 581

Fax: (+351) 266 744 677

email: iifa@uevora.pt