



Departamento de Métodos Quantitativos para Gestão e Economia

Modelos com Trajetória Latente em Painéis Rotativos com Dados Omissos

Paula Cristina Ribeiro Vicente

Tese especialmente elaborada para obtenção do grau de
Doutor em Métodos Quantitativos Aplicados

Orientadora:

Prof^ª Doutora Maria de Fátima Salgueiro, Professora Associada com Agregação
Instituto Universitário de Lisboa, ISCTE-IUL

maio 2014

Departamento de Métodos Quantitativos para Gestão e Economia

Modelos com Trajetória Latente em Painéis Rotativos com Dados Omissos

Paula Cristina Ribeiro Vicente

Tese especialmente elaborada para obtenção do grau de
Doutor em Métodos Quantitativos Aplicados

Júri:

Doutora Elizabeth de Azevedo Reis, Professora Catedrática
do ISCTE-Instituto Universitário de Lisboa

Doutora Maria Manuela Costa Neves Figueiredo, Professora Catedrática
do Instituto Superior de Agronomia da Universidade de Lisboa

Doutora Maria Eugénia Neto Ferrão da Silva, Professora Auxiliar com Agregação
da Universidade da Beira Interior

Doutora Maria da Conceição Esperança Amado, Professora Auxiliar
do Instituto Superior Técnico da Universidade de Lisboa

Doutor José Manuel Gonçalves Dias, Professor Associado
do ISCTE-Instituto Universitário de Lisboa

Doutora Maria de Fátima Ramalho Fernandes Salgueiro (orientadora), Professora
Associada com Agregação do ISCTE-Instituto Universitário de Lisboa

maio 2014

Esta tese é dedicada às minhas filhas Maria e Carolina

Agradecimentos

A elaboração de uma tese de doutoramento é um longo caminho que envolve pessoas e instituições, e sem o apoio das quais este trabalho não teria sido possível. As próximas linhas são-lhes dedicadas.

Em primeiro lugar quero expressar o meu agradecimento à orientadora, Professora Maria de Fátima Salgueiro, pelo seu apoio e disponibilidade ao longo do tempo em que esta tese foi realizada. Sem o seu empenho e amizade este trabalho não teria sido o mesmo. Também, aos Professores Elizabeth Reis e José Dias um agradecimento muito especial pelas suas preciosas contribuições ao longo dos vários painéis em que este trabalho foi sendo avaliado.

Agradeço ainda, à unidade de investigação BRU-IUL que me proporcionou a participação em congressos nacionais e internacionais.

Aos colegas da Universidade Lusófona, e aos amigos que, ao longo do tempo que demorou a elaboração desta tese, me incentivaram com uma palavra amiga, no momento certo, é também devida uma palavra de agradecimento.

Por fim, devo agradecer à minha mãe, ao meu marido e às minhas filhas, que me ajudaram neste percurso, e sem o amor, o apoio e a paciência dos quais, este trabalho não teria sido possível.

A todos,

Muito Obrigado

Resumo

Com o intuito de modelar longitudinalmente trajetórias de privação material, com dados provenientes do painel rotativo ICOR (Inquérito às Condições de Vida e ao Rendimento), são apresentados resultados de diversos estudos, tendo por base diferentes opções metodológicas no *framework* dos modelos com equações estruturais, designadamente modelos com trajetória latente.

Para lidar com as omissões resultantes da rotatividade do painel são utilizadas as abordagens estatísticas *full information maximum likelihood* e imputação múltipla, assumindo um mecanismo de omissão de dados ignorável. Porém, neste painel existem não respostas, cujo mecanismo de omissão pode ser considerado não ignorável. Assim, foram também utilizadas as abordagens *pattern mixture* e *selection models*.

Foi realizado um estudo de simulação, recorrendo ao pacote estatístico Mplus, no sentido de avaliar o efeito da existência de omissões, de desvios face à normalidade da distribuição dos dados, do número de momentos temporais, da dimensão da amostra e do método de estimação, nas estimativas dos parâmetros do modelo.

Assim, os principais contributos metodológicos desta tese são a modelação longitudinal de dados provenientes de um painel rotativo com recurso a diferentes abordagens estatísticas para lidar com as omissões, o desenvolvimento de rotinas para implementação de um estudo de simulação em Mplus, e a apresentação de um conjunto de orientações a serem seguidas por um investigador que pretenda usar modelos com trajetória latente em painéis rotativos com dados omissos.

Palavras chave: Dados Omissos; Modelos com Trajetória Latente; Painel Rotativo; Estudo de Simulação

Classificação JEL: C33; C63

Abstract

In order to model longitudinal trajectories of material deprivation, with data from the rotating panel ICOR (Statistics on Income and Living Conditions), different studies results are presented, based on distinct methodological options in the framework of structural equation models, namely latent growth curve models.

To deal with the non-responses resulting from the rotation of the panel the statistical approaches full information maximum likelihood and multiple imputation are used, if is assumed an ignorable missing data mechanism. However, there is another type of non-responses in this panel whose missing data mechanism cannot be considered ignorable. Thus the approaches pattern mixture and selection models were also considered.

A simulation study was performed using the statistical software Mplus, in order to evaluate the effect of the existence of missing data, the deviations from the normality distribution, the number of time points, the sample size and estimation method, on model parameters estimates.

Thus, the main methodological contributions of this thesis are modeling longitudinal data from a rotating panel using different statistical approaches in order to deal with the non-responses, the development of routines for implementing a simulation study in Mplus, and presenting a set of guidelines to be followed by a researcher wishing to use latent growth curve models in rotating panels with missing data.

Keywords: Latent Growth Curve Models; Missing Data; Rotating Panel; Simulation Study

JEL Classification System: C33; C63

Conteúdo

Glossário de termos	xvii
Glossário de notação	xix
1 Introdução	1
1.1 Dados longitudinais	3
1.2 Inquérito às Condições de Vida e ao Rendimento - ICOR	5
1.3 Modelos com variáveis latentes contínuas para análise de dados longitudinais	6
1.4 Dados omissos	7
1.5 Privação material	8
1.6 Objetivo	9
1.7 Apresentação dos estudos	11
1.7.1 Estudos com dados completos	12
1.7.2 Estudos com dados com omissões MAR	14
1.7.3 Estudos com dados com omissões MNAR	15
1.7.4 Estudo de simulação	17
1.8 Estrutura da tese	17
2 O Painel ICOR e a Medição da Privação Material	19
2.1 O Painel ICOR	19
2.2 A Medição da Privação Material	23
3 Modelações com Dados Completos	35
3.1 Modelos com Trajetória Latente: evolução histórica, estimação do modelo e medidas de qualidade de ajustamento	35
3.2 Modelos de Análise Fatorial Confirmatória: Um caso particular de um modelo com equações estruturais	42
3.2.1 Modelos de Análise Fatorial Confirmatória Multigrupos	42
3.2.2 Resultados do estudo da Privação Material utilizando um Modelo de Análise Fatorial Confirmatória Multigrupos	44
3.3 Modelos com Equações Estruturais	48
3.3.1 Modelos com Equações Estruturais com Efeitos Cruzados e Desfasados	48
3.3.2 Resultados do estudo da Privação Material utilizando um Modelo com Equações Estruturais com Efeitos Cruzados e Desfasados . . .	49
3.4 Modelos com Trajetória Latente	50

3.4.1	Modelos com Trajetória Latente de Primeira Ordem	50
3.4.2	Resultados do estudo de Privação Material utilizando um Modelo com Trajetória Latente Bivariada	59
3.4.3	Modelos com Trajetória Latente de Segunda Ordem	60
3.4.4	Resultados do estudo da Privação Material utilizando um Modelo com Trajetória Latente de Segunda Ordem com Trajetórias Condi- cionadas	64
4	Modelações considerando a existência de dados omissos	69
4.1	Omissões por desenho do estudo	69
4.2	Mecanismo de omissão dos dados	74
4.3	Modelos para lidar com omissões MCAR ou MAR	76
4.3.1	<i>Full Information Maximum Likelihood</i>	77
4.3.2	Resultados do estudo da Privação Material utilizando <i>Full Infor- mation Maximum Likelihood</i>	78
4.3.3	Imputação Múltipla	83
4.3.4	Resultados do estudo utilizando <i>Full Information Maximum Likeli- hood</i> e Imputação Múltipla	86
4.4	Modelos para lidar com omissões MNAR	92
4.4.1	<i>Pattern Mixture Models</i>	95
4.4.2	<i>Selection Models</i>	96
4.4.3	Resultados do estudo utilizando <i>Pattern Mixture Models</i> e <i>Selection Models</i>	99
5	Estudo de simulação	113
5.1	Aspectos introdutórios e principais etapas	113
5.1.1	Objetivo do estudo	115
5.1.2	O modelo	115
5.1.3	Software utilizado	116
5.2	Opções metodológicas	117
5.2.1	Planejar as questões de investigação	117
5.2.2	Valores dos parâmetros populacionais	119
5.2.3	Critérios de avaliação	123
5.3	Resultados do estudo de simulação para dados gerados com distribuição normal	124
5.3.1	Dados gerados com distribuição normal - completos	125
5.3.2	Dados gerados com distribuição normal - com omissões	146
5.4	Resultados do estudo de simulação para dados gerados a partir de uma mistura de distribuições normais	155
5.4.1	Dados gerados a partir de uma mistura de distribuições normais - completos	155
5.4.2	Dados gerados a partir de uma mistura de distribuições normais - com omissões	168
5.5	Principais conclusões do estudo de simulação	173
5.5.1	Conclusões do estudo para dados gerados com distribuição normal .	173

5.5.2	Conclusões do estudo para dados gerados a partir de uma mistura de distribuições normais	178
5.5.3	Conclusões do estudo para dados gerados com distribuição normal vs dados gerados a partir de uma mistura de distribuições normais .	181
6	Conclusões	183
6.1	Principais conclusões dos estudos realizados	184
6.2	Principais limitações	189
6.3	Perspetivas para trabalho futuro	190
	Bibliografia	193
	Anexo A Fórmulas das medidas de qualidade do ajustamento modelo-dados disponíveis no Mplus	205
	Anexo B Comandos de sintaxe em Mplus	209
	Anexo C Diagrama e tabela de Análise Fatorial Confirmatória	229
	Anexo D Tabelas do estudo de simulação: Dados com o desenho das subamostras 2 a 7	231
	Anexo E Diagrama de um procedimento <i>Bootstrap</i>	233
	Anexo F Tabelas do estudo de simulação: Dados com distribuição normal e $T=3$	235
	Anexo G Tabelas do estudo de simulação: Dados com distribuição mistura de normais e $T=3$	243
	Anexo H Tabelas do estudo de simulação: Dados com distribuição normal e $T=5$	247
	Anexo I Tabelas do estudo de simulação: Dados com distribuição mistura de normais e $T=5$	255

Lista de Figuras

2.1	ICOR - Desenho do painel (número de agregados familiares)	20
2.2	Gráficos das quatro dimensões da privação material: constrangimentos económicos, posse de bens duradouros, condições da habitação e vizinhança, nos anos de 2007 e 2008	25
2.3	ICOR-Desenho do estudo, dentro da caixa os agregados considerados no estudo 4	29
2.4	Percentagem de agregados com itens em privação, para o caso completo (845 agregados), em cada um dos quatro momentos temporais, 2006 a 2009	30
2.5	Percentagem de agregados com itens em privação, para o caso com todas as observações (1693 agregados), em cada um dos quatro momentos temporais, 2006 a 2009	30
2.6	Desenho amostral do ICOR e diferentes subamostras consideradas no estudo 5	31
2.7	Percentagem de agregados com diferente número de itens em privação, nas várias subamostras consideradas, nos anos de 2006 a 2009	33
2.8	Número médio de itens em privação em cada ano, de 2006 a 2009, para os diferentes padrões de omissão	34
3.1	Diagrama de um modelo com equações estruturais com quatro variáveis latentes independentes (ξ) e três dependentes (η)	43
3.2	Diagrama de um modelo de análise fatorial confirmatória com a privação material medida por quatro indicadores	45
3.3	Diagrama de um modelo de análise fatorial confirmatória com a privação material medida por 3 indicadores	46
3.4	Diagrama de um modelo com equações estruturais com a privação material medida por 3 indicadores e 3 variáveis explicativas	47
3.5	Diagrama de um modelo com equações estruturais com efeitos cruzados e desfasados	49
3.6	Diagrama de um modelo com equações estruturais com efeitos cruzados e desfasados	49
3.7	Diagrama de um modelo com trajetória latente	51
3.8	Diagrama de um modelo com trajetória latente com variáveis y_t ordinais, sendo y_{it}^* as variáveis latentes contínuas subjacentes	57
3.9	Diagrama de um modelo com trajetória latente bivariada para dois processos y e w	57

3.10	Diagrama de um modelo com trajetória latente com dois processos: capacidade financeira e posse de bens duradouros	60
3.11	Diagrama de um modelo com trajetória latente de segunda ordem, em que a variável latente η é medida por três indicadores, em quatro momentos temporais. Os termos residuais das medidas repetidas do primeiro indicador encontram-se correlacionados	61
3.12	Diagrama de um modelo com trajetória latente de segunda ordem condicionada, com variável explicativa invariante (x) e variável explicativa variante no tempo (w_t)	63
3.13	Diagrama de um modelo com trajetória latente de segunda ordem da privação material medida por 3 dimensões consideradas como variáveis ordinais	65
3.14	Diagrama de um modelo com trajetória latente de segunda ordem condicionada para a privação material. As variáveis explicativas são consideradas como não variantes no tempo	65
3.15	Diagrama de um modelo com trajetória latente de segunda ordem condicionada para a privação material. A variável explicativa da trajetória, logaritmo do rendimento, é considerada como variante no tempo. As variáveis dimensão do agregado e área de urbanização são consideradas invariantes	67
4.1	Diagrama de um modelo com trajetória latente	79
4.2	Diagrama das três etapas da imputação múltipla	83
4.3	Desenho amostral do ICOR e diferentes subamostras consideradas no estudo	87
4.4	Diagrama de um <i>Pattern Mixture Model</i> - abordagem multigrupos (c corresponde aos grupos pré-definidos)	96
4.5	Diagrama de um <i>Pattern Mixture Model</i> - parametrização convencional	97
4.6	Diagrama de um <i>Selection Model</i> - abordagem de Wu-Carroll	98
4.7	Diagrama de um <i>Selection Model</i> - abordagem de Diggle-Kenward	99
4.8	ICOR-Desenho do painel, dentro da caixa azul os agregados considerados no estudo, dentro da caixa encarnada o caso completo	100
4.9	Trajетórias estimadas para cada um dos dois padrões considerados: “caso completo+MAR” (a encarnado) e “MNAR” (a azul)	107
4.10	Trajетórias estimadas para cada um dos três padrões considerados: “caso completo”(a encarnado), “MAR”(a azul) e “MNAR”(a verde)	109
4.11	Trajетórias estimadas para cada um dos quatro padrões considerados: “caso completo”(a encarnado), “entradas tardias”(a azul), “saídas antes do tempo”(a verde) e “outras”(a preto)	110
5.1	Distribuição univariada das variáveis Y_1, Y_2, Y_3 e Y_4 , na primeira amostra gerada a partir do modelo H1	156
5.2	Distribuição univariada das variáveis Y_1, Y_2, Y_3 e Y_4 , em cada uma das classes definidas pelo modelo H1	156
5.3	Distribuição univariada das variáveis Y_1, Y_2, Y_3 e Y_4 , na primeira amostra gerada a partir do modelo H2	157
5.4	Distribuição univariada das variáveis Y_1, Y_2, Y_3 e Y_4 , em cada uma das classes definidas pelo modelo H2	157

5.5	Distribuição univariada das variáveis Y_1, Y_2, Y_3 e Y_4 , na primeira amostra gerada a partir do modelo H3	157
5.6	Distribuição univariada das variáveis Y_1, Y_2, Y_3 e Y_4 , em cada uma das classes definidas pelo modelo H3	157
5.7	Distribuição univariada das variáveis Y_1, Y_2, Y_3 e Y_4 , na primeira amostra gerada a partir do modelo H4	158
5.8	Distribuição univariada das variáveis Y_1, Y_2, Y_3 e Y_4 , em cada uma das classes definidas pelo modelo H4	158
5.9	Distribuição univariada das variáveis Y_1, Y_2, Y_3 e Y_4 , na primeira amostra gerada a partir do modelo H5	158
5.10	Distribuição univariada das variáveis Y_1, Y_2, Y_3 e Y_4 , em cada uma das classes definidas pelo modelo H5	158
C.1	Diagrama de um modelo de análise fatorial confirmatória com a privação material medida por 3 indicadores (CE, PBD e CH)	229
E.1	Diagrama de um procedimento de <i>Bootstrap</i>	233

Lista de Tabelas

1.1	Abordagens estatísticas, variáveis e o tipo de dados utilizados em cada um dos estudos que compõem esta tese	12
2.1	Percentagem de agregados em privação material (mais itens corresponde a maior privação), nos anos de 2007 e 2008	27
2.2	Percentagem de observações omissas nas diferentes subamostras e a respectiva dimensão, para os anos de 2006 a 2009	32
2.3	Diferentes padrões de omissão existentes na amostra em estudo (\checkmark = valor observado; - = valor omissão)	33
3.1	Medidas de ajustamento: AFC da privação material medida por quatro dimensões	45
3.2	Estimativas, numa solução estandardizada, dos pesos fatoriais (valores dos testes t): AFC da privação material medida por quatro dimensões. A <i>bold</i> estão os valores que se mostraram significativos.	45
3.3	Estimativas, numa solução estandardizada, (valores do teste t) dos pesos fatoriais: AFC da privação material medida por três dimensões. A <i>bold</i> estão os valores que se mostraram significativos.	46
3.4	Efeito das variáveis explicativas (valores dos testes t) sobre a privação material medida por três dimensões. A <i>bold</i> estão os valores que se mostraram significativos	47
3.5	Estimativas (valores dos testes t), numa solução estandardizada, para os parâmetros de um modelo com equações estruturais com efeitos cruzados e desfasados. A <i>bold</i> estão os valores que se mostraram significativos, a um nível de 5%.	50
3.6	Estimativas (valores do teste t) para os parâmetros aleatórios do modelo com trajetória latente bivariada. A <i>bold</i> estão os valores que se mostraram significativos, a um nível de 5%.	60
3.7	Estimativas (valores dos testes t) da média, variâncias e covariância do intercepto e declive para um modelo com trajetória latente de segunda ordem para a privação material. A <i>bold</i> estão os valores que se mostraram significativos	66

3.8	Estimativas (valores dos testes t) dos coeficientes de regressão das variáveis explicativas do intercepto e do declive para um modelo com trajetória latente de segunda ordem condicionada. A <i>bold</i> estão os valores que se mostraram significativos	67
3.9	Estimativas (valores do teste t) dos coeficientes de regressão da variável explicativa (logaritmo do) rendimento do fator latente privação material, e dos coeficientes de regressão das variáveis explicativas dimensão do agregado e área de urbanização do intercepto e do declive, para um modelo com trajetória latente de segunda ordem condicionada. A <i>bold</i> estão os valores que se mostraram significativos	68
4.1	<i>Multiple matrix design</i> (✓ = questão colocada; - = questão não colocada; A a G dizem respeito a diferentes grupos de questões)	70
4.2	<i>3-form design</i> (✓ = questão colocada; - = questão não colocada)	72
4.3	<i>Split questionnaire design</i> (✓ = questão colocada; - = questão não colocada)	72
4.4	<i>Cohort-sequential design</i> (✓ = questão colocada; - = questão não colocada)	73
4.5	Medidas da qualidade do ajustamento (“caso completo” e “caso com todas as observações”)	79
4.6	Estimativas dos parâmetros obtidas no ajustamento de um modelo com trajetória latente (“caso completo” e “caso com todas as observações”). A <i>bold</i> estão os valores que se mostraram significativos	79
4.7	Resultados obtidos no estudo de simulação gerando 1 000 amostras com 10% de omissões no ano de 2009	81
4.8	Resultados obtidos no estudo de simulação gerando 1 000 amostras com 20% de omissões no ano de 2009	81
4.9	Resultados obtidos no estudo de simulação gerando 1 000 amostras com 30% de omissões no ano de 2009	81
4.10	Resultados obtidos no estudo de simulação gerando 1 000 amostras com 40% de omissões no ano de 2009	82
4.11	Resultados obtidos no estudo de simulação gerando 1 000 amostras com 50% de omissões no ano de 2009	82
4.12	Enviesamento relativo na estimativa média dos parâmetros quando são geradas 1 000 amostras com 50%, 40%, 30%, 20% e 10% de omissões no ano de 2009	82
4.13	Percentagem de observações omissas nas diferentes subamostras e a respetiva dimensão	87
4.14	Medidas de qualidade do ajustamento para as diferentes subamostras consideradas. Na imputação múltipla foi considerada a imputação de 5 e 20 valores distintos, abordagens IM(5) e IM(20), respetivamente	90
4.15	Estimativas dos parâmetros obtidas no ajustamento de um modelo com trajetória latente a diferentes subamostras. A <i>bold</i> estão os valores que se mostraram significativos	91
4.16	Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 1	92

4.17	Diferentes padrões de omissão existentes na base de dados em estudo (\checkmark = valor observado; - = valor omissão)	100
4.18	Medidas da qualidade do ajustamento (caso completo e com todas as observações)	101
4.19	Estimativas dos parâmetros obtidas no ajustamento de um modelo com trajetória latente (“caso completo” e “caso com todas as observações”). A <i>bold</i> estão os valores que se mostraram significativos;* 1% nível de significância.	102
4.20	Caso completo: Intervalos de confiança não paramétricos obtidos por <i>bootstrap</i>	103
4.21	Caso com todas as observações: Intervalos de confiança não paramétricos obtidos por <i>bootstrap</i>	103
4.22	Estimativas dos parâmetros obtidas recorrendo ao <i>Pattern Mixture Model</i> (parametrização multigrupos - 2 grupos). A <i>bold</i> estão os valores que se mostraram significativos; * 1% nível de significância.	104
4.23	Estimativas dos parâmetros obtidas recorrendo ao <i>Pattern Mixture Model</i> (parametrização multigrupos - 4 grupos). A <i>bold</i> estão os valores que se mostraram significativos; *1% nível de significância.	105
4.24	Estimativas dos parâmetros obtidas recorrendo ao <i>Pattern Mixture Model</i> (3 grupos-restrição 1). A <i>bold</i> estão os valores que se mostraram significativos.	105
4.25	Estimativas dos parâmetros obtidas recorrendo ao <i>Pattern Mixture Model</i> (3 grupos-restrição 2). A <i>bold</i> estão os valores que se mostraram significativos.	106
4.26	Estimativas dos parâmetros obtidas recorrendo ao <i>Pattern Mixture Model</i> (3 grupos-restrição 3). A <i>bold</i> estão os valores que se mostraram significativos.	106
4.27	Medidas de qualidade do ajustamento utilizando <i>Pattern Mixture Model</i>	106
4.28	Estimativas dos parâmetros obtidas recorrendo ao <i>Pattern Mixture Model</i> (parametrização multigrupos - 3 grupos). A <i>bold</i> estão os valores que se mostraram significativos; *1% nível de significância.	108
4.29	Estimativas dos parâmetros obtidas recorrendo ao <i>Pattern Mixture Model</i> (parametrização convencional - 3 grupos). A <i>bold</i> estão os valores que se mostraram significativos; *1% nível de significância.	109
4.30	Estimativas dos parâmetros obtidas recorrendo ao <i>Pattern Mixture Model</i> (parametrização convencional - 4 grupos). A <i>bold</i> estão os valores que se mostraram significativos;* 1% nível de significância.	110
4.31	Medidas de qualidade do ajustamento utilizando um <i>Pattern Mixture Model</i> (parametrização multigrupos)	111
4.32	Medidas de qualidade do ajustamento utilizando um <i>Pattern Mixture Model</i> (parametrização convencional)	111
5.1	<i>Planned missing design</i> para modelos com $T = 3$ (\checkmark = valor observado; - = valor omissão)	117
5.2	<i>Planned missing design</i> para modelos com $T = 4$ (\checkmark = valor observado; - = valor omissão)	118
5.3	<i>Planned missing design</i> para modelos com $T = 5$ (\checkmark = valor observado; - = valor omissão)	118

5.4	Parâmetros dos Modelos A1 a G4, para cada um dos modelos $\psi_{\alpha\beta}$ é função da correlação (α, β) , a qual assumiu valores -0.9, -0.5, 0, 0.5, 0.9	120
5.5	Parâmetros para os Modelos H1 a H5	121
5.6	Níveis de assimetria e achatamento	122
5.7	Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros, dos modelos A, com correlação $(\alpha, \beta) = -0.9$	127
5.8	Erro quadrático médio para os parâmetros dos modelos A com correlação $(\alpha, \beta) = -0.9$	128
5.9	<i>Coverage</i> para os parâmetros dos modelos A, com correlação $(\alpha, \beta) = -0.9$.	128
5.10	Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros, dos modelos A, com correlação $(\alpha, \beta) = -0.5$	129
5.11	Erro quadrático médio para os parâmetros dos modelos A com correlação $(\alpha, \beta) = -0.5$	130
5.12	<i>Coverage</i> para os parâmetros dos modelos A, com correlação $(\alpha, \beta) = -0.5$.	130
5.13	Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros, dos modelos A, com correlação $(\alpha, \beta) = 0$	132
5.14	Erro quadrático médio para os parâmetros dos modelos A com correlação $(\alpha, \beta) = 0$	132
5.15	<i>Coverage</i> para os parâmetros dos modelos A, com correlação $(\alpha, \beta) = 0$. .	133
5.16	Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A, com correlação $(\alpha, \beta) = 0.5$	134
5.17	Erro quadrático médio para os parâmetros dos modelos A, com correlação $(\alpha, \beta) = 0.5$	134
5.18	<i>Coverage</i> para os parâmetros dos modelos A, com correlação $(\alpha, \beta) = 0.5$. .	135
5.19	Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A, com correlação $(\alpha, \beta) = 0.9$	136
5.20	Erro quadrático médio para os parâmetros dos modelos A com correlação $(\alpha, \beta) = 0.9$	136
5.21	<i>Coverage</i> para os parâmetros dos modelos A com correlação $(\alpha, \beta) = 0.9$. .	137
5.22	Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A, com fiabilidade dos indicadores de 0.4	141
5.23	Erro quadrático médio para os parâmetros dos modelos A com fiabilidade dos indicadores de 0.4	142
5.24	<i>Coverage</i> para os parâmetros dos modelos A com fiabilidade dos indicadores de 0.4	142
5.25	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.5$, $T = 4$ e omissões	148
5.26	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0$, $T = 4$ e omissões	148
5.27	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.5$, $T = 4$ e omissões	149

5.28	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos A1 e A4, com correlação(α, β) = 0.5, $T = 3$ e omissões	151
5.29	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos A1 e A4, com correlação(α, β) = 0, $T = 3$ e omissões	152
5.30	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos A1 e A4, com correlação(α, β) = -0.5, $T = 3$ e omissões	152
5.31	Valores da assimetria e achatamento para cada uma das variáveis univariadas geradas pelos modelos H1 a H5 (com 100 000 observações)	159
5.32	Enviesamento bruto (enviesamento relativo) da estimativa média para os parâmetros dos modelos H1 a H5, e $T = 4$	160
5.33	Erro quadrático médio de estimação para os parâmetros dos modelos H1 a H5, e $T = 4$	161
5.34	<i>Coverage</i> para os parâmetros dos modelos H1 a H5, e $T = 4$	162
5.35	<i>Coverage</i> para os parâmetros dos modelos H1 a H5, com $T = 4$ e estimador não robusto	163
5.36	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos H1 e H5, com $T = 4$ e omissões	169
C.1	Estimativas, numa solução estandardizada, dos pesos fatoriais, para os quatro modelos de AFC	229
D.1	Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 2	231
D.2	Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 3	231
D.3	Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 4	232
D.4	Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 5	232
D.5	Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 6	232
D.6	Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 7	232
F.1	Enviesamento bruto(enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = -0.9 e $T = 3$	235
F.2	Enviesamento bruto(enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = -0.5 e $T = 3$	235
F.3	Enviesamento bruto(enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = 0 e $T = 3$	236
F.4	Enviesamento bruto(enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = 0.5 e $T = 3$	236
F.5	Enviesamento bruto(enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = 0.9 e $T = 3$	236

F.6	Erro quadrático médio dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.9$ e $T = 3$	237
F.7	Erro quadrático médio dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.5$ e $T = 3$	237
F.8	Erro quadrático médio dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0$ e $T = 3$	237
F.9	Erro quadrático médio dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.5$ e $T = 3$	238
F.10	Erro quadrático médio dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.9$ e $T = 3$	238
F.11	<i>Coverage</i> dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.9$ e $T = 3$	238
F.12	<i>Coverage</i> dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.5$ e $T = 3$	239
F.13	<i>Coverage</i> dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0$ e $T = 3$	239
F.14	<i>Coverage</i> dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.5$ e $T = 3$	239
F.15	<i>Coverage</i> dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.9$ e $T = 3$	240
F.16	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.5$, $T = 3$ e omissões	240
F.17	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0$, $T = 3$ e omissões	240
F.18	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.5$, $T = 3$ e omissões	241
G.1	Enviesamento bruto (enviesamento relativo) na estimativa média dos parâmetros dos modelos H1 a H5, e $T = 3$	243
G.2	Erro quadrático médio de estimação para os parâmetros, dos modelos H1 a H5, e $T = 3$	244
G.3	<i>Coverage</i> para os parâmetros, dos modelos H1 a H5, e $T = 3$	244
G.4	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos H1 e H5, com $T = 3$ e omissões	245
H.1	Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.9$ e $T = 5$	247
H.2	Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.5$ e $T = 5$	247
H.3	Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0$ e $T = 5$	248
H.4	Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.5$ e $T = 5$	248
H.5	Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.9$ e $T = 5$	248
H.6	Erro quadrático médio, dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.9$ e $T = 5$	249

H.7	Erro quadrático médio, dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.5$ e $T = 5$	249
H.8	Erro quadrático médio, dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0$ e $T = 5$	249
H.9	Erro quadrático médio, dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.5$ e $T = 5$	250
H.10	Erro quadrático médio, dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.9$ e $T = 5$	250
H.11	<i>Coverage</i> , dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.9$ e $T = 5$. . .	250
H.12	<i>Coverage</i> , dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.5$ e $T = 5$. . .	251
H.13	<i>Coverage</i> , dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0$ e $T = 5$	251
H.14	<i>Coverage</i> , dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.5$ e $T = 5$	251
H.15	<i>Coverage</i> , dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.9$ e $T = 5$	252
H.16	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0.5$, $T = 5$ e omissões	252
H.17	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos A1 e A4, com correlação $(\alpha, \beta) = 0$, $T = 5$ e omissões	252
H.18	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.5$, $T = 5$ e omissões	253
I.1	Enviesamento bruto (enviesamento relativo) na estimativa média dos parâmetros dos modelos H1 a H5, e $T = 5$	255
I.2	Erro quadrático médio de estimação para os parâmetros, dos modelos H1 a H5, e $T = 5$	256
I.3	<i>Coverage</i> para os parâmetros, dos modelos H1 a H5, e $T = 5$	257
I.4	Enviesamento bruto (enviesamento relativo), erro quadrático médio e <i>coverage</i> para os parâmetros, dos modelos H1 e H5, com $T = 5$ e omissões . . .	258

Glossário de termos

Akaike Information Criteria (AIC)	Critério de informação de Akaike
Available case missing variable restriction	Restrição de valores omissos de informação disponível
Balanced incomplete blocks spiral design	Planeamento de blocos incompletos fracionados em espiral
Bayesian Information Criteria (BIC)	Critério de informação de Bayes
Cohort-sequential design	Planeamento com grupos sequenciais
Comparative Fit Index (CFI)	Índice de ajustamento comparado
Complete case missing variable restriction	Restrição de valores omissos de casos completos
Coverage	Cobertura
Data augmentation algorithm	Algoritmo de ampliação de dados
Data fusion design	Planeamento de fusão de dados
Dropout	Atrito, Desistência
Dummy variable	Variável muda ou fictícia
Expectation-Maximization algorithm	Algoritmo de Estimaco-Maximizaco
Fractional block design	Planeamento de blocos fracionados
Full Information Maximum Likelihood (FIML)	Mxima verosimilhana de informaco completa
Item Response Theory (IRT)	Teoria da resposta ao item
Item non response	No resposta ao item
Listwise deletion	Eliminaco <i>Listwise</i>
Matrix sampling design	Planeamento da matriz amostral
Missing Completely At Random (MCAR)	Omisso completamente aleatria
Missing At Random (MAR)	Omisso aleatria
Missing Not At Random (MNAR)	Omisso no aleatria
Multiple matrix design	Planeamento de matriz mltipla
Neighboring case missing variable restriction	Restrição de valores omissos de casos vizinhos
Pairwise deletion	Eliminaco <i>Pairwise</i>
Pattern mixture model	Modelo de mistura de padres
Planned missing design	Desenho com omisses planeadas

Root Mean Square Error of Approximation(RMSEA)	Raiz do erro quadrático médio de aproximação
Root Mean Square Residual(RMR)	Raiz do resíduo quadrático médio
Selection model	Modelo de seleção
Split questionnaire design	Planeamento de questionário subdividido
Standardized Root Mean Square Residual (SRMR)	Raiz padronizada do resíduo quadrático médio
Subsampling design	Planeamento de subamostras
Threshold model	Modelo com limiar
Time sampling design	Planeamento de amostras no tempo
Unit non response	Não resposta à unidade
Weighted Root Mean Square Residual (WRMR)	Raiz ponderada do resíduo quadrático médio
3-form design	Planeamento <i>3-form</i>

Glossário de notação

θ	parâmetros do modelo (em termos genéricos)
Σ	matriz de variâncias-covariâncias populacionais
S	matriz de variâncias-covariâncias amostrais
μ	vetor das médias populacionais
X	vetor das variáveis observadas das latentes independentes (modelo com equações estruturais)
Y	vetor das variáveis observadas das latentes dependentes (modelo com equações estruturais)
ξ	vetor das variáveis latentes independentes (modelo com equações estruturais)
η	vetor das variáveis latentes dependentes (modelo com equações estruturais)
δ e ε	vetores dos termos residuais dos modelos de medida, X e Y, respectivamente (modelo com equações estruturais)
ζ	vetor dos termos residuais das relações estruturais (modelo com equações estruturais)
Λ_X e Λ_Y	matrizes dos parâmetros (pesos fatoriais) nos modelos de medida (modelo com equações estruturais)
B e Γ	matrizes dos parâmetros nas relações estruturais (modelo com equações estruturais)
Φ	matriz de variâncias-covariâncias das variáveis latentes independentes (modelo com equações estruturais)
Ψ	matriz de variâncias-covariâncias das variáveis latentes dependentes (modelo com equações estruturais)
Θ_δ e Θ_ε	matrizes de variâncias-covariâncias dos termos residuais dos modelos de medida (modelo com equações estruturais)
k	número de variáveis
$t = 1, \dots, T$	número de momentos temporais
$i = 1, \dots, N$	número de elementos
y_{it}	variável observada i , no momento t
α_i	intercepto aleatório do elemento i
β_i	declive aleatório do elemento i
ε_{it}	termo residual no modelo de trajetória latente de primeira ordem

λ_t	constante que define a trajetória, $\lambda_t = t - 1$ se a trajetória é linear
Θ_ε	matriz de variâncias-covariâncias do termo residual (modelo com trajetória latente)
$\theta_{\varepsilon t}$	elementos da diagonal da matriz Θ_ε (modelo com trajetória latente)
μ_α	média do intercepto
μ_β	média do declive
ζ_{α_i}	variabilidade no momento inicial (termo residual do intercepto)
ζ_{β_i}	variabilidade em torno da taxa de crescimento (termo residual do declive)
$\psi_{\alpha\alpha}$	variância do intercepto
$\psi_{\beta\beta}$	variância do declive
$\psi_{\alpha\beta}$	covariância entre o intercepto e o declive
l	número de fatores latentes do modelo
\mathbf{Y}	vetor das observações
Λ	matriz dos pesos fatoriais
$\boldsymbol{\eta}$	vetor dos fatores latentes (modelo com trajetória latente)
$\boldsymbol{\mu}_\eta$	vetor das médias dos fatores latentes
$\boldsymbol{\varepsilon}$	vetor dos termos residuais (modelo com trajetória latente)
$\boldsymbol{\zeta}$	vetor dos termos residuais dos fatores latentes (modelo com trajetória latente)
$c = 1, \dots, C$	número de categorias ordenadas (modelo de thresholds)
τ_{c-1} e τ_c	thresholds inferior e superior da categoria c , respetivamente
Ψ_{yw}	matriz de variâncias-covariâncias entre intercepto e declive de dois processos (y e w) (modelo com trajetória latente bivariada)
$\Theta_{\varepsilon\varepsilon}$	matriz de variâncias-covariâncias dos termos residuais de dois processos (modelo com trajetória latente bivariada)
Λ_{jt}	peso fatorial do indicador j no momento t (modelo com trajetória latente de segunda ordem)
ν_{jt}	intercepto do indicador j , no momento t
y_{ijt}	valor do indicador j , para o elemento i , no momento t
η_{it}	variável latente para o elemento i , no momento t (modelo com trajetória latente de segunda ordem)
v_{ijt}	termo residual no modelo com trajetória latente de segunda ordem
γ_α e γ_β	coeficientes de regressão dos efeitos aleatórios sobre a variável x
x	variável explicativa invariante no tempo
γ_t	coeficiente de regressão das medidas repetidas de η_t sobre as medidas repetidas da variável explicativa w_t
w_t	variável explicativa variante no tempo
\mathbf{Y}_{obs}	valores observados do vetor \mathbf{Y}
\mathbf{Y}_{miss}	valores omissos do vetor \mathbf{Y}
\mathbf{Y}_i	vetor das observações para o elemento i
L	função de verosimilhança
L_i	função de verosimilhança para o elemento i

$B_R(\hat{\theta})$	enviesamento relativo do valor estimado de θ
$B(\hat{\theta})$	enviesamento absoluto do valor estimado de θ
h	iteração nos passos I e P do procedimento de imputação múltipla
\mathbf{Y}_h^*	valores imputados no passo I na iteração h
$\boldsymbol{\theta}_h^*$	valores dos parâmetros simulados do passo P iteração h
$\boldsymbol{\Sigma}^*$	matriz de variâncias-covariâncias simulada
$\boldsymbol{\mu}^*$	vetor de médias simuladas
m	número de amostras geradas
\mathbf{R}	indicador de omissão
n_i	número de observações para o elemento i
$\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$	estimativa para o parâmetro de interesse em cada uma das B amostras de <i>bootstrap</i>
Φ	parâmetros que descrevem a propensão para as omissões (numa abordagem para lidar com dados omissos MNAR)

Capítulo 1

Introdução

Nas últimas décadas tem aumentado o interesse pela recolha e análise de dados longitudinais, isto é, dados que medem a evolução dos acontecimentos durante um determinado período de tempo. Como consequência deste crescente interesse, o número de publicações reportando resultados de análises de dados longitudinais, nas mais variadas áreas, como por exemplo, psicologia, sociologia, medicina, entre outras, tem seguido esta tendência.

De acordo com vários autores, Singer e Willett [145], Bollen e Curran [21], Rogosa et al. [127], Finkel [48] e Duncan et al. [34], os dados longitudinais são considerados indispensáveis para examinar questões de causalidade e mudança. Todavia, deve ser salientada a diferença entre estudo longitudinal e dados longitudinais, sendo possível uma análise longitudinal utilizando dados transversais (Taris [152], McArdle [88]). Por exemplo, para aferir sobre a qualidade dos serviços de um determinado centro de ocupação de tempos livres para crianças, não será necessário inquirir os mesmos pais em distintos momentos temporais para avaliar como se alterou a qualidade dos serviços neste centro. Por outro lado, em estudos na área do marketing é frequente a utilização de um conjunto de consumidores, usualmente designado por painel de consumidores, os quais são inquiridos em vários momentos temporais, obtendo-se assim dados longitudinais. Contudo, nem sempre estes dados são analisados longitudinalmente.

Uma das grandes vantagens dos dados longitudinais, apontada por Hsiao [69], é a possibilidade de analisar alterações ou mudanças de fenómenos sociais, económicos ou ainda de outra natureza, bem como considerar essas mudanças ao nível individual. No seu livro, Singer e Willett [145] salientam as vantagens destes relativamente aos dados transversais dizendo: “... *research questions about change and event occurrence can be addressed with longitudinal data. In doing so, we reveal research opportunities unavailable in the world of cross-sectional data*”. Na área da economia Verbeek [160] enfatiza as vantagens dos dados longitudinais da seguinte forma: “*The availability of repeated observations on the same units allows economists to specify and estimate more complicated and more realistic models than a single cross-section or a single time series would do*”.

Por outro lado, numa introdução aos métodos de pesquisa em Ciências Sociais e do

Comportamento sobre estudos analisando dados longitudinais Robson [126] diz: “... *it tends to be difficult to carry out and is demanding on the time and resources of the investigator*”, para evidenciar que o desenvolvimento computacional foi uma contribuição decisiva no crescente aumento do número de bases de dados longitudinais disponíveis para análise, e consequente aumento do número de publicações reportando resultados da modelação deste tipo de dados.

De acordo com Skrondal e Rabe-Hesketh [148], também a análise de dados com estrutura latente representa uma forte tendência na modelação estatística em diferentes áreas, aplicando-se sempre que existe a necessidade de introduzir a noção de variável latente, uma vez que conceitos, como a definição de um trabalhador como “eficiente”, ou um estudante como “brilhante”, ou ainda a definição de “privação material”, não conseguem ser observados diretamente, sendo medidos através de indicadores. Quando os construtos em causa são considerados como assumindo valores num *continuum* as correspondentes variáveis latentes são contínuas. Exemplos de abordagens estatísticas na presença de variáveis latentes contínuas incluem, entre outros, modelos de análise fatorial confirmatória (Jöreskog [74]), modelos com equações estruturais (Jöreskog [75], McArdle e McDonald [92]) e modelos com trajetória latente (Meredith e Tisak [94]), sendo de destacar estes últimos pelas vantagens que apresentam no estudo da mudança. De acordo com Preacher et al. [117], “... *even more appropriate can be the method of latent growth curve models, with which the data analyst can not only deal with unobserved heterogeneity but also model tendencies ... over the life course*”.

Os dados para modelação estatística podem ser obtidos de várias formas, sendo uma das formas mais utilizada na área das ciências sociais o inquérito. Um inquérito consiste na recolha sistemática de dados, com o objetivo de dar resposta a um determinado problema, baseando-se numa série de perguntas a serem aplicadas a uma amostra que se deseja representativa do grupo que se pretende estudar, num ou em vários momentos temporais (tratando-se neste último caso de um inquérito longitudinal). No entanto, os inquéritos podem apresentar problemas, tais como, a existência de não respostas, a uma pergunta ou a todas as perguntas, e em um ou vários momentos temporais. Pode também acontecer, em inquéritos longitudinais, que os elementos selecionados para a amostra deixem de responder a partir de determinado momento. Por outro lado, o desenho do inquérito pode apresentar omissões que resultam da vontade do investigador, e que são planeadas com o objetivo de não cansar o inquirido e evitar o seu consequente abandono do estudo.

Assim, o desafio que a modelação de dados longitudinais obtidos por inquérito apresenta, devido à existência de diferentes tipos de omissões nos dados, e as vantagens apresentadas na literatura da área sobre a utilização de modelos com trajetória latente no estudo da mudança, são uma forte motivação para a elaboração desta tese, que pretende ser um contributo para auxiliar outros investigadores na modelação longitudinal recorrendo a modelos com trajetória latente, com dados provenientes de um inquérito que apresenta distintos tipos de não resposta. Assim, o desenho do inquérito subjacente à recolha dos dados é uma motivação para este trabalho, devido à particularidade de apre-

sentar omissões planeadas pelo investigador, para além de outros tipos de omissão mais comuns. Em especial, pretende-se também lidar com as omissões resultantes da entrada tardia de elementos no inquérito, omissão essa que não aparece habitualmente descrita na literatura. Salientar as vantagens da utilização dos modelos com trajetória latente na modelação longitudinal de dados face a outros modelos com variáveis latentes contínuas, é outra das contribuições que esta tese pretende dar.

Esta tese é composta por vários estudos que foram sendo realizados ao longo do tempo, à medida que os dados foram sendo disponibilizados pelo INE (Instituto Nacional de Estatística), tendo-se utilizado os dados mais recentes disponíveis no momento da elaboração de cada estudo. Nos primeiros estudos apenas foram consideradas as observações válidas, nos seguintes foram considerados os dados com omissões. Quando comparados os resultados obtidos na modelação longitudinal das observações válidas com os resultados obtidos na modelação longitudinal com dados omissos, estes mostraram-se contraditórios, constituindo este facto motivação para a realização de um estudo de simulação recorrendo a procedimentos de Monte Carlo, e que é apresentado por último nesta tese.

A temática considerada ao longo deste trabalho é a privação material ao nível dos agregados familiares, em grande parte devido à emergência do tema e da necessidade premente que existia de calcular e estudar o comportamento de indicadores de privação material aquando do início desta tese, uma vez que 2010 foi considerado o ano europeu do combate à pobreza e à exclusão social.

No que se refere à estrutura do presente capítulo começa-se por fazer a distinção entre as diferentes formas que os dados longitudinais podem assumir e apresenta-se o painel longitudinal ICOR (Inquérito às Condições de Vida e ao Rendimento) cujos dados são recolhidos por inquérito. De seguida, são apresentados os modelos com variáveis latentes contínuas para dados longitudinais que serão utilizados nesta tese, bem como as diferentes abordagens estatísticas para lidar com as omissões que podem existir num inquérito longitudinal. Por fim, é apresentado o conceito de privação material, o objetivo desta tese e os vários estudos que a compõem, terminando o capítulo com a apresentação da estrutura da tese.

1.1 Dados longitudinais

De acordo com Taris [152] existem, basicamente, dois tipos de dados longitudinais: i) dados em painel; e ii) medidas repetidas de dados transversais. Os dados em painel apresentam informação acerca do que acontece a um mesmo conjunto de indivíduos, agregados ou entidades durante um determinado número de momentos temporais, que são designados por vagas. O número de momentos temporais é geralmente pequeno e o intervalo entre eles pode ser qualquer, de semanas a décadas. O número de participantes no estudo, em cada momento temporal, é usualmente grande, sendo igualmente grande o número observado de variáveis. As medidas repetidas de dados transversais consistem na mesma

informação obtida em vários momentos temporais para conjuntos diferentes de indivíduos.

Por outro lado, numa série cronológica, medidas repetidas são observadas para o mesmo conjunto de participantes tal como acontece para os dados em painel, mas ao contrário destes o número de participantes é limitado e estes são seguidos ao longo do tempo num grande número de ocasiões e para um pequeno número de variáveis (Finkel [48]). Deste modo, os dados de uma série cronológica podem ser considerados um caso particular de dados em painel.

Do mesmo modo, os dados obtidos de um *cohort-sequential design* podem ser considerados dados longitudinais, sendo definidos como conjuntos de indivíduos, por exemplo com idade diferente, que experimentam o mesmo acontecimento dentro do mesmo intervalo de tempo, ou seja, são observados num determinado número de momentos temporais. A análise da mudança é posteriormente realizada ao nível da idade e não do período de tempo em que os indivíduos estiveram no estudo, ou seja, recorre-se à utilização de uma métrica de tempo alternativa (Nesselroad e Baltes [108], Duncan et al. [36]).

Uma das desvantagem dos dados em painel, e uma vez que estes são recolhidos de uma forma prospetiva, é a necessidade de se ter de esperar o passar do tempo para se obter os dados e proceder a uma análise longitudinal, sendo necessário ter informação para pelo menos dois momentos temporais. No entanto, os dados longitudinais podem ser recolhidos num estudo com uma única vaga, colocando questões acerca do que aconteceu no passado, questões ditas retrospectivas.

Outra das desvantagens dos dados em painel é a existência de dados omissos, resultantes do abandono do painel após alguns momentos temporais, isto é, os indivíduos podem recusar continuar a cooperar, os agregados podem ter mudado de residência e deixa de ser possível localizá-los ou as empresas podem ter encerrado a sua atividade, etc. Assim, uma forma de minimizar o problema é considerar um painel rotativo, em que em cada momento temporal (vaga) uma fração dos elementos da amostra sai e entram outros elementos para a amostra em sua substituição, sendo o número de elementos que entra igual ao que sai. Como consequência o painel resultante já não é retangular (como um painel genuíno), uma vez que se o número de elementos da amostra for N e o número de momentos temporais T , o número de observações agora é inferior a NT , isto é, o painel rotativo apresenta omissões devido ao esquema de rotatividade adotado (Verbeek [160]).

Nesta tese a base de dados longitudinal considerada nos diversos estudos que a compõem tem subjacente um painel rotativo.

1.2 Inquérito às Condições de Vida e ao Rendimento - ICOR

Neste trabalho o principal objeto de estudo são dados em painel rotativo. Os dados em painel surgiram em meados dos anos 60 do século XX nos EUA com o *Panel Study of Income Dynamics* (PSID) criado pelo *Survey Research Center* na Universidade de Michigan, com o objetivo de estudar o bem-estar económico e a pobreza, e também com o *National Longitudinal Survey of Labor Market Experience* (NLS) cujo objetivo consistia no estudo de dados do trabalho e do emprego (Hsiao [69]). Posteriormente apareceram outras bases de dados em painel, sendo de destacar na Europa, pela sua grande importância e aplicação o BHPS -*British Household Panel Survey* que surgiu em 1991 e para o qual existem dados até ao ano de 2007 (altura em que o BHPS passou a integrar o *Understanding Society*), bem como, o *Netherlands Socio-Economic Panel* (SEP), o *German Social Economics Panel* (GSOEP) e o *Luxembourg Social Economic Panel* (PSELL), entre outros. Em 1994, devido à crescente necessidade de assegurar informação sobre rendimento, trabalho e emprego, pobreza, exclusão social, condições de habitabilidade e saúde, em condições de comparabilidade entre os estados membros surgiu o ECHP (*European Community Household Panel*), painel coordenado pelo EUROSTAT com o objetivo de representar a população dos países membros da União Europeia, ao nível individual e das famílias, mas que foi descontinuado em 2001.

Atualmente, o EUROSTAT tem a decorrer vários estudos longitudinais, sendo de destacar o EU-SILC (*Statistics on Income and Living Conditions*), pela sua grande utilização por parte da comunidade científica (embora, na grande maioria dos casos utilizando apenas os dados transversais) em grande parte devido à importância que atualmente as temáticas subjacentes apresentam, e que teve início em 2003. A criação do EU-SILC teve por objetivo a obtenção de indicadores sobre rendimento, condições de vida e exclusão social em condições de comparabilidade entre os vários países da União Europeia. O estudo de indicadores de coesão social assim apurados representa um *input* fundamental ao processo de tomada de decisão nesta área. Por outro lado, o EU-SILC permite uma melhor compreensão de áreas que até agora não foram satisfatoriamente estudadas e a implementação de novas medidas (Eurostat [45]). Este painel, que sucedeu ao ECHP, apresenta a particularidade de ser um painel rotativo, isto é, uma fração da amostra é renovada todos os anos.

O ICOR - Inquérito às Condições de Vida e ao Rendimento para o qual existem, neste momento, dados longitudinais relativos aos anos de 2004 a 2010 e dados transversais para os anos de 2004 a 2011, foi implementado pelo Instituto Nacional de Estatística (INE) com o objetivo de assegurar a participação da população portuguesa na base de dados estatística europeia EU-SILC.

1.3 Modelos com variáveis latentes contínuas para análise de dados longitudinais

Recentemente tem-se assistido a uma crescente utilização de modelos com variáveis latentes para dados longitudinais com o objetivo de descrever, explicar e quantificar a mudança. De acordo com Singer e Willet [145], Collins [25], Bollen [19] e McArdle [88], existe uma grande variedade de modelos disponíveis para analisar dados longitudinais dependendo a escolha das questões de investigação, pois os diferentes modelos fornecem distintas, mas não necessariamente opostas, perspectivas na descrição da mudança. Por outro lado, o aumento da utilização destes modelos resulta da existência de um número crescente de dados longitudinais disponíveis para investigação. Também o grande número de trabalhos mostrando novas formas de aplicação dos modelos para dados longitudinais ajudou a tornar estes métodos mais acessíveis a investigadores que fazem trabalho aplicado, tendo o desenvolvimento de *software* estatístico tido um papel importante nesta questão (Davidov et al. [28], Duncan et al. [34],[36], Ferrer et al [47], Humphreys e Janson [70]).

Dos modelos com variáveis latentes contínuas mais utilizados na modelação deste tipo de dados destacam-se os modelos com trajetória latente. Estes modelos permitem descrever trajetórias individuais, descrever uma trajetória média para o conjunto dos elementos em estudo, bem como a variabilidade em torno dessa trajetória média, possibilitam ainda, a incorporação de variáveis explicativas dessas trajetórias (Bollen e Curran [21], Duncan et al. [34]). A trajetória considerada é latente e não observada, sendo estimada a partir de um estrutura de médias e de variâncias-covariâncias entre as medidas repetidas das variáveis observadas, isto é, não é modelada a evolução das variáveis, mas sim os processos que lhes estão subjacentes.

Assim, dado que, o modelo com trajetória latente assume a trajetória das variáveis observadas ao longo do tempo como latente, é possível parametrizar um modelo com trajetória latente tendo por base o *framework* dos modelos com equações estruturais (Stoel et al. [151])(autores como Raudenbush e Bryck [118] abordam estes modelos no âmbito dos modelos de regressão multinível). Deste modo, porque assentam a sua estrutura e método na lógica dos modelos com equações estruturais, os modelos com trajetória latente partilham algumas das suas vantagens e desvantagens (Bollen e Curran [21], Duncan et al. [34]). Como vantagens, saliente-se a capacidade para testar a adequabilidade global do modelo em estudo, permitindo avaliar a qualidade do ajustamento modelo-dados, a utilização de variáveis latentes medidas por múltiplos indicadores, ou ainda permitir a modelação de fenómenos multidimensionais e complexos. As desvantagens são a necessidade de utilização de amostras de grande dimensão e a impossibilidade de lidar com variáveis categoriais ordinais que alguns pacotes estatísticos utilizados nesta área apresentam.

Além dos modelos com trajetória latente, também outros modelos que integram o *framework* dos modelos com equações estruturais permitem uma análise longitudinal, tais

como, modelos de análise fatorial confirmatória multigrupos ou modelos com equações estruturais com efeitos cruzados e desfasados. Os primeiros permitem o estudo da diferença entre os grupos considerados (estes grupos podem ser distintos momentos temporais), enquanto que com os segundos, o estudo da mudança em determinado momento é sempre função do que acontece no momento imediatamente anterior, para o conjunto dos elementos em estudo. Desta forma, estas abordagens estatísticas não permitem aferir da possível existência de heterogeneidade entre os elementos do conjunto, ao contrário do que acontece com o modelo com trajetória latente. Podendo dizer-se que constituem uma perspectiva mais tradicional dos modelos com equações estruturais na modelação longitudinal de dados. De acordo com McArdle [88], é necessário sublinhar que o estudo da diferença e o estudo da mudança são questões distintas, utilizando dados de diferentes tipos e recorrendo a modelos diferentes.

O crescimento que se tem verificado na utilização dos modelos com trajetória latente, bem como de outros modelos com variável latente contínua para dados longitudinais não é alheio ao aparecimento e desenvolvimento de diferentes pacotes estatísticos, sendo de destacar o *software* Mplus (Muthén e Muthén [105]). Modelações com dados transversais ou longitudinais, de natureza discreta ou contínua, análise multigrupos, diferentes formas de lidar com não respostas e ainda procedimentos de Monte Carlo são algumas das potencialidades deste *software* estatístico cuja utilização por parte da comunidade estatística internacional tem verificado um grande crescimento, tendo-se tornado uma referência incontornável nesta área da estatística, quando comparado com outros pacotes estatísticos, tais como, EQS (Bentler e Wu [14]), LISREL (Jöreskog e Sörbom [77]) e AMOS (Arbuckle [6]).

1.4 Dados omissos

Apesar das inúmeras vantagens que os dados longitudinais apresentam, tais como a possibilidade de analisar alterações ou mudanças de fenómenos sociais, ou outros, a existência de dados omissos constitui uma problemática recorrente em estudos com dados longitudinais em diferentes áreas do conhecimento, assistindo-se cada vez mais ao aparecimento de novas técnicas para lidar com este problema. Todavia, na aplicação destas técnicas é necessário ter em atenção a relação entre as variáveis observadas e a probabilidade de omissão, isto é, o mecanismo de omissão. De acordo com Rubin [131] o mecanismo de omissão pode ser: i) completamente aleatório (MCAR - *missing completely at random*), se não existe relação entre as variáveis observadas e a não resposta; ii) aleatório (MAR - *missing at random*), se a não resposta está relacionada com outra variável presente na análise; ou iii) não aleatório (MNAR - *missing not at random*), se a não resposta se relaciona com o valor que essa mesma variável assumiria. Dependendo ou não, da necessidade de introduzir a relação entre omissão e variáveis observadas na modelação dos dados, as duas primeiras classificações dizem-se ignoráveis e a última não ignorável.

As omissões que podem ser encontradas na literatura são de diferentes tipos e de-

pendem do tipo de dados em estudo. Assim, em dados transversais é usual encontrar as omissões: *item non response*, caso em que o indivíduo não responde a uma pergunta; *unit non response*, caso em que um indivíduo não responde a todas as perguntas. Para dados longitudinais, é usual encontrar omissões devido a atrito ou *dropout*, caso em que os indivíduos abandonam o estudo, ou não respostas intermitentes, caso em que o indivíduo responde em determinado momento, não responde no seguinte, mas volta a responder mais tarde. Num *planned missing design* há uma estrutura de omissão de dados que ocorre de forma intencional e de acordo com o planejado pelo investigador, sendo o objetivo deste planejamento minimizar o esforço de inquirição por parte dos respondentes e consequente abandono. Assim, o painel rotativo ICOR pode ser considerado um exemplo de um *planned missing design* (Enders [41]), mas além das omissões devidas à rotatividade do painel, existem ainda outros tipos de omissões que devem ser tidas em atenção.

De acordo com Molenberghs e Fitzmaurice [95], três métodos são usualmente utilizados para lidar com dados omissos em estudos longitudinais: i) métodos de imputação; ii) métodos baseados na verosimilhança; e iii) métodos com recurso a ponderação. Por outro lado, Schaffer e Graham [143] sublinham a importância da utilização de métodos mais recentes para lidar com dados omissos, tais como, *full information maximum likelihood* e imputação múltipla, apresentando estas múltiplas vantagens sobre os procedimentos clássicos de *listwise* ou substituição pela média, entre outros. Todavia, estas duas abordagens estatísticas, *full information maximum likelihood* e imputação múltipla, conduzem a bons resultados nos casos em que é possível assumir um mecanismo ignorável de omissão dos dados. Abordagens utilizando técnicas como os *pattern mixture models* ou *selection models* são descritas na literatura como mais indicadas sempre que o mecanismo de omissão dos dados é não ignorável.

De referir que a não existência de pesos longitudinais na base de dados ICOR impossibilita a utilização de métodos com recurso a ponderação para lidar com a problemática das não respostas, aquando da modelação estatística dos dados. A incorporação de pesos longitudinais na modelação permitiria a estimação dos modelos tendo em conta as características do desenho amostral complexo, uma vez que os pesos longitudinais são calculados, em cada momento temporal, de forma a refletir a estrutura da população em estudo.

1.5 Privação material

Os estudos publicados sobre pobreza nos últimos anos têm recorrentemente por base a definição de Townsend [158], a qual estabelece que os indivíduos ou agregados se dizem em pobreza quando perdem os recursos para obter os tipos de alimentação, participar em atividades, ter as condições de vida e comodidades que são habituais, ou pelo menos usuais, nas comunidades em que se inserem. Todavia, esta definição de pobreza está longe de ser consensual, tendo presente a complexidade deste conceito, bem como as dificuldades associadas à sua quantificação.

Embora o estudo da pobreza e da qualidade de vida das famílias seja grande parte das vezes feito com base no rendimento disponível, esta medida pode, por si só, não ser satisfatória e não refletir de forma assertiva a realidade para alguns grupos da população, tais como trabalhadores por conta própria ou pessoas cujos rendimentos são gerados pela chamada economia paralela. Por outro lado, alguns agregados podem ter disponibilidade de outros recursos, como por exemplo, poupanças ou ajudas de familiares e amigos.

Assim, o nível e a qualidade de vida das famílias podem ser medidos através de outros indicadores, tais como a capacidade para aceder a um conjunto de necessidades básicas, a posse de bens duradouros, as condições de habitabilidade e mesmo ainda as condições ambientais do local onde residem, ou seja, através de um conceito de privação material (Guio [60], Whelan et al. [166], Whelan e Maître [167], Dekkers [29], Halleröd et al. [62]).

A privação material corresponde à ausência forçada de uma combinação de itens que contribuem para condições materiais de vida ou de qualidade de vida, tais como condições de habitação, posse de bens, capacidade para suportar necessidades básicas na sociedade a que os indivíduos pertencem, ou mesmo condições ambientais. Este conceito de ausência forçada foi introduzido por Mack e Lansley [86], contrariamente a Townsend [158] que simplesmente considerou a falta de um item como implicando privação.

Autores como Whelan e Maître [167], Guio [60] e Halleröd et al. [62], entre outros, consideram várias dimensões na definição do conceito de privação material, embora não exista entre eles um consenso quanto ao número de dimensões a considerar, nem quanto ao número de itens que compõem cada uma das dimensões.

Nos estudos 1 a 3 que compõem esta tese nem sempre foi considerado o mesmo número de dimensões do conceito latente de privação material. No estudo 1 foram consideradas as dimensões: constrangimentos económicos; posse de bens; condições da habitação; e vizinhança, tendo-se esta última mostrado não relevante. Assim, no estudo 3 apenas foram consideradas as três primeiras dimensões. No estudo 2 foram consideradas as dimensões: capacidade financeira (constrangimentos económicos) e posse de bens, as dimensões com maior peso na definição do conceito de privação material. Nos estudos 4 a 7, com base nas análises que foram sendo conduzidas, foi construído um indicador de privação material consistindo numa soma dos itens que constituíam as dimensões que mostraram ter maior peso na definição do conceito de privação material (ver secção 1.7).

1.6 Objetivo

O objetivo deste trabalho é auxiliar o investigador na modelação longitudinal de dados que resultam de um painel rotativo recorrendo a modelos com trajetória latente. O painel rotativo ICOR contém omissões que são planeadas pelo investigador (omissões devido ao esquema de rotatividade implementado no painel), podendo para estas ser assumido

um mecanismo MAR. Por outro lado, no painel ICOR existem ainda outras omissões, para as quais não pode ser descartado um mecanismo MNAR. As abordagens estatísticas consideradas para lidar com as omissões existentes neste painel foram: *full information maximum likelihood* e imputação múltipla, se é assumido um mecanismo de omissão de dados MAR; *pattern mixture model* e *selection model*, se é assumido um mecanismo de omissão MNAR. A utilização de um *pattern mixture model* foi feita recorrendo a duas diferentes parametrizações, a multigrupos e a convencional. Neste painel, tal como acontece com a maioria dos dados em painel (painel genuíno), as omissões podem ser resultado de elementos que abandonam o estudo ou podem ser não respostas intermitentes, mas ao contrário dos painéis genuínos, também existem entradas tardias, um tipo de omissão que não se encontra descrito na literatura. Assim pretendeu-se estimar um modelo com trajetória latente utilizando os dados do painel rotativo ICOR, tendo em atenção este tipo de omissão. Com este propósito um novo indicador, um indicador de padrão de omissão, foi proposto recorrendo à parametrização convencional do *pattern mixture model*, em alternativa aos indicadores de omissão usualmente descritos na literatura. De referir ainda que, embora existam vários exemplos da utilização dos *pattern mixture model* e *selection model* na literatura, muito poucos são na área das ciências sociais.

Por outro lado, os procedimentos de Monte Carlo foram utilizados na implementação de um estudo de simulação que teve como objetivo averiguar o efeito da variação de diferentes fatores nas estimativas dos parâmetros de um modelo com trajetória latente. Os fatores investigados foram desvios face à normalidade da distribuição dos dados, a dimensão da amostra, o método de estimação utilizado, o número de momentos temporais do modelo, diferentes modelos e ainda diferentes percentagens de omissões em diferente número de momentos temporais, resultantes de um *planned missing design*, assim como distintas combinações destes fatores. Com a realização deste estudo de simulação procurou conhecer-se a razão (ou razões) para a existência de resultados contraditórios, na modelação longitudinal da privação material com as observações válidas, e com dados em que existem diferentes tipos de omissões.

Num dos estudos que compõem esta tese a reamostragem pelo método não paramétrico de *bootstrap* foi também considerada, no sentido de esclarecer dúvidas suscitadas pelas conclusões contraditórias obtidas, e já referidas.

Foram também considerados para modelação longitudinal da privação material, outros modelos, que tal como o modelo com trajetória latente partilham o *framework* dos modelos com equações estruturais, tais como, o modelo de análise fatorial confirmatória multigrupos e modelo com equações estruturais com efeitos cruzados e desfasados. As vantagens do modelo com trajetória latente a descrever, explicar e quantificar a mudança são apresentadas.

1.7 Apresentação dos estudos

Esta tese compila resultados e conclusões de vários estudos, que foram sendo apresentados em congressos estatísticos, visando a modelação de dados longitudinais, ao longo do tempo. No entanto a realização destes estudos foi condicionada pelo tipo de dados que foram sendo disponibilizados, transversais ou longitudinais, bem como, pelo momento em que estes foram disponibilizados pelo INE. Nomeadamente, o primeiro estudo, no qual foi analisada a diferença na privação material dos agregados nos anos de 2007 e 2008, utilizando-se as bases de dados transversais do ICOR, as únicas disponíveis na altura.

Ao longo do tempo em que esta tese foi realizada, o INE foi disponibilizando de forma faseada os dados do inquérito ICOR. Consequentemente os dados subjacentes aos diferentes estudos realizados não são sempre os mesmos, mas sim os mais recentes no momento da realização desse estudo. A exceção é o estudo três que foi repetido, já quando se escrevia o texto final desta tese, por esta razão foram utilizados os dados de 2007 a 2010, ao invés dos dados de 2006 a 2008 que tinham sido utilizados na comunicação oral da JOCLAD 2011, uma vez que é desejável a utilização de um maior número de momentos temporais na modelação da trajetória latente da privação material. No último estudo apresentado, o estudo de simulação, os dados foram gerados com uma estrutura específica, por forma a responder às diferentes questões de investigação.

À exceção do primeiro estudo, todos os outros foram realizados utilizando o pacote estatístico Mplus que é uma referência, no panorama internacional, na análise longitudinal recorrendo a modelos com variáveis latentes contínuas, apresentando inúmeras potencialidades, tais como, modelar dados transversais ou longitudinais, variáveis discretas ou contínuas, utilização de diferentes abordagens estatísticas para lidar com as não respostas, permitir análise multigrupos e procedimentos de Monte Carlo. Assim, com o propósito de modelar longitudinalmente o conceito de privação material, com dados completos ou com dados apresentando omissões, várias rotinas foram programadas recorrendo a este *software*. Em alguns estudos o conceito de privação material foi considerado como variável latente, sendo medido por distintos conjuntos de itens nas suas diferentes dimensões. Por outro lado, noutros estudos foi calculado um indicador de privação material como uma soma de itens, uma vez que a maioria das variáveis existentes no ICOR são do tipo sim/não. De salientar que, este é um procedimento usual em estudos sobre privação material, ver por exemplo Guio [60] ou Whelan e Maître [167].

Na tabela 1.1 são apresentados os vários estudos realizados e a metodologia utilizada em cada um deles, bem como as variáveis consideradas e o tipo de dados subjacentes.

Os estudos que constituem esta tese podem ser considerados como pertencentes a quatro diferentes grupos, de acordo com os dados utilizados nas modelações. Um primeiro grupo em que apenas foram consideradas as observações válidas, um segundo grupo em que se consideraram além das observações válidas, as omissões que resultam da rotatividade do painel, consequentemente os dados apresentam um mecanismo de omissão

	Metodologias	Variáveis	Tipo de dados
Estudo 1	Análise Fatorial Confirmatória Análise Fatorial Confirmatória Multigrupos	Var.latente: Privação Material Dimensões: Constrang. económicos; Posse de bens; Condições da habitação; Vizinhaça	Completos
Estudo 2	Modelo com equações estruturais com efeitos cruzados e desfasados Modelo com trajetória latente bivariada	Capacidade Financeira; Posse de bens	Completos
Estudo 3	Modelo com trajetória latente de segunda ordem Modelo com trajetória latente de segunda ordem condicionada Análise Fatorial Confirmatória	Var.latente: Privação Material Dimensões: Constrang. económicos; Posse de bens; Condições da habitação	Completos
Estudo 4	Modelo com trajetória latente <i>Full Information Maximum Likelihood</i> Simulações de Monte Carlo	Indicador de Privação Material	Omissões ignoráveis
Estudo 5	Modelo com trajetória latente <i>Full Information Maximum Likelihood</i> Imputação Múltipla Simulações de Monte Carlo	Indicador de Privação Material	Omissões ignoráveis
Estudo 6	Modelo com trajetória latente <i>Full Information Maximum Likelihood</i> <i>Pattern Mixture Model</i> Técnica de <i>Bootstrap</i>	Indicador de Privação Material	Omissões ignoráveis e não ignoráveis
Estudo 7	Modelo com trajetória latente <i>Full Information Maximum Likelihood</i> <i>Pattern Mixture Model</i> <i>Selection Model</i>	Indicador de Privação Material	Omissões ignoráveis e não ignoráveis
Estudo de Simulação	Modelo com trajetória latente Modelo de Mistura <i>Full Information Maximum Likelihood</i> Simulações de Monte Carlo		Completos e com omissões ignoráveis

Tabela 1.1: Abordagens estatísticas, variáveis e o tipo de dados utilizados em cada um dos estudos que compõem esta tese

MAR. Num terceiro grupo foram consideradas todas as observações da base de dados, consequentemente todas as omissões aí existentes, as resultantes da rotatividade e outras, sendo o mecanismo de omissão de dados MNAR. Por fim, é apresentado um estudo de simulação, no qual são gerados dados com e sem omissões. Cada um destes quatro grupos de estudos é de seguida apresentado com maior detalhe.

1.7.1 Estudos com dados completos

Primeiro estudo

Num primeiro estudo¹ foram utilizados os dados transversais do ICOR dos anos 2007 e 2008 para analisar o conceito de privação material em quatro dimensões: i) capacidade financeira do agregado para satisfazer necessidades básicas; ii) posse de bens duradouros; iii) condições da habitação; e iv) vizinhança, nos dois diferentes momentos temporais.

¹Estudo adaptado do trabalho apresentado em comunicação oral na JOCLAD 2010

Neste sentido, para cada um dos momentos temporais, são calculados indicadores correspondentes às quatro dimensões da privação material. O indicador dos constrangimentos económicos corresponde a uma soma de itens respeitantes à capacidade financeira para suportar os requisitos mais básicos do agregado. O indicador da posse de bens duradouros é construído como uma soma de itens relativos à disponibilidade económica para possuir determinados bens, tais como um aparelho de TV a cores. O indicador das condições da habitação é calculado considerando os itens que caracterizam a deterioração ou falta de equipamento básico do alojamento. Finalmente, o indicador da vizinhança considera itens como barulho, crime e poluição na área envolvente à que o agregado habita. Um modelo de análise fatorial confirmatória foi considerado para cada um dos diferentes anos, 2007 e 2008, com o objetivo de validar a estrutura fatorial proposta, bem como avaliar o peso das diferentes dimensões da privação material na definição deste conceito, não se tendo mostrado relevante a dimensão vizinhança. Com o intuito de explicar o conceito de privação material nas suas diferentes dimensões, é proposto um modelo com equações estruturais considerando as três seguintes variáveis caracterizadoras dos agregados: a área de urbanização do local onde o agregado reside, a dimensão do agregado e a existência de crianças no agregado.

Posteriormente, e tendo por objetivo uma análise longitudinal dos dados, é utilizada uma análise multigrupos para aferir da existência de comportamentos diferentes por parte dos agregados nos dois momentos temporais considerados, 2007 e 2008, sendo que os dados utilizados são transversais. A modelação estatística é feita recorrendo ao *software* LISREL e os resultados apresentados na secção 3.2.2.

Segundo estudo

Numa segunda abordagem proposta²(Vicente e Salgueiro [162]) utilizam-se dados longitudinais do ICOR dos anos 2004 a 2007 para estudar o conceito de privação material em duas dimensões: i) capacidade financeira do agregado para satisfazer necessidades básicas; e ii) posse de bens duradouros. Para cada uma destas dimensões, os indicadores utilizados são calculados como somas de itens, seguindo a prática comum em trabalhos sobre esta temática.

Neste estudo o objetivo é modelar longitudinalmente a evolução da privação material de 2004 a 2007, comparando os resultados obtidos através de duas técnicas de modelação diferentes: modelos com equações estruturais com efeitos cruzados e desfasados e modelos com trajetória latente. No capítulo 3 este estudo é apresentado dividido em duas partes, nas secções 3.3.2 e 3.4.2, de acordo com a metodologia utilizada. A modelação estatística é feita recorrendo ao *software* estatístico Mplus.

²Estudo apresentado em comunicação oral na SPE 2010

Terceiro estudo

Neste trabalho³ utilizam-se os dados do ICOR correspondentes a quatro momentos temporais, de 2007 a 2010, com o objetivo de modelar longitudinalmente a privação material, considerando este conceito como medido em três dimensões: i) constrangimentos económicos; ii) posse de bens duradouros; e iii) condições da habitação. Com este propósito são calculados *scores* para cada uma das três dimensões da privação (obtidos como somas de itens), utilizados como indicadores ordinais de medida do conceito de privação material em cada um dos momentos temporais. Esta modelação da privação material é feita recorrendo a um modelo com trajetória latente de segunda ordem, ou modelo com trajetória latente com múltiplos indicadores, que permite descrever a trajetória de uma variável latente medida por um conjunto de variáveis observadas. Uma vez descrita a trajetória da privação material no período em análise, são consideradas como variáveis explicativas desta trajetória, o rendimento, a dimensão e a área de urbanização do local de residência do agregado. É discutida a inclusão destas variáveis no modelo condicional quando consideradas como variantes e como invariantes no tempo.

O modelo com trajetória latente de segunda ordem, não condicionada e condicionada, medida por indicadores ordinais, proposto neste trabalho é estimado com recurso ao pacote estatístico Mplus e os resultados são apresentados na secção 3.4.4.

1.7.2 Estudos com dados com omissões MAR

Quarto estudo

Neste trabalho⁴ é proposta a modelação através de um modelo com trajetória latente de um indicador de privação material para os anos de 2006 a 2009. Ao longo dos quatro momentos temporais em análise permaneceram no painel um conjunto de agregados, para os quais existe informação completa. Existe ainda um outro conjunto de agregados que participaram no estudo entre 2006 e 2008 e que não estiveram em 2009 devido ao esquema de rotatividade do painel.

A utilização de procedimentos de *full information maximum likelihood* disponíveis no programa estatístico Mplus permite estimar um modelo com trajetória latente para modelar a trajetória do *score* de privação material entre 2006 e 2009, considerando a totalidade dos agregados, ao invés do conjunto inicialmente considerado através da aplicação de uma técnica *listwise*, uma vez que no ano de 2009 existem omissões que resultam do desenho do estudo. Diferentes conjuntos de estimativas são então obtidas para os parâmetros do modelo, sendo discutido qual das duas opções metodológicas deve o analista tomar.

Recorrendo a um estudo de simulação de Monte Carlo, e assumindo como verdadeiros os valores dos parâmetros estimados pelo modelo com trajetória latente para os dados

³Estudo adaptado do trabalho apresentado em comunicação oral na JOCLAD 2011

⁴Estudo apresentado em comunicação oral na JOCLAD 2012

completos, são geradas amostras com um padrão de omissão de 50% no último momento temporal, ano de 2009. Para cada uma das diferentes amostras são estimados utilizando o método do *full information maximum likelihood* os parâmetros do modelo com trajetória latente, sendo o valor médio dessas estimativas comparado com o “verdadeiro” valor usado para gerar as réplicas. Com o objetivo de avaliar o enviesamento das estimativas é calculada a proporção de réplicas para as quais um intervalo de confiança a 95% contém o verdadeiro parâmetro. O *software* estatístico utilizado é o Mplus e os resultados são apresentados na secção 4.3.2.

Quinto estudo

Neste trabalho⁵ o objetivo é comparar os resultados obtidos na estimação dos parâmetros de um modelo com trajetória latente (com quatro momentos temporais) para um *score* de privação material, na presença de dados omissos que configuram um *planned missing design*, isto é, os dados do painel rotativo ICOR. Distintas subamostras que apresentam diferentes percentagens de omissões são consideradas. São usadas duas abordagens estatísticas (*full information maximum likelihood* e imputação múltipla) para lidar com as omissões. As estimativas dos parâmetros obtidas por aplicação destas duas diferentes abordagens estatísticas face à existência de dados omissos, são assim comparadas entre si, bem como com aquelas que resultam da estimação do modelo com dados completos. Um estudo de simulação utilizando procedimentos de Monte Carlo disponíveis no pacote estatístico Mplus permite aferir a qualidade da estimação dos parâmetros de um modelo com trajetória latente com dados omissos. Os resultados são apresentados na secção 4.3.4.

1.7.3 Estudos com dados com omissões MNAR

Sexto estudo

Num *planned missing design* a maioria das omissões ocorre de acordo com o planeado pelo investigador, sendo o objetivo da utilização de um desenho deste tipo aumentar a qualidade dos dados recolhidos (Enders [41]). O painel rotativo ICOR pode ser considerado um exemplo de um *planned missing design*. Todavia, neste painel ocorrem ainda outros tipos de omissões, tais como saídas antes do tempo, entradas tardias ou agregados que não respondem em determinado momento temporal e voltam a responder no seguinte, resultando assim em diferentes padrões de omissão.

O objetivo neste estudo⁶ é lidar com os diferentes padrões de omissão que podem ocorrer num painel rotativo como o ICOR, utilizando duas diferentes abordagens estatísticas para efeitos de estimação do modelo: *full information maximum likelihood* e *pattern mixture model*, tendo em consideração o mecanismo de omissão dos dados, ignorável ou

⁵Estudo apresentado em comunicação oral no European Congress of Methodology 2012

⁶Estudo apresentado em comunicação oral na SPE, 2012

não ignorável.

Assim, este estudo propõe a modelação através de um modelo com trajetória latente de um *score* de privação material para os anos de 2006 a 2009. Do total dos agregados que estão na base de dados entre 2006 a 2009, apenas se mantiveram no painel durante os quatro anos uma pequena fração do total dos agregados, verificando-se a existência de diferentes padrões de omissão.

A utilização de procedimentos de *full information maximum likelihood* permite modelar a trajetória latente do score de privação material entre 2006 e 2009, considerando a totalidade dos agregados. O recurso a uma abordagem *bootstrap* permite obter intervalos de confiança não paramétricos para os parâmetros estimados, por forma a aferir a qualidade da estimação.

Uma abordagem *pattern mixture model* é considerada com o objetivo de incorporar na modelação informação sobre a distribuição das omissões. Para tal, os diferentes padrões de omissões observados na amostra em estudo são divididos em dois grupos, os que resultam da rotatividade do painel e os outros, uma vez que se acredita que estes dois grupos têm mecanismos de omissão distintos. O pacote estatístico utilizado é o Mplus e os resultados são apresentados na secção 4.4.3. juntamente com os resultados do estudo seguinte, uma vez que estes se complementam.

Sétimo estudo

Neste estudo⁷ é modelada a trajetória de um indicador de privação material, no período de 2006 a 2009, recorrendo a modelos com trajetória latente. Os dados utilizados são a base de dados longitudinal 2006 a 2009 do ICOR. Assim, nesta base de dados existem omissões devido ao desenho do estudo, existindo no entanto, outro tipo de não respostas.

O objetivo deste estudo consiste em lidar com a existência dos vários tipos de omissões presentes nos dados recorrendo a diferentes abordagens estatísticas. Uma abordagem com *full information maximum likelihood* quando se assume um mecanismo de omissão de dados ignorável, caso contrário, é discutida a utilização do *pattern mixture model* e do *selection model*. Ambas as abordagens, *pattern mixture model* e *selection model* permitem incorporar na modelação informação sobre a distribuição das omissões. No *pattern mixture model* duas parametrizações distintas são utilizadas e comparadas. Uma das parametrizações é uma abordagem multigrupos, tal como no estudo seis, a outra parametrização consiste numa abordagem convencional no âmbito dos *pattern mixture models*, na qual um indicador de padrão de omissão é proposto como alternativa ao indicador de omissão que é usualmente descrito na literatura. Padrões de omissão distintos são tidos em consideração na modelação de uma trajetória de mudança do indicador de

⁷Estudo apresentado em comunicação oral no EU-SILC Users Conference 2013

privação material e a estabilidade dos resultados obtidos nas diferentes abordagens é comparada. A modelação estatística é realizada recorrendo ao *software* estatístico Mplus e os resultados são apresentados na secção 4.4.3, uma vez que este estudo se apresenta como uma sequência lógica do estudo 6.

1.7.4 Estudo de simulação

Neste estudo de simulação⁸ são gerados dados com uma determinada estrutura com o propósito de responder a algumas questões que foram sendo suscitadas pelos resultados obtidos nos estudos anteriores. A influência de diferentes fatores nas estimativas de um modelo com trajetória latente, tais como, desvios à normalidade da distribuição dos dados utilizados, o efeito da dimensão da amostra, o efeito da existência de diferentes percentagens de omissões nos dados, o efeito do número de momentos temporais ou ainda o estimador utilizado na estimação do modelo são avaliados.

O estudo de simulação foi realizado em Mplus pela razão de que este *software* permite, de forma integrada, gerar os dados e estimar os modelos com a estrutura definida pelo utilizador. O procedimento do Mplus para gerar dados encontra-se no âmbito dos modelos de mistura. Para gerar dados com distribuição normal foi considerado um modelo de mistura com apenas uma classe, na qual foi definido um modelo com trajetória latente. Para gerar dados com distribuição resultante de uma mistura de normais foi utilizado um modelo de mistura com duas classes, com diferentes percentagens de observações, sendo que em cada uma destas classes foi definido um modelo com trajetória latente diferente, isto é, modelos cujos parâmetros assumiram valores distintos. É assim possível em Mplus gerar dados com distintos graus de assimetria e achatamento. Foram gerados dados com distribuição, normal e não normal, com diferentes percentagens de não respostas em diferente número de momentos temporais, por forma a obter nas amostras padrões de omissão semelhantes aos que ocorrem nos *planned missing design* (base de dados ICOR). As amostras de dados foram geradas com dimensão a variar entre 50 e 1000 observações. Para cada uma das amostras de dados gerados foi estimado um modelo com trajetória latente (tendo sido utilizados diferentes métodos de estimação) e os resultados obtidos analisados. Os resultados deste estudo são apresentados detalhadamente no capítulo 5.

1.8 Estrutura da tese

No capítulo 1 são apresentados a motivação para a realização desta tese e o objetivo da sua realização. São ainda apresentados os diversos estudos que compõem a tese e o tipo de dados subjacentes a estes estudos. Também neste capítulo, é realizada uma introdução aos diferentes modelos com variáveis latentes contínuas para dados longitudinais utilizados nesta tese, assim como ao conceito de privação material, temática utilizada nos

⁸Parte deste estudo foi apresentado em duas comunicações orais na JOCLAD 2014

diferentes estudos.

No capítulo 2 é apresentado em detalhe o inquérito ICOR, incluindo o respetivo desenho, bem como as variáveis que o constituem e as que foram sendo utilizadas nos diferentes estudos. Neste capítulo são sintetizados os resultados da revisão da literatura sobre o conceito de privação material e sobre as diversas formas em que este tem sido estudado.

No capítulo 3 são apresentados os estudos que fizeram uso essencialmente dos modelos com trajetória latente, assim como, outras metodologias no *framework* dos modelos com equações estruturais. As vantagens e desvantagens destas diferentes abordagens estatísticas são apresentadas e discutidas.

No capítulo 4 são apresentados os estudos que procuraram lidar com as omissões resultantes da rotatividade do painel, assim como com outras não respostas que ocorrem neste painel. As diferentes abordagens estatísticas consideradas para lidar com as omissões existentes, são apresentadas e discutidas.

No capítulo 5 é apresentado o estudo de simulação que procurou responder às questões que se afiguraram pertinentes face aos resultados contraditórios obtidos em alguns dos estudos do capítulo 4, tais como, o efeito nas estimativas dos parâmetros de um modelo com trajetória latente de: diferentes níveis de desvio face à distribuição normal dos dados; existência de omissões; diferentes dimensões amostrais; e diferentes métodos de estimação.

As conclusões de todo o trabalho realizado são sintetizadas no capítulo 6, as limitações e perspectivas para trabalho futuro são ainda discutidas.

Capítulo 2

O Painel ICOR e a Medição da Privação Material

2.1 O Painel ICOR

O Inquérito às Condições de Vida e Rendimento das Famílias (ICOR) é um painel com uma periodicidade anual, tendo sido implementado com o propósito de assegurar a participação portuguesa na base de dados estatística europeia denominada EU-SILC (*European Statistics on Income and Living Conditions*), cujos objetivos são a produção de estatísticas sobre a distribuição do rendimento, as condições de vida e exclusão social, em condições de comparabilidade entre os vários países da União Europeia e outros estados vizinhos que participam no inquérito.

A participação portuguesa no EU-SILC é assegurada pelo Instituto Nacional de Estatística (INE) desde 2004. Neste sentido, o INE tem atualmente em curso o painel ICOR para o qual existem (disponíveis aos utilizadores mediante a assinatura de um protocolo de colaboração, tendo em vista a sua utilização em projetos de investigação) as bases de dados transversais relativas aos anos de 2004 a 2011 e as bases de dados longitudinais relativas aos seguintes períodos: 2004 a 2007; 2005 a 2008; 2006 a 2009; e 2007 a 2010. Por razões de confidencialidade que se encontram regulamentadas, os dados são disponibilizados por forma a não possibilitar ao utilizador a junção dos dados transversais, para obtenção dos dados longitudinais.

Este painel apresenta a particularidade de ser um painel rotativo, com uma dinâmica de rotatividade de $\frac{1}{4}$ da amostra em cada ano, isto é, 25% da amostra é renovada todos os anos. Assim, nenhum agregado familiar permanecerá na amostra por mais do que quatro anos consecutivos, procurando-se deste modo evitar o esforço de inquirição sobre os respondentes do agregado e o conseqüente abandono antes do fim dos quatro anos (INE [72]).

A amostra total está então dividida em quatro subamostras, sendo que anualmente cada subamostra que já foi observada quatro vezes é substituída por uma nova (com exceção das três primeiras subamostras do primeiro ano, 2004), conforme representado na

figura 2.1. Os alojamentos que saem da amostra são substituídos por outros da mesma área geográfica, sendo que as áreas permanecem fixas ao longo do tempo. Este esquema permite estudos longitudinais e transversais.

De salientar que, devido ao esquema de rotatividade implementado no ICOR, para análise longitudinal com dados completos em quatro momentos temporais existem apenas 25% das observações, em três momentos temporais existem 50% das observações, e 75% das observações considerando somente dois momentos temporais. A dimensão total da amostra em cada ano situa-se entre os 4 e os 5 mil agregados.

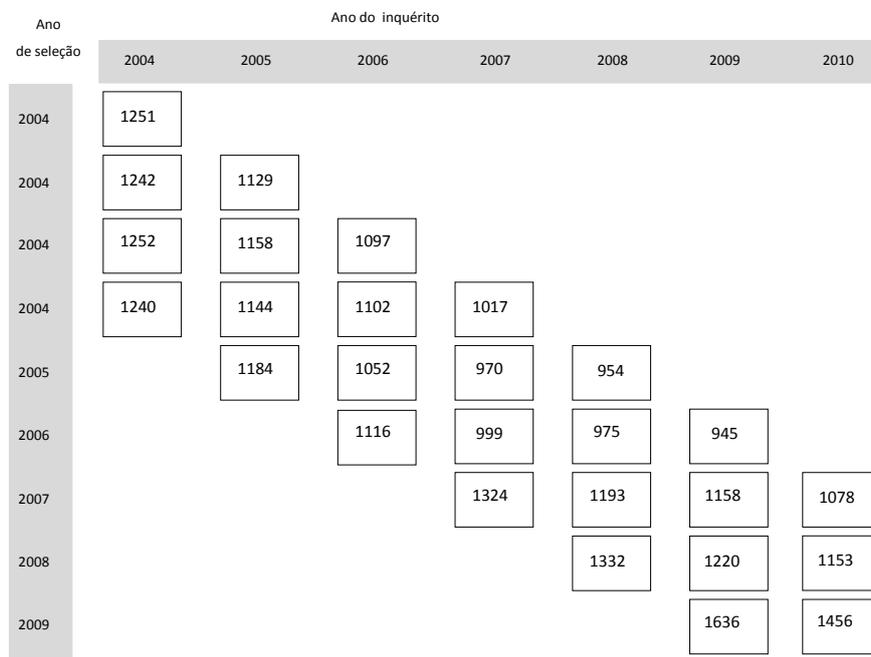


Figura 2.1: ICOR - Desenho do painel (número de agregados familiares)

A amostra do ICOR é selecionada a partir de uma base de amostragem designada por “amostra-mãe” e constituída por alojamentos familiares que foram obtidos a partir dos dados do recenseamento da população e habitação de 2001. O universo de referência é o conjunto de todos os indivíduos residentes no território nacional que, no período em questão, viviam em alojamentos familiares. O esquema de amostragem utilizado é estratificado multietapas (áreas e depois alojamentos). No primeiro momento temporal, a amostra foi selecionada a partir da “amostra-mãe” introduzindo uma nova etapa no processo de seleção escolhendo aleatoriamente 542 áreas de entre as 1408 que constituem esta base. Na etapa seguinte, selecionaram-se em cada uma das áreas da etapa anterior 12 alojamentos familiares de residência principal, para os quais se recolhe informação sobre todos os agregados e indivíduos que aí tenham a sua residência principal (INE [72]).

Cada base de dados anual do ICOR é constituída por quatro ficheiros:

- D - Ficheiro com informação básica dos agregados (amostragem, resultado do contacto, outra informação);
- R - Ficheiro com informação básica dos indivíduos (informação demográfica, relações de parentesco, elegibilidade, resultado da entrevista, cuidados com a criança);
- H - Ficheiro com informação completa dos agregados (informação da entrevista ao agregado propriamente dita, com dados do alojamento, rendimentos, exclusão social e geral);
- P - Ficheiro com informação completa dos indivíduos (informação da entrevista ao indivíduo propriamente dita, com dados demográficos, rendimentos, condição perante o trabalho, histórico profissional, saúde, educação).

Em cada um destes quatro ficheiros existem vários blocos de informação, sendo o ficheiro D (registo dos agregados) constituído apenas por um único bloco B de informação básica. No ficheiro R (registo de todos os indivíduos) constam dois blocos, B - informação básica e L - trabalho. No ficheiro H (dados dos agregados) constam quatro blocos, B - informação básica, H - alojamento, S - exclusão social e Y - rendimento. No ficheiro P (dados dos indivíduos com mais de 16 anos) constam dois blocos, B - informação básica, E - educação, H - saúde, L - trabalho, Y - rendimento. Os ficheiros D e R destinam-se sobretudo a análises de carácter metodológico, como, por exemplo, identificação de agregados/indivíduos que estão em condições de elegibilidade para a vaga seguinte. Os ficheiros H e P destinam-se a análises temáticas sobre a privação material, pobreza e desigualdade, e constituem a verdadeira essência dos dados.

As variáveis utilizadas nos diferentes estudos realizados ao longo do presente trabalho foram:

- HS040 - Capacidade do agregado para pagar uma semana de férias, por ano, fora de casa;
- HS050 - Capacidade do agregado para ter uma refeição de carne ou de peixe (ou equivalente vegetariano), pelo menos de dois em dois dias;
- HS060 - Capacidade do agregado para suportar despesas inesperadas;
- HS070 - Disponibilidade de telefone fixo ou móvel;
- HS080 - Disponibilidade de TV a cores;
- HS100 - Disponibilidade de máquina de lavar roupa;
- HS110 - Disponibilidade de veículo ligeiro de passageiros ou misto;
- HS120 - Capacidade do agregado para fazer face às despesas e encargos usuais;
- HS160 - Luz insuficiente no alojamento;

- HS170 - Vizinhos barulhentos ou ruídos exteriores (comércio, trânsito, indústrias);
- HS180 - Poluição, sujidade ou outros problemas ambientais causados pelo trânsito ou indústria;
- HS190 - Crime, violência ou vandalismo na área de residência;
- HH040 - Telhado que deixa passar água, paredes/fundações/chão húmido, ou persianas de janelas ou o chão apodrecido;
- HH050 - Capacidade financeira para ter aquecimento adequado em casa;
- HH080 - Instalações de banho ou duche no interior (ou HH081 a partir de 2009);
- HH090 - Retrete no interior (ou HH091 a partir de 2009);
- DB100 - Grau de urbanização;
- HX040 - Dimensão do agregado em número de indivíduos;
- HX060 - Composição familiar do agregado.

As variáveis HX são variáveis derivadas e fazem parte do ficheiro H (dados dos agregados). As variáveis HS040, HS050, HS060, HS160, HS170, HS180, HS190, HH040, HH050, HH080, HH090 são nominais, sendo a resposta possível sim ou não. A resposta para a variável HS120 é considerada numa escala de 1, com grande dificuldade, a 6, com muita facilidade. Para as variáveis HS070, HS080, HS100 e HS110 a resposta possível corresponde a uma das três categorias: 1) sim; 2) não, não tem possibilidades económicas; 3) não, outra razão. A variável grau de urbanização DB100 pode tomar três valores, 1, 2 e 3, que correspondendo respetivamente a área densamente povoada, área intermédia e área pouco povoada. A variável HX040 toma valores inteiros de 1 ao número de indivíduos que constituem o agregado. A composição familiar do agregado considera as seguintes situações:

- um adulto;
- dois adultos sem crianças dependentes, ambos com menos de 65 anos;
- dois adultos sem crianças dependentes, pelo menos um com 65 ou mais anos;
- outros agregados sem crianças dependentes;
- um adulto com uma ou mais crianças dependentes;
- dois adultos com uma criança dependente;
- dois adultos com duas crianças dependentes;
- dois adultos com três ou mais crianças dependentes;

- outros agregados com uma ou mais crianças dependentes.

Esta última variável foi utilizada para construir a variável nominal existência de crianças no agregado, que assume os valores 0 (agregado sem crianças, isto é, as primeiras quatro categorias) ou 1 (agregado com crianças, isto é, as 5 últimas categorias).

2.2 A Medição da Privação Material

As dimensões da privação material, tal como o número de itens em cada dimensão, têm sido objeto de vários estudos. Estes trabalhos nem sempre se têm mostrado consensuais quanto ao número de dimensões a considerar na definição do conceito de privação material, sendo ainda naturalmente condicionados pelas perguntas disponíveis nas bases de dados utilizadas.

Betti e Verma [16] consideraram cinco dimensões da privação material: 1) privação básica do estilo de vida (ausência de capacidade financeira para assegurar bens essenciais, tais como, manter a casa aquecida); 2) privação secundária do estilo de vida (por exemplo, possuir TV); 3) condições da habitação (por exemplo, ter casa de banho); 4) estado da habitação (por exemplo, telhado que deixa passar água); 5) problemas ambientais (por exemplo, ter vizinhança ruidosa).

Por outro lado, Eroglu [44] considerou a privação material medida em três dimensões: 1) monetária (rendimento, poupanças, etc); 2) consumo (alimentação, educação, despesas correntes com a habitação, etc); e 3) trabalho (nº de horas de trabalho, seguro, etc.).

Em 2007, Whelan e Maître [167] consideraram cinco dimensões de privação material: 1) constrangimentos económicos (capacidade financeira para suportar despesas com alimentação, etc); 2) consumo (posse de bens, como por exemplo, TV); 3) condições da habitação (por exemplo, ter casa de banho); 4) vizinhança (ruído, poluição, etc); e 5) estado da saúde (ter doença crónica, ser portador de deficiência, etc).

Nolan e Whelan [110] salientam o facto de o número de questões disponíveis na base de dados condicionar o número de dimensões a considerar no estudo do conceito de privação material. Deste modo, no seu trabalho com a base de dados ECHP consideram cinco dimensões na definição da privação material, as mesmas que foram consideradas por Betti e Verma [16]. Todavia, se a base de dados subjacente ao estudo é o EU-SILC, consideram a privação material medida em apenas três dimensões: 1) privação no consumo (em que os itens considerados são, por exemplo, alimentação, casa, etc); 2) bens do agregado (possuir carro, casa de banho, etc); e 3) vizinhança (poluição, crime, etc); isto devido à limitação das questões disponíveis.

Guio [61] no seu trabalho de desenvolvimento de um indicador de privação material com base nos dados do EU-SILC considerou três dimensões da privação material: 1)

constrangimentos económicos; 2) posse de bens duradouros; e 3) condições da habitação, sendo que o recurso à análise fatorial confirmatória lhe permitiu concluir pela importância das duas primeiras dimensões na definição do conceito em estudo. Assim, o indicador de privação material utilizado pelo EUROSTAT consiste na soma dos itens considerados nestas duas dimensões.

Nos vários estudos realizados ao longo deste trabalho, a privação material nem sempre foi estudada da mesma forma, isto porque se foi concluindo pela não relevância de algumas das dimensões, propostas por autores como Whelan e Maître [167] entre outros, na definição deste conceito. Posteriormente, os itens considerados nas dimensões que mostraram ter maior peso na definição deste conceito foram utilizados na construção de um indicador de privação material. Assim, no primeiro estudo este conceito foi considerado medido em quatro diferentes dimensões, constrangimentos económicos (ou ausência de capacidade financeira), posse de bens duradouros, condições da habitação e vizinhança, sendo para cada uma delas calculada uma soma de itens:

1. Constrangimentos económicos (ausência de capacidade financeira para)
 - ter uma refeição de carne ou de peixe, pelo menos de dois em dois dias;
 - pagar uma semana de férias, por ano, fora de casa, a todo o agregado;
 - suportar despesas inesperadas, sem recorrer a crédito;
 - ter a casa aquecida;
 - fazer face às despesas e encargos usuais;
2. Posse de bens duradouros (sem disponibilidade económica para)
 - telefone fixo ou móvel;
 - TV a cores;
 - máquina lavar roupa;
 - veículo ligeiro de passageiros ou misto;
3. Condições da habitação
 - telhado que deixa passar água, paredes/fundações/chão húmido, caixilhos ou chão apodrecido;
 - luz insuficiente;
 - ausência de retrete;
 - ausência de instalações de banho ou de duche no interior;
4. Vizinhança
 - vizinhos barulhentos ou ruídos exteriores;
 - poluição, sujidade ou outros problemas ambientais causados pelo trânsito ou indústria;

- crime, violência ou vandalismo na área de residência;

De salientar que no ICOR as questões sobre a posse de bens duradouros permitem distinguir entre a falta do item (motivada pela opção de o não possuir) e a falta forçada do item (famílias que o queriam ter, mas que não têm disponibilidade económica para tal). Apenas este último grupo é considerado como estando em privação, de forma a excluir preferências e estilos de vida do conceito de privação (Mack e Lansley [86]).

Este primeiro estudo teve por base respostas válidas de 4 304 agregados em 2007 e de 4 450 agregados em 2008. Comparando as respostas dos agregados familiares nos anos de 2007 e 2008, em relação à dimensão constrangimentos económicos é de salientar o aumento da proporção de agregados que afirma não ter capacidade financeira para i) satisfazer as despesas e encargos usuais; ii) fazer face às despesas inesperadas; iii) ter uma refeição de carne ou peixe pelo menos de dois em dois dias; e iv) fazer uma semana de férias uma vez por ano fora de casa. Importa ainda sublinhar a elevada percentagem de famílias que não tem capacidade financeira para satisfazer as despesas e encargos usuais, 76.9% em 2007 e 82.1% em 2008, bem como a percentagem de famílias que não consegue fazer férias fora de casa uma vez por ano, 65% em 2007 e 66.8% em 2008. Por outro lado, refira-se a diminuição percentual de famílias que afirma não ter capacidade para manter a casa aquecida (figura 2.2).

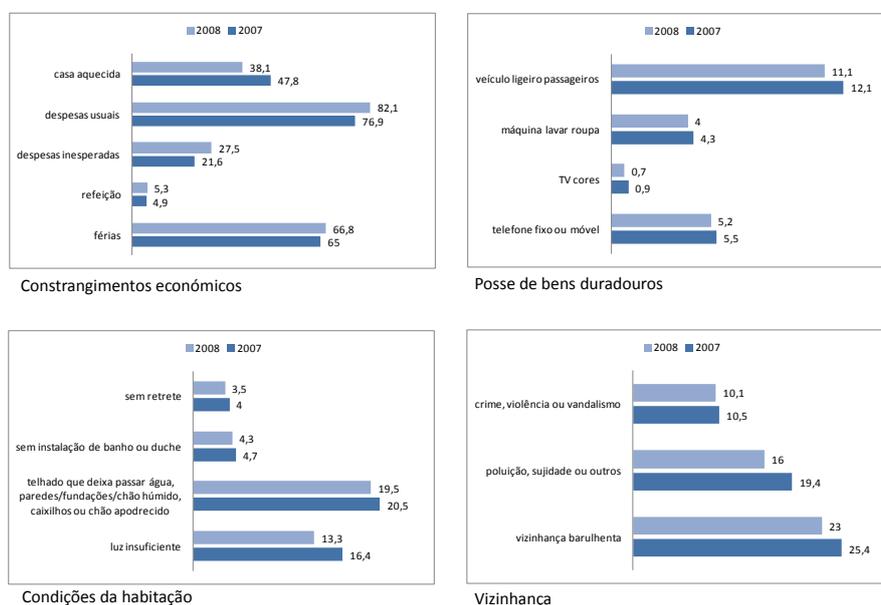


Figura 2.2: Gráficos das quatro dimensões da privação material: constrangimentos económicos, posse de bens duradouros, condições da habitação e vizinhança, nos anos de 2007 e 2008

Na dimensão posse de bens duradouros refira-se a diminuição em 2008 da percentagem de agregados que afirma não ter disponibilidade económica para possuir os diferentes bens considerados (figura 2.2). O bem veículo ligeiro de passageiros é aquele que mais agregados afirmam não ter disponibilidade económica para possuir: 12.1% em 2007 e 11.1% em 2008.

Na dimensão condições da habitação, embora a percentagem de agregados com problemas tenha diminuído de 2007 para 2008, merece destaque a elevada proporção de agregados com problemas com o telhado, fundações, caixilhos e chão: 20.5% em 2007 e 19.5% em 2008 (figura 2.2). Na dimensão vizinhança os problemas de uma forma geral tenderam a baixar percentualmente em 2008 relativamente ao ano de 2007 (figura 2.2).

Na tabela 2.1 é apresentada a percentagem de agregados afetados pela privação material, isto é, o número de itens em falta em cada uma das quatro dimensões: i) constrangimentos económicos, ii) posse de bens, iii) condições de habitação e iv) vizinhança, nos anos de 2007 e 2008. Note-se que zero itens significa não estar em privação nessa dimensão, ao passo que cinco itens significa estar privado de todos os itens. De referir para a dimensão constrangimentos económicos que 67.1% dos agregados afirma não ter capacidade financeira para conseguir satisfazer dois ou mais itens em 2007, sendo esse valor de 67.9% em 2008. Em sentido inverso, é de salientar a elevada percentagem de agregados que não está privada de qualquer item na dimensão posse de bens, 82.7% e 84%, respetivamente, em 2007 e 2008. Também as dimensões condições da habitação e vizinhança apresentam um número elevado de agregados que não estão em privação, 67.2% e 70.5% nas condições da habitação, e 63.3% e 67.3% na dimensão da vizinhança, em 2007 e 2008, respetivamente.

Neste trabalho foram utilizadas como explicativas do conceito de privação material as variáveis existência de crianças no agregado, área de urbanização do local da residência do agregado e dimensão do agregado. Existem 36.8% de agregados com crianças em 2007, sendo que em 2008 esta percentagem baixa para 33.7%. Estes agregados são compostos em média por 2.7 pessoas em 2007 e 2008, respetivamente, podendo variar de 1 a 15 indivíduos em 2007 e de 1 a 12 em 2008. A área de urbanização - área pouco povoada, área intermédia e área densamente povoada - apresenta valores semelhantes nos dois momentos temporais considerados, respetivamente, 32.8%, 33.1% e 34.1% em 2007 e 32.4%, 32.6% e 35.0% em 2008.

No estudo 1 recorrendo a uma análise fatorial confirmatória foi possível verificar que as dimensões capacidade financeira e posse de bens duradouros são as que apresentam um maior peso no conceito de privação material. Assim, no segundo estudo, e por forma a aferir sobre o comportamento da privação material, entre 2004 e 2007, duas diferentes dimensões deste conceito foram considerados: capacidade financeira e posse de bens duradouros. Os *scores* para cada dimensão são calculados como soma de itens do inquérito ICOR. Na dimensão da capacidade financeira são considerados cinco itens: i) tem capacidade financeira para pagar uma semana de férias a todo o agregado, uma vez por ano;

CAPÍTULO 2. O PAINEL ICOR E A MEDIÇÃO DA PRIVAÇÃO MATERIAL

	Constrangimentos Económicos		Posse de Bens		Condições da habitação		Vizinhança	
	2007	2008	2007	2008	2007	2008	2007	2008
0 itens	16.6%	13.8%	82.7%	84%	67.2%	70.5%	63.3%	67.3%
1 item	16.3%	18.4%	13.2%	11.9%	23.8%	21.5%	21.7%	19.7%
2 itens	20.6%	23.8%	3.1%	3.3%	6.3%	5.8%	11.4%	9.5%
3 itens	29.2%	25.7%	0.8%	0.7%	1.6%	1.5%	3.6%	3.5%
4 itens	15%	15.2%	0.2%	0.1%	1.1%	0.7%		
5 itens	2.3%	3.2%						
	100%	100%	100%	100%	100%	100%	100%	100%

Tabela 2.1: Percentagem de agregados em privação material (mais itens corresponde a maior privação), nos anos de 2007 e 2008

ii) tem capacidade financeira para assegurar uma refeição de carne ou peixe, pelo menos dois dias por semana; iii) tem capacidade financeira para manter a casa aquecida; iv) tem capacidade financeira para assegurar as despesas regulares; e v) tem capacidade financeira para fazer face a despesas inesperadas sem recorrer ao crédito. A variável capacidade financeira calculada como a soma de itens que o agregado tem capacidade financeira para assegurar toma valores de 0 a 5. A dimensão da posse de bens duradouros considera os seguintes itens: i) aparelho de tv; ii) telefone; iii) máquina de lavar roupa; e iv) veículo ligeiro de passageiros, e toma valores de 0 a 4.

A amostra em estudo inclui 991 agregados com respostas válidas a todas as questões acerca da capacidade financeira e posse de bens, entre 2004 e 2007. Em 2004, 20.7% dos agregados mostram ter capacidade financeira para assegurar as despesas regulares, 49.6% conseguem manter a habitação aquecida, 94.2% conseguem assegurar uma refeição de carne ou peixe pelo menos duas vezes por semana, 31.8% têm capacidade para ter uma semana férias e 77.3% conseguem enfrentar despesas inesperadas sem recorrer a crédito. Em termos da posse de bens duradouros, 65.9%, 91.7%, 98.7% e 89.8% dos agregados considerados possuem veículo ligeiro de passageiros, máquina de lavar roupa, televisão e telefone, respetivamente.

No terceiro estudo foi considerada a modelação longitudinal da privação material, entre 2007 e 2010, recorrendo a um modelo com trajetória latente de segunda ordem. Três dimensões do conceito latente de privação material foram considerados: constrangimentos económicos, posse de bens duradouros, e condições da habitação, sendo calculado um *score* para cada uma destas dimensões como uma soma de itens. Posteriormente, as variáveis rendimento, área de urbanização e dimensão do agregado foram utilizadas na modelação realizada como possíveis determinantes da trajetória latente da privação material.

A amostra em estudo inclui 967 agregados com respostas válidas a todas as questões sobre constrangimentos económicos, posse de bens duradouros e condições de habitação, nos anos de 2007 a 2010. Deste conjunto de agregados, 46.9% não têm capacidade financeira para assegurar mais de dois itens do *score* constrangimentos económicos, no ano de 2007, 43.7% em 2008, 41.7% em 2009 e 42.5% em 2010. Relativamente à posse de

bens duradouros, 82.3%, 85.3%, 85.7% e 86.1% dos agregados, possuem todos os bens considerados neste *score*, nos anos de 2007, 2008, 2009 e 2010, respetivamente. Quanto às condições da habitação, 77.2%, 77.6%, 78.9% e 78.6% dos agregados não apresentam qualquer um dos problemas dos considerados no cálculo do *score*, nos anos de 2007, 2008, 2009 e 2010, respetivamente.

Foram ainda utilizadas as variáveis dimensão do agregado, rendimento disponível do agregado e área de urbanização, como explicativas da trajetória da privação material. A dimensão média dos agregados é de 2.74, 2.76, 2.71 e 2.67 elementos, com um desvio padrão de 1.298, 1.304, 1.274 e 1.265, nos anos de 2007 a 2010, respetivamente. O rendimento médio disponível das famílias aumentou entre 2007 e 2010, com valores de 16 283.467, 17 334.729, 17 821.481 e 18 619.823 euros, respetivamente. O desvio padrão do rendimento tomou valores de 12 892.597, 15 578.820, 17 723.414 e 15 728.051, nos anos de 2007 a 2010, respetivamente. Quanto à área de urbanização, os 967 agregados que se mantiveram no estudo entre 2007 e 2010 dividem-se em número aproximadamente igual pelas três categorias, área muito povoada, intermédia e pouco povoada.

Nos restantes estudos (à exceção do estudo de simulação) foi decidido modelar longitudinalmente um indicador de privação material. Este indicador calculado como uma soma de itens, considerou os itens pertencentes às dimensões, constrangimentos económicos (ou ausência de capacidade financeira) e posse de bens duradouros. Deste modo, os itens considerados para o cálculo do indicador foram:

- ter uma refeição de carne ou de peixe, pelo menos de dois em dois dias;
- pagar uma semana de férias, por ano, fora de casa, a todo o agregado;
- suportar despesas inesperadas, sem recorrer a crédito;
- ter a casa aquecida;
- fazer face às despesas e encargos usuais;
- telefone móvel ou fixo;
- TV a cores;
- máquina lavar roupa;
- veículo ligeiro de passageiros ou misto.

No quarto estudo para o indicador de privação material foram considerados 10 itens, ao contrário dos outros estudos nos quais foram apenas considerados os 9 itens apresentados, incluindo-se a variável disponibilidade económica para ter computador, no cálculo do indicador de privação material. Todavia, a inclusão deste item no cálculo de um indicador de privação material nem sempre se tem mostrado consensual. De acordo com, Guio [60] este item não deve ser considerado, sendo a sua decisão justificada pelos dados do Eurobarómetro, inquérito no qual foi colocada uma questão sobre a importância de possuir

aquele bem, e cujos resultados para a posse de computador ter sido apontada apenas por uma pequena percentagem dos inquiridos.

Neste estudo foram consideradas duas subamostras, uma que consiste nos 845 agregados que se mantiveram no painel entre 2006 e 2009 (designado por caso completo) e outra com 1693 observações, que além dos agregados que se mativeram entre os anos de 2006 e 2009, inclui ainda os agregados que estiveram no painel entre os anos 2006 e 2008 (designado por caso com todas as observações), ver figura 2.3. Consequentemente, a subamostra com 1693 agregados contém aproximadamente 50% de omissões no quarto momento temporal (2009).

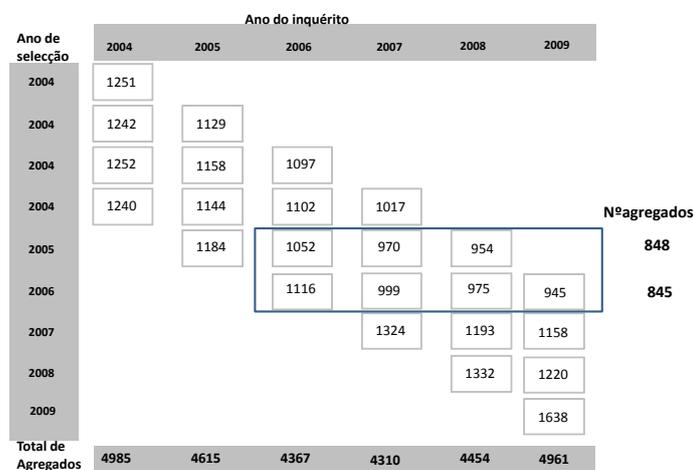


Figura 2.3: ICOR-Desenho do estudo, dentro da caixa os agregados considerados no estudo 4

Nos gráficos 2.4 e 2.5 é possível ver a percentagem de agregados com diferentes números de itens em privação nos quatro momentos temporais considerados, quer para a subamostra com todas as observações (1693 observações), quer para o caso completo (845 observações). A análise destes gráficos permite concluir que existe uma percentagem de agregados em cada escalão de privação semelhante nas duas subamostras em estudo, exceto no ano de 2008, em que a percentagem de agregados com 1 item em privação é bastante superior na subamostra com todas as observações, relativamente ao caso completo.

No quinto estudo foram consideradas sete diferentes subamostras da amostra principal que corresponde à base de dados longitudinal de 2006 a 2009. Na primeira subamostra foram considerados os agregados que permaneceram na amostra entre os anos de 2006 e 2009, o que constitui 25% da amostra original observados em quatro momentos temporais, correspondendo a 845 agregados, a qual se designa por caso completo (figura 2.6). Foram ainda consideradas mais sete diferentes subamostras dos dados originais (restantes

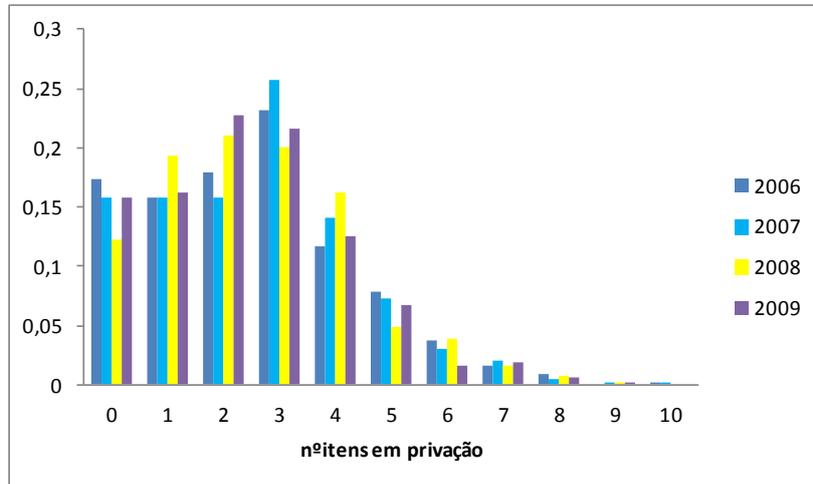


Figura 2.4: Percentagem de agregados com itens em privação, para o caso completo (845 agregados), em cada um dos quatro momentos temporais, 2006 a 2009

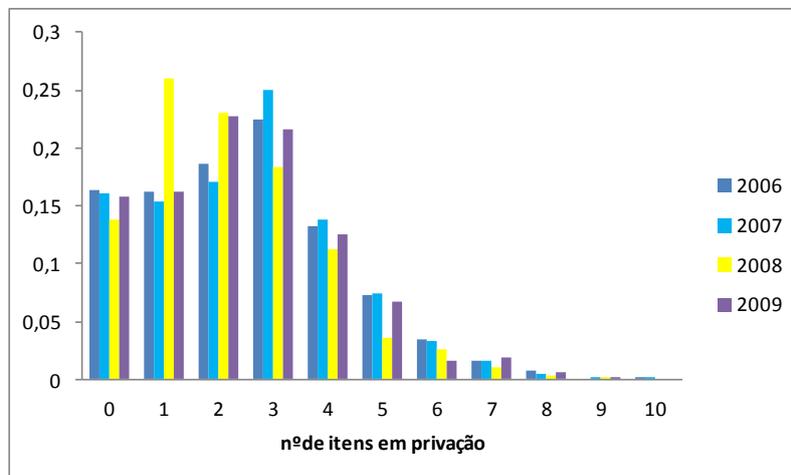


Figura 2.5: Percentagem de agregados com itens em privação, para o caso com todas as observações (1693 agregados), em cada um dos quatro momentos temporais, 2006 a 2009

figuras). Por exemplo, na subamostra 1, para além dos 845 agregados que se mantiveram nos quatro momentos temporais, foram ainda considerados os agregados que permaneceram no estudo de 2006 a 2008, num total de 1762 agregados (figura 4.3). Na subamostra 2 consideraram-se os 845 agregados que permaneceram entre 2006 e 2009 e ainda os agregados que entraram em 2007 para o painel, um total de 1929 agregados (figura 2.6).

CAPÍTULO 2. O PAINEL ICOR E A MEDIÇÃO DA PRIVAÇÃO MATERIAL

De salientar que as subamostras 1 e 2 têm um desenho equivalente em termos de omissões, pois enquanto na subamostra 1 foram considerados os agregados que estiveram nos quatro momentos temporais e os agregados que saíram em 2008, na subamostra 2 estão os agregados que se mantiveram de 2006 a 2009, e os agregados que devido à rotatividade do painel só entraram em 2007. Em termos percentuais a subamostra 1 tem 52% de omissões em 2009, enquanto a subamostra 2 tem 56% de omissões em 2006 (tabela 2.2).

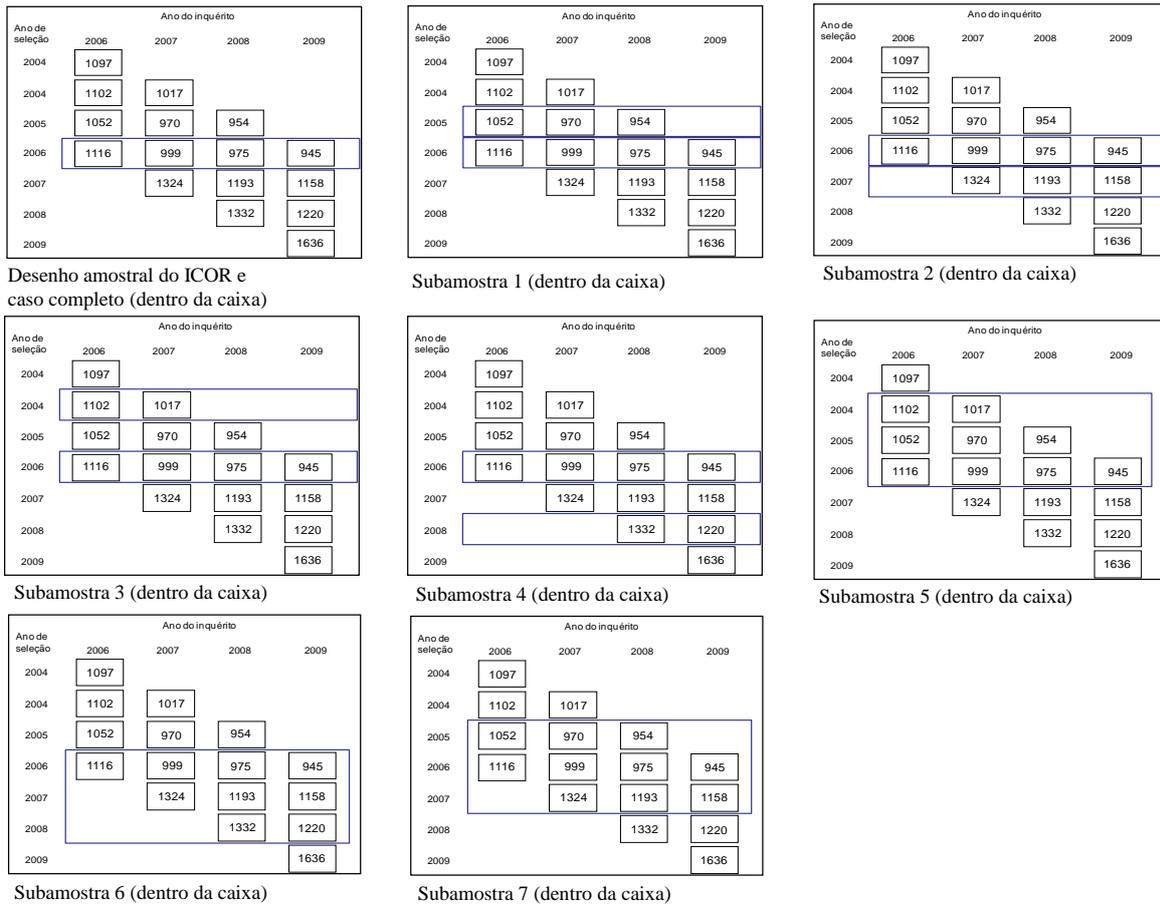


Figura 2.6: Desenho amostral do ICOR e diferentes subamostras consideradas no estudo 5

De igual modo, podem-se considerar equivalentes as subamostras 3 e 4, assim como as 5 e 6. A subamostra 3 tem um total de 1 702 agregados com 50% de omissões nos momentos temporais 2008 e 2009. Um total de 2061 agregados compõe a subamostra 4, que apresenta 59% de omissões nos dois primeiros momentos temporais (tabela 2.2). A subamostra 5 é constituída pelos agregados que saíram do estudo em 2008, os que saíram em 2009 e ainda os agregados que se mantiveram no estudo durante os anos de 2006 a 2009, totalizando um total de 2 619 agregados com 33% de omissões em 2008 e 68% em 2009 (tabela 2.2). Por fim, a subamostra 7 que consiste no caso completo mais os agregados que saíram em 2008 e os que entraram em 2007, num total de 2 846 agregados, com 38% de dados omissos em 2006, e 32% em 2009.

Na modelação longitudinal dos dados é considerado um *score* de privação material que

ano	subamostra						
	1	2	3	4	5	6	7
2006	0%	56%	0%	59%	0%	73%	38%
2007	0%	0%	0%	59%	0%	39%	0%
2008	0%	0%	50%	0%	33%	0%	0%
2009	52%	0%	50%	0%	68%	0%	32%
N	1 762	1 929	1 702	2 061	2 619	3 145	2 846

Tabela 2.2: Percentagem de observações omissas nas diferentes subamostras e a respetiva dimensão, para os anos de 2006 a 2009

corresponde a uma soma de itens. Assim, este *score* toma valores numa escala de 0 a 9 itens em privação, sendo que 0 itens em privação significam que o agregado não está privado de nenhum destes itens considerados e 9 itens em privação que os agregados estão totalmente privados. Este *score* tal como definido é calculado em cada um dos quatro momentos temporais em estudo.

Uma análise descritiva das subamostras consideradas mostrou que mais de 95% dos agregados tem menos de 6 itens em privação, dos 9 itens que compõe o *score* de privação material, independentemente da subamostra considerada. De igual modo, é possível dizer que a maior percentagem de agregados tem entre 2 a 3 itens em privação, qualquer que seja a subamostra e o momento temporal considerado (figura 2.7).

Nos estudos seis e sete são consideradas duas amostras, uma que consiste nos 845 agregados que se mantiveram no estudo nos anos de 2006 a 2009, caso completo, e a outra composta por todos os 6101 agregados, logo com omissões. Estas omissões são devidas à rotatividade do painel, existindo no entanto outras omissões. São considerados 15 diferentes padrões (tabela 2.3), o primeiro padrão corresponde aos 845 agregados que não apresentam omissões, designado por caso completo. Os quatro padrões seguintes da tabela 2.3 são devido à rotatividade do painel e os últimos 10 são devido a outro tipo de omissões.

Um *score* de privação material, análogo ao do estudo anterior, com 9 itens, foi considerado na modelação longitudinal dos dados. Uma análise descritiva dos dados (figura 2.8) permitiu ver como evolui o número médio de itens em privação para cada padrão de omissão presente nas amostras em estudo ao longo dos quatro momentos temporais, sendo de destacar a discida do número médio de itens em privação para os padrões com entradas tardias (exceto o padrão 9, em que o número médio de itens em privação se mantém análogo, nos anos de 2007 e 2008), e para os padrões com não respostas intermitentes.

CAPÍTULO 2. O PAINEL ICOR E A MEDIÇÃO DA PRIVAÇÃO MATERIAL

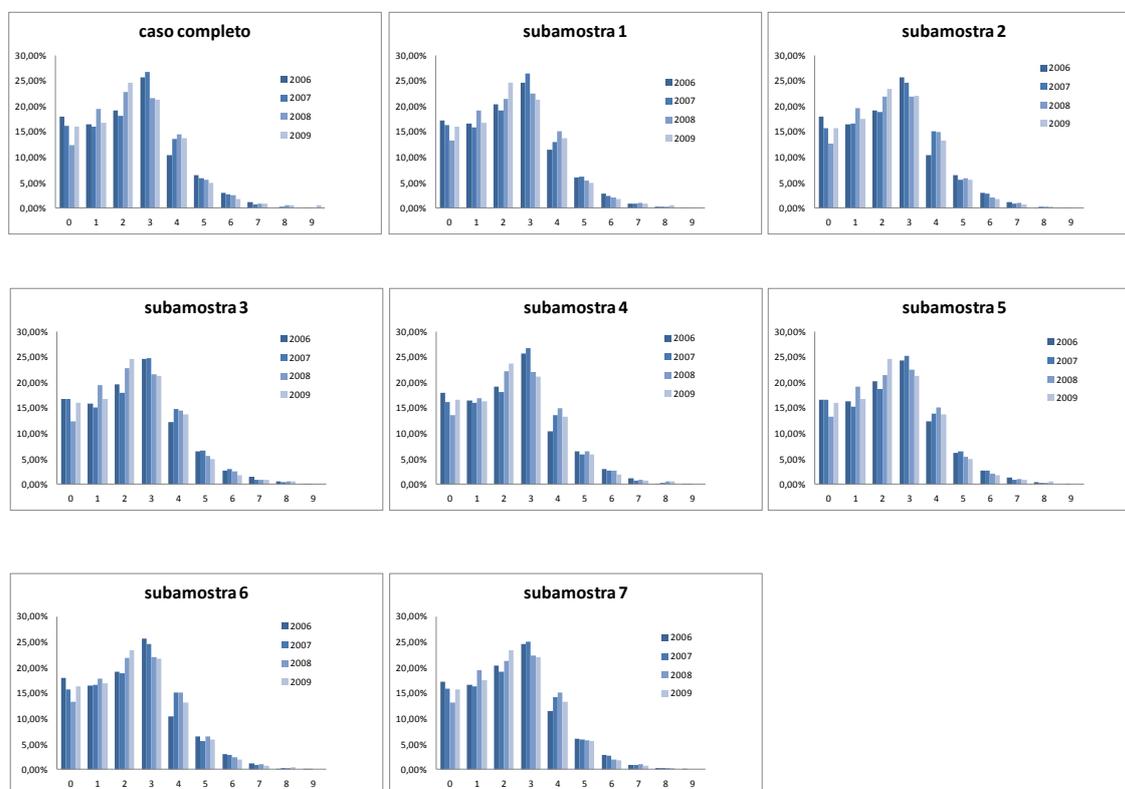


Figura 2.7: Percentagem de agregados com diferente número de itens em privação, nas várias subamostras consideradas, nos anos de 2006 a 2009

2006	2007	2008	2009	N
✓	✓	✓	✓	845
-	✓	✓	✓	1 084
-	-	✓	✓	1 216
✓	✓	-	-	857
✓	✓	✓	-	917
-	-	-	✓	51
✓	-	✓	-	59
✓	-	-	-	531
-	✓	✓	-	137
-	✓	-	-	147
✓	✓	-	✓	25
✓	-	✓	✓	33
-	-	✓	-	159
-	✓	-	✓	40
✓	-	-	✓	0

Tabela 2.3: Diferentes padrões de omissão existentes na amostra em estudo (✓ = valor observado; - = valor omissão)

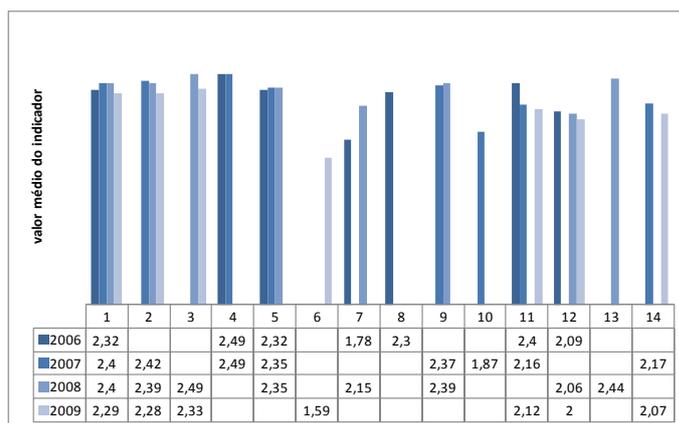


Figura 2.8: Número médio de itens em privação em cada ano, de 2006 a 2009, para os diferentes padrões de omissão

Capítulo 3

Modelações com Dados Completos

Nesta tese um dos objetivos é modelar longitudinalmente a privação material com recurso a modelos com trajetória latente. Todavia, o atraso na disponibilização dos dados longitudinais (painel ICOR) por parte do INE, face ao inicialmente acordado, fez com que, também uma modelação longitudinal com dados transversais tivesse sido inicialmente considerada. Assim, neste capítulo as metodologias utilizadas foram a análise fatorial confirmatória, modelos com equações estruturais com efeitos cruzados e desfasados e os modelos com trajetória latente, todos eles modelos com variáveis latentes contínuas. Em todos os estudos apresentados neste capítulo só foram consideradas as observações válidas, tendo sido ignorado a existência de dados omissos.

Neste capítulo é primeiro apresentada a evolução história dos modelos com trajetória latente no *framework* dos modelos com equações estruturais, o procedimento de estimação destes modelos e as medidas de qualidade do ajustamento usualmente utilizadas. De seguida, são apresentadas as diferentes metodologias que foram utilizadas nas modelações em que foram ignoradas as observações omissas, e os resultados obtidos nas aplicações práticas.

3.1 Modelos com Trajetória Latente: evolução histórica, estimação do modelo e medidas de qualidade de ajustamento

Evolução histórica dos Modelos com Trajetória Latente no *framework* dos Modelos com Equações Estruturais

Os modelos com equações estruturais são uma técnica estatística usada para testar e estimar relações lineares entre variáveis, considerando a componente do erro de medida. As variáveis no modelo podem ser variáveis observadas ou variáveis latentes. Assim, um modelo com equações estruturais consiste num conjunto hipotético de relações lineares diretas e não diretas entre um conjunto de variáveis observadas e latentes. As relações

diretas implicam uma influência de uma variável noutra, enquanto que as relações indiretas são correlações e não implicam influência. As relações entre as variáveis latentes do modelo especificam a componente estrutural do modelo. Por outro lado, a relação das variáveis latentes com as variáveis observadas é designada por modelo de medida.

A análise fatorial confirmatória (AFC), também conhecida por modelo de medida no âmbito dos modelos com equações estruturais, sendo um seu caso particular, tem sido usada em diversas áreas do conhecimento com o intuito de testar a estrutura fatorial das variáveis observadas utilizadas para medir os conceitos latentes, Hair et al. [63]. Esta estrutura fatorial é estabelecida *a priori*, de acordo com o modelo teórico previamente aceite, sendo depois validada, face à amostra obtida, com recurso à AFC.

Desde os anos 70 que o desenvolvimento dos modelos com equações estruturais e de *software* aplicacional tem verificado um grande crescimento, nas mais variadas áreas do conhecimento, tais como Gestão, Economia, Marketing, Sociologia e Psicologia. A multidisciplinaridade de origens desta técnica pode ser encontrada em Blalock [11] e Golderberg e Duncan [50]. Todavia, o seu aparecimento deve-se aos trabalhos pioneiros do biométrico Sewall Wright [138], [139], [140], [141] no âmbito da análise de caminhos. Estes trabalhos de Wright deram origem aos trabalhos dos investigadores das ciências sociais, tais como, Blalock [11] e Golderberg e Duncan [50]. De salientar também o contributo do trabalho de Spearman [150] no início do século XX, estabelecendo a relação dos fatores latentes com as variáveis observadas, isto é, o estudo do que se designa atualmente por modelo de medida.

O desenvolvimento de um *framework* de Modelos com Equações Estruturais ficou a dever-se aos trabalhos de Jöreskog [75], Keesling [80], Wiley [165], englobando entre outras a análise de caminhos e a análise fatorial, assim como um modelo de equações simultâneas com variáveis latentes. O crescimento exponencial da utilização destes modelos e a sua grande popularidade é indissociável do aparecimento do *software* estatístico LISREL (Linear Structural Relations) desenvolvido por Jöreskog e Sörbom, nos anos 70, tendo inclusivamente dado origem ao designado por notação LISREL e que é amplamente utilizada na literatura da área, Bollen [18], Finkel [48], Kline [81]. Inúmeras aplicações de Jöreskog e seus colaboradores podem ser encontradas em distintas áreas do conhecimento.

Uma breve resenha histórica e alguns desenvolvimentos metodológicos podem ser encontrados em Bentler [12], Bentler [13]. De igual modo, MacCallum e Austin [85] apresentam um resumo de trabalhos utilizando os modelos com equações estruturais. Uma outra referência a ter em consideração é Kaplan [79]. Por outro lado, também Salgueiro [135] faz uma apresentação teórica desta metodologia e algumas das suas extensões, complementando com aplicações práticas. A utilização destes modelos na análise da mudança pode ser encontrada em Bollen e Curran [21] e McArdle [88].

No trabalho de McArdle [88] é apresentada uma grande variedade de modelos para a análise da estabilidade e da mudança no contexto dos modelos com variável latente e uma discussão destes diferentes modelos é considerada, salientando as respetivas vanta-

gens e desvantagens. De acordo com este autor, é necessário sublinhar que o estudo da diferença e o estudo da mudança são questões distintas levadas a cabo por várias abordagens estatísticas, utilizando dados de diferentes tipos (longitudinais ou transversais). Um desses modelos foi proposto por Jöreskog [76], nas análises preliminares com dados longitudinais, o modelo autorregressivo que assenta no princípio das séries cronológicas. Por outro lado, também, Bollen e Curran [20] e Finkel [48] mostraram como incorporar na análise características de modelação da mudança e uma abordagem autoregressiva num único modelo. Este modelo designado por modelo com equações estruturais com efeitos cruzados e desfasados, usa uma ou mais variáveis medida num dado momento temporal como predictoras de si própria e de outras variáveis no momento temporal seguinte.

Por outro lado, a análise fatorial confirmatória multigrupos permite aferir a existência de comportamentos diferentes por parte dos diferentes grupos em relação ao conceito em estudo, através de um teste de invariância dos parâmetros do modelo considerado nos distintos grupos em análise, isto é, é verificada a estabilidade do modelo nos diferentes grupos. Diversas aplicações utilizando este procedimento podem ser encontradas em Bollen [18], Byrne [24], Finkel [48], Kline [81], entre outros. De acordo com McArdle [88], dado que, uma análise fatorial confirmatória multigrupos permite testar diferenças entre os grupos considerados, sendo os grupos diferentes momentos temporais este procedimento pode ser encarado como uma análise longitudinal, embora os dados subjacentes ao estudo sejam transversais.

Em Collins [25] e Collins e Horn [26] é apresentada uma discussão de diferentes formas de modelos matemáticos, no âmbito dos modelos com variáveis latentes contínuas, para análise de dados longitudinais, sendo um desses casos os modelos com trajetória latente. Estes modelos têm sido alvo de grandes desenvolvimentos nos últimos tempos, sendo bastante utilizados nas mais variadas áreas científicas, visando lidar com a existência de omissões, com a natureza categorial das variáveis, com diferentes tipos de dados, múltiplos indicadores, entre outros. Trabalhos com destaque visando este tipo de modelação são: Bollen e Curran [21], Duncan et al. [36], Duncan et al. [38], Enders e Bandalos [43], McArdle e Epstein [90], McArdle e Hamagami [91], Muthén e Muthén [106], Salgueiro [134], [136], Sayer e Collins [137], Raykov [124]. De acordo com Preacher et al. [117], os modelos com trajetória latente assentam na ideia de que os indivíduos mudam ao longo do tempo, mas não necessariamente da mesma forma.

O desenvolvimento de um modelo matemático para o estudo da mudança teve início no século XIX com os trabalhos de Gompertz. Gompertz [51] e [52] tentou encontrar um modelo matemático com uma única trajetória explicativa das trajetórias das taxas de mortalidade ao longo do tempo para um grupo de indivíduos. Posteriormente, e devido aos trabalhos de Griliches [58] e Wishart [169] foi possível passar da estimação da trajetória para um grupo, para a estimação de uma trajetória específica para cada elemento desse grupo, recorrendo a modelos de análise da variância e de covariância. Deste modo, permitindo-se a análise de diferenças individuais nas trajetórias e a explicação destas por outras variáveis.

Assim, até meados do século XX, os dados longitudinais foram analisados estatisticamente recorrendo maioritariamente a modelos de análise de variância e de análise de covariância, nos quais as estimativas das trajetórias individuais determinadas e as diferenças das estimativas examinadas considerando os indivíduos como membros de um grupo.

Por outro lado, sensivelmente na mesma altura, os trabalhos de Baker [8], Rao [122] e Tucker [159] sugeriram a utilização de variáveis latentes no âmbito dos modelos de crescimento, tendo mais tarde Meredith e Tisak [93] proposto a modelação das trajetórias no contexto da análise fatorial confirmatória. O trabalho de Meredith e Tisak assentou nas potencialidades dos modelos com equações estruturais para estimar e testar uma variedade de modelos com trajetória latente. Um dos princípios básicos dos modelos com trajetória latente é o de que existe um fator latente, não observado, que explica a estrutura de associações entre as variáveis observadas. Assim, o interesse não está especificamente nas medidas repetidas observadas mas sim na estrutura de associações, a qual é explicada pelos efeitos aleatórios da trajetória latente.

Aqueles autores podem assim ser considerados como os precursores da utilização de modelos de trajetórias no contexto dos modelos com equações estruturais e embora Meredith e Tisak [94] apresentem uma discussão teórica destes modelos, também Duncan et al. [36], MacCallum e Austin [85], Willet e Sayer [168] fazem uma descrição destes modelos, mas de uma forma menos teórica. O objetivo de tais modelos é capturar informação sobre as diferenças *inter-individual* na mudança *intra-individual*, Nesselroad [107]. Nos dias de hoje, uma referência incontornável na apresentação e discussão dos modelos com trajetória latente é sem dúvida o livro de Bollen e Curran [21]. Também os livros de Duncan et al. [34] e Preacher et al. [117], devem ser referências a ter em consideração para quem pretende trabalhar nesta área da modelação estatística.

A estimação do modelo

A estimação de um modelo com equações estruturais consiste em encontrar estimativas para os parâmetros do modelo que resultem numa matriz de variâncias-covariâncias que reproduza o melhor possível a matriz de variâncias-covariâncias implícita ao modelo teórico considerado. Assim, sendo Σ a matriz de variâncias-covariâncias subjacente ao modelo com equações estruturais, pretende-se que a matriz de variâncias-covariâncias entre as variáveis observadas \mathbf{S} , esteja o mais próximo possível da matriz Σ . Representando por θ os parâmetros do modelo, os valores estimados para θ são determinados por forma a minimizar uma função distância entre $\Sigma(\hat{\theta})$ e \mathbf{S} . Esta função distância pode assumir diferentes formas, mas todas elas verificando as seguintes condições: 1) $F(\mathbf{S}, \Sigma)$ é um escalar; 2) $F(\mathbf{S}, \Sigma) \geq 0$; 3) $F(\mathbf{S}, \Sigma) = 0$ se e só se $\Sigma = \mathbf{S}$; e 4) $F(\mathbf{S}, \Sigma)$ é contínua

em \mathbf{S} e Σ . A escolha da função distância depende do método de estimação selecionado⁸ recorrendo-se tradicionalmente ao método da máxima verosimilhança para a estimação dos parâmetros nos modelos com equações estruturais. Este método permite obter estimadores para os parâmetros, θ , que gozam das seguintes propriedades: a) consistentes; b) assintoticamente não enviesados; e c) assintoticamente eficientes; sendo necessário que as variáveis observadas apresentem distribuição normal multivariada

A função de distância tem a seguinte forma

$$F_{ML} = \log|\Sigma(\theta)| - \log|\mathbf{S}| + \text{tr}[\mathbf{S}\Sigma^{-1}(\theta)] - k, \quad (3.4)$$

em que se assume que $\Sigma(\theta)$ e \mathbf{S} são matrizes definidas positivas, e k indica o número de variáveis. A função F_{ML} toma o valor zero quando o modelo estima perfeitamente os valores da matriz de covariâncias amostrais, isto é, $\Sigma(\hat{\theta}) = \mathbf{S}$.

De sublinhar que, na estimação de um modelo com equações estruturais são simultaneamente estimadas as relações entre as variáveis latentes e as variáveis observadas (componente de medida) e as relações especificadas entre as variáveis latentes (componente estrutural do modelo). A identificação de um modelo deste tipo verifica-se sempre que cada elemento de θ pode ser escrito de forma única em função dos elementos de Σ (Wiley [165], Bollen [18]).

De modo análogo, para a estimação de parâmetros de um modelo com trajetória latente, que é definido por uma estrutura de variâncias-covariâncias, bem como por uma estrutura de médias, a ideia subjacente visa seleccionar valores para os parâmetros do modelo que reproduzem o mais próximo possível a estrutura de médias e de variâncias-covariâncias das variáveis observadas. O vetor das médias amostrais ($\bar{\mathbf{Y}}$) e a matriz de variâncias-covariâncias amostrais (\mathbf{S}) substituem o vetor de médias populacionais (μ) e a matriz de covariâncias populacionais (Σ), que são desconhecidos. Representando os parâmetros estimados do modelo por $\hat{\theta}$, o objetivo é o de escolher valores de $\hat{\theta}$, de forma

⁸A função distância para o método dos mínimos quadrados generalizados (GLS) é dada por

$$F_{GLS} = \left(\frac{1}{2}\right)\text{tr}(\{[\mathbf{S} - \Sigma(\theta)]\mathbf{W}^{-1}\}^2), \quad (3.1)$$

em que \mathbf{W}^{-1} uma matriz de pesos, sendo usualmente $\mathbf{W}^{-1} = \mathbf{S}^{-1}$.

Para o método dos mínimos quadrados não ponderados (ULS) a função distância é dada por

$$F_{ULS} = \left(\frac{1}{2}\right)\text{tr}[(\mathbf{S} - \Sigma(\theta))^2]. \quad (3.2)$$

Para o método dos mínimos quadrados ponderados (WLS) a função distância é dada por

$$F_{WLS} = [\mathbf{s} - \sigma]^T \mathbf{W}^{-1} [\mathbf{s} - \sigma]. \quad (3.3)$$

em que \mathbf{s}^T é o vetor dos elementos da parte inferior da matriz de variâncias-covariâncias das variáveis observadas \mathbf{S} , σ^T é o vetor dos elementos correspondentes na matriz de variâncias-covariâncias subjacente ao modelo Σ e \mathbf{W}^{-1} é uma matriz de pesos

a que $\boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$ fique próximo de $(\bar{\mathbf{Y}})$ e $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ próximo de \mathbf{S} .

No método de estimação da máxima verosimilhança, esta otimização é definida usando a função de verosimilhança normal multivariada. Isto é, a estimação de máxima verosimilhança resulta num conjunto de parâmetros estimados $\hat{\boldsymbol{\theta}}$ que maximizam o logaritmo da função de verosimilhança

$$\log L = -\frac{1}{2} \sum_{i=1}^N \{k \log(2\pi) + \log|\boldsymbol{\Sigma}| + (\mathbf{Y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \boldsymbol{\mu})\}. \quad (3.5)$$

De acordo com Jorëskog[74], pode ser mostrado que a função $\log L$ é maximizada quando a seguinte função “distância” é minimizada

$$F_{ML} = \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \log|\mathbf{S}| + \text{tr}[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{S}] - k - [\bar{\mathbf{Y}} - \boldsymbol{\mu}(\boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) [\bar{\mathbf{Y}} - \boldsymbol{\mu}(\boldsymbol{\theta})] \quad (3.6)$$

ou seja, dado \mathbf{S} e $(\bar{\mathbf{Y}})$, o método procura o vetor de estimativas dos parâmetros, $\hat{\boldsymbol{\theta}}$, que produzem matrizes $\boldsymbol{\Sigma}$ e $\boldsymbol{\mu}$ que minimizem F_{ML} . Esta função tem um mínimo em zero e não tem máximo. Aproxima-se do seu mínimo quando $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) \rightarrow \mathbf{S}$ e $\boldsymbol{\mu}(\hat{\boldsymbol{\theta}}) \rightarrow \bar{\mathbf{Y}}$, e é zero quando estas relações são iguais.

Os erros padrão assintóticos das estimativas dos parâmetros (elevados ao quadrado) estão na diagonal da inversa da matriz de informação de Fisher

$$I_{ML} = - \sum_{i=1}^N \frac{\partial^2 \log L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}. \quad (3.7)$$

De acordo com Browne [23] o estimador de máxima verosimilhança mantém as suas propriedades assintóticas quando as variáveis observadas apresentam um achatamento semelhante ao da distribuição normal multivariada. Todavia, se os dados apresentam um desvio da normalidade é desejável a utilização de uma versão robusta do estimador da máxima verosimilhança.

Assim, sendo a função de verosimilhança para os dados observados dada pela seguinte expressão

$$\log L = \sum_{i=1}^N \log L_i. \quad (3.8)$$

e definindo

$$\mathbf{A} = - \sum_{i=1}^N \frac{\partial^2 \log L_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}, \quad (3.9)$$

e

$$\mathbf{B} = \sum_{i=1}^N \frac{\partial \log L_i}{\partial \boldsymbol{\theta}} \times \frac{\partial \log L_i}{\partial \boldsymbol{\theta}^T}, \quad (3.10)$$

a abordagem pelo método de estimação de máxima verosimilhança robusta (MLR) faz uso da matriz de informação de Fisher do seguinte modo,

$$I_{MLR} = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}. \quad (3.11)$$

Esta abordagem é discutida por Yuan e Bentler [172] e recomendada sempre que a dimensão da amostra é pequena ou média.

A identificação de um modelo com trajetória latente verifica-se sempre que cada elemento de θ pode ser escrito de forma única em função dos elementos de μ e de Σ . Todavia, esta questão será discutida com maior detalhe mais à frente (ver estudo de simulação).

Medidas de qualidade do ajustamento modelo-dados

No âmbito dos modelos com trajetória latente e modelos com equações estruturais existem várias medidas que permitem aferir sobre a qualidade do ajustamento modelo-dados, embora com diferentes abordagens. As medidas de qualidade do ajustamento consideradas nesta tese podem ser classificadas em três grupos: a) medidas de ajustamento incremental; b) medidas de parcimónia; e c) medidas de ajustamento absoluto.

As medidas de ajustamento incremental comparam o modelo nulo ao modelo estimado, isto é, comparam um modelo mais restritivo, geralmente tomando como parâmetros livres as variâncias das variáveis observadas e fixando a zero as suas covariâncias, com o modelo estimado. Neste grupo de medidas, os índices mais frequentemente utilizados na aplicação dos modelos com equações estruturais, bem como nos modelos com trajetória latente são o *Tucker-Lewis Index* (TLI) e o *Comparative Fit Index* (CFI). Estas medidas tomam valores entre 0 (ausência total de ajustamento) e 1 (ajustamento ideal), sendo considerado um bom ajustamento resultados a partir de 0.90.

Os critérios de parcimónia, *Akaike Information Criteria* (AIC) e *Bayesian Information Criteria* (BIC), consideram a complexidade do modelo e o número de parâmetros a estimar. Estas medidas que permitem a comparação de modelos, favorecem os modelos mais parcimoniosos, e conseqüentemente penalizam os modelos mais complexos. Para estes critérios quanto menor o valor obtido melhor o ajustamento aos dados do modelo considerado. Por outro lado, como estes critérios são calculados a partir da função de máxima verosimilhança, apenas são disponibilizados para os modelos com variáveis contínuas.

As medida de ajustamento absoluto usualmente utilizadas são o *Root Mean Square Error of Approximation* (RMSEA) que é uma medida baseada no erro de aproximação à população, o *Root Mean Square Residual* (RMR) que é uma medida de média dos resíduos ajustados, e a estatística de χ^2 que é um teste estatístico cujo objetivo é a não rejeição da hipótese nula de que a estrutura de variâncias-covariâncias entre as variáveis observadas reproduz a estrutura de variâncias-covariâncias do modelo especificado. De acordo com Schumaker e Lomax [147], valores de RMSEA inferiores a 0.05 indicam um bom ajustamento, mas valores até 0.08 podem ser considerados aceitáveis. A medida RMR toma valores entre 0 e 1 se for usada uma matriz de correlações e é desejável que assuma valores inferiores a 0.05. Por outro lado, se for utilizada uma matriz de variâncias-covariâncias, esta medida não apresenta limite superior, embora seja sempre positiva. A sua inter-

pretação é feita relativamente à dimensão das variâncias e covariâncias observadas, não podendo ser utilizada em modelos com variáveis diferentes.

No pacote estatístico Mplus, além das medidas referidas (exceto a medida RMR), estão ainda disponíveis medidas baseadas nos resíduos, tais como, *Weighted Root Mean Square Residual* (WRMR) e *Standardized Root Mean Square Residual* (SRMR), sendo os valores recomendados para $SRMR < 0.08$ e para $WRMR < 0.90$, (Yu e Muthén [171]). De acordo com Muthén [99], a medida WRMR é adequada para modelos em que as estatísticas amostrais têm variâncias muito dispersas e quando estas estatísticas estão em diferentes escalas, tal como acontece em modelos com estruturas de médias e/ou *thresholds*. A medida WRMR é também indicada nos casos em que as variáveis observadas não têm distribuição normal. A medida SRMR, sendo as variáveis observadas categoriais só é disponibilizada nos casos em que não existe uma estrutura de thresholds ou não são consideradas variáveis explicativas no modelo (Muthén [99]).

Para uma apresentação mais formal das medidas de ajustamento disponíveis no pacote estatístico Mplus consultar apêndice I.

3.2 Modelos de Análise Fatorial Confirmatória: Um caso particular de um modelo com equações estruturais

3.2.1 Modelos de Análise Fatorial Confirmatória Multigrupos

O modelo com equações estruturais é uma das mais importantes técnicas estatísticas da análise multivariada e pode ser vista como uma combinação de análise fatorial confirmatória (modelo de medida) e regressão (modelo estrutural). O modelo de medida representa a relação entre as variáveis observadas e as variáveis latentes, enquanto que o modelo estrutural representa a relação entre as variáveis latentes do modelo (ver figura 3.1). Apresenta a vantagem de permitir a estimação simultânea de um conjunto de relações de dependência entre variáveis que podem ser latentes, considerando a componente erro de medida, Salgueiro [135].

O modelo com equações estruturais pode ser definido da seguinte forma

$$\boldsymbol{\eta} = \mathbf{B} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad (3.12)$$

$$\mathbf{X} = \boldsymbol{\Lambda}_X\boldsymbol{\xi} + \boldsymbol{\delta} \quad (3.13)$$

$$\mathbf{Y} = \boldsymbol{\Lambda}_Y\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad (3.14)$$

em que, a primeira equação constitui a componente estrutural do modelo e as duas equações seguintes constituem, cada uma delas, um modelo de medida. $\boldsymbol{\xi}$ e $\boldsymbol{\eta}$ representam os vetores das variáveis latentes independentes e dependentes, respetivamente. \mathbf{X} e \mathbf{Y}

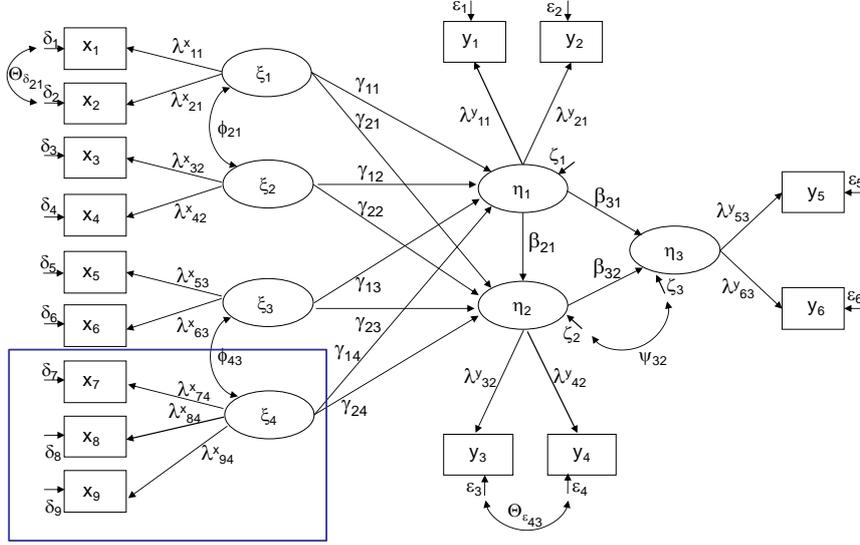


Figura 3.1: Diagrama de um modelo com equações estruturais com quatro variáveis latentes independentes (ξ) e três dependentes (η)

representam os vetores de variáveis observadas das latentes independentes e dependentes, respetivamente. Os vetores dos termos residuais dos modelos de medida são designados por δ e ε , sendo o vetor dos termos residuais das relações estruturais designado por ζ . As matrizes dos parâmetros nos modelos de medida são representadas por Λ_X e Λ_Y e os seus elementos designam-se por pesos fatoriais, sendo usual impor restrições para estes coeficientes. B e Γ são as matrizes dos parâmetros nas relações estruturais. As matrizes de variância-covariância das variáveis latentes independentes e dos termos residuais das relações estruturais são designadas por Φ e Ψ , isto é, $E(\xi\xi^T) = \Phi$ e $E(\zeta\zeta^T) = \Psi$, respetivamente. As matrizes de variância-covariância dos termos residuais dos modelos de medida δ e ε são Θ_δ e Θ_ε , respetivamente.

Os pressupostos subjacentes ao modelo com equações estruturais são

- as variáveis são medidas como desvios face aos seus valores médios, então $E(\eta) = E(\xi) = E(\zeta) = E(X) = E(Y) = E(\varepsilon) = E(\delta) = \mathbf{0}$;
- os erros de medida não estão correlacionados com as variáveis latentes, isto é, ε não está correlacionado com η , nem δ está correlacionado com ξ , assim temos que $E(\xi\delta^T) = E(\delta\xi^T) = E(\eta\varepsilon^T) = E(\varepsilon\eta^T) = \mathbf{0}$
 $E(\xi\varepsilon^T) = E(\varepsilon\xi^T) = E(\eta\delta^T) = E(\delta\eta^T) = \mathbf{0}$;
- os erros de medida não estão correlacionados entre si, nem com os termos residuais das equações estruturais, ou seja $E(\delta\varepsilon^T) = E(\varepsilon\delta^T) = \mathbf{0}$
 $E(\zeta\varepsilon^T) = E(\varepsilon\zeta^T) = E(\zeta\delta^T) = E(\delta\zeta^T) = \mathbf{0}$;

- ξ e ζ não estão correlacionados, isto é, $(\xi\zeta^T) = E(\zeta\xi^T) = \mathbf{0}$;
- \mathbf{B} tem zeros na diagonal e $(\mathbf{I}-\mathbf{B})$ é uma matriz não singular, pelo que $(\mathbf{I}-\mathbf{B})^{-1}$ existe.

Se o pretendido é apenas a utilização de uma relação do tipo

$$\mathbf{X} = \mathbf{\Lambda}_X \xi + \delta, \quad (3.15)$$

isto é, um modelo de medida, diz-se que se está perante um modelo de análise fatorial confirmatória (por exemplo, ver caixa na figura 3.1).

Uma AFC multigrupos permite aferir a existência de comportamentos diferentes por parte dos diferentes grupos em relação ao conceito em estudo, podendo estes grupos ser diferentes momentos temporais. Assim, para cada um destes grupos (diferentes momentos temporais) deve ser definido um modelo de medida. Posteriormente são impostas restrições de igualdade entre parâmetros dos vários grupos, designadamente ao nível dos pesos fatoriais. Do ponto de vista estatístico, dois modelos, em que um deles é submodelo do outro, podem ser comparados calculando a diferença entre os valores de χ^2 dos dois modelos ($\Delta\chi^2$). Os correspondentes graus de liberdade (Δdf) são dados pela diferença de graus de liberdade dos dois modelos. A hipótese nula em teste é a de invariância dos parâmetros nos dois modelos, devendo o valor de $\Delta\chi^2$ obtido ser comparado com o valor crítico de uma distribuição de χ^2 com graus de liberdade dados por Δdf . Rejeitar a hipótese nula corresponde a concluir que os dois modelos não são equivalentes, ou seja, que existem diferenças no conceito em estudo nos vários momentos temporais considerados. Note-se que o teste da diferença de χ^2 pressupõe a normalidade das variáveis observadas.

3.2.2 Resultados do estudo da Privação Material utilizando um Modelo de Análise Fatorial Confirmatória Multigrupos

Tendo como objectivo o estudo da privação material nas suas quatro dimensões, recorreu-se à AFC para estimar, em cada momento temporal, um modelo de medida para a privação material, em que esta se considera medida por quatro indicadores (figura 3.2). Cada indicador corresponde à soma dos itens considerados na respetiva dimensão.

As medidas de ajustamento (tabela 3.1) mostram ajustamentos modelo-dados razoáveis para os anos de 2007 e 2008. Para o ano de 2007, o valor da estatística de qui-quadrado é de 51.57 com 2 graus de liberdade e o valor do RMSEA é de 0.076. Para o modelo ajustado aos dados de 2008, o valor da estatística de qui-quadrado é de 65.26 com 2 graus de liberdade e o RMSEA é de 0.084, valor este que se encontra no limite da razoabilidade. Da análise dos valores de CFI e RMR obtidos para os modelos estimados em cada um dos anos, e da comparação dos valores de AIC obtidos nos dois anos, é possível concluir que existe um melhor ajustamento modelo-dados para o ano de 2007.

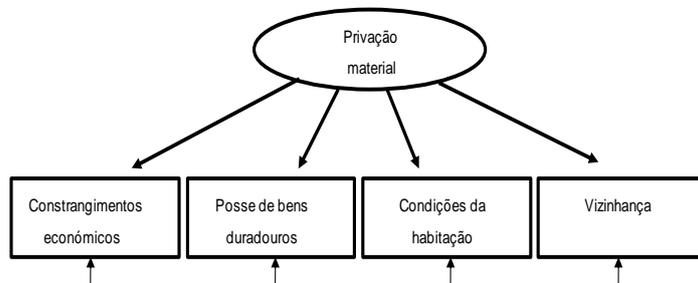


Figura 3.2: Diagrama de um modelo de análise fatorial confirmatória com a privação material medida por quatro indicadores

	χ^2	df	RMSEA	AIC	CFI	RMR
2007	51.57	2	0.076	67.57	0.97	0.043
2008	65.26	2	0.084	81.26	0.96	0.050

Tabela 3.1: Medidas de ajustamento: AFC da privação material medida por quatro dimensões

Os pesos factoriais mostraram-se todos estatisticamente significativos, embora seja de salientar que em ambos os anos se verificou um peso fatorial estimado maior para a variável posse de bens duradouros (0.79 em 2007 e 0.77 em 2008), relativamente aos pesos factoriais das restantes dimensões. No entanto, a variável vizinhança mostrou ter um peso fatorial muito pequeno, igual a 0.09 nos dois anos, valor esse bastante inferior ao 0.6 considerado na literatura como aceitável. De referir ainda que também os pesos factoriais para a dimensão condições da habitação estão no limite do aceitável. A tabela 3.2 apresenta as estimativas obtidas para os pesos factoriais nos dois anos em estudo, numa solução estandardizada. Entre parênteses encontram-se os valores dos testes t à hipótese nula de que o peso fatorial é nulo.

	2007	2008
Constrangimentos Económicos	0.72 (35.12)	0.72 (31.55)
Posse de bens	0.79 (31.85)	0.77 (28.06)
Condições da habitação	0.57 (27.08)	0.53 (23.58)
Vizinhança	0.09 (3.82)	0.09 (3.67)

Tabela 3.2: Estimativas, numa solução estandardizada, dos pesos factoriais (valores dos testes t): AFC da privação material medida por quatro dimensões. A *bold* estão os valores que se mostraram significativos.

Face aos resultados obtidos, foi tomada a opção de não considerar a vizinhança como indicador de medida da privação material, tendo sido estimado um novo modelo con-

	2007	2008
Constrangimentos Económicos	0.73 (33.33)	0.73 (29.45)
Posse de bens	0.78 (30.29)	0.76 (26.53)
Condições da habitação	0.53 (24.81)	0.49 (21.36)

Tabela 3.3: Estimativas, numa solução estandardizada, (valores do teste t) dos pesos fatoriais: AFC da privação material medida por três dimensões. A *bold* estão os valores que se mostraram significativos.

siderando a privação material como medida apenas por três dimensões: constrangimentos económicos, posse de bens duradouros e condições da habitação (figura 3.3). Os valores obtidos para as estimativas dos pesos fatoriais (tabela 3.3) mostraram mais uma vez um maior valor para a estimativa do peso factorial correspondente à variável posse de bens, 0.78 em 2007 e 0.76 em 2008, embora todas as variáveis tenham apresentado estimativas dos pesos fatoriais significativas. A variável condições de habitação é aquela que apresenta um peso factorial menor nos dois anos, 0.53 e 0.49, respetivamente em 2007 e 2008.

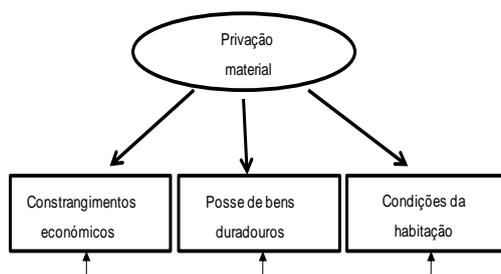


Figura 3.3: Diagrama de um modelo de análise fatorial confirmatória com a privação material medida por 3 indicadores

Posteriormente foram considerados dois grupos, um correspondendo ao ano de 2007 com 4 303 observações válidas e outro correspondendo ao ano de 2008 com 4 450 observações válidas, tendo-se realizado uma AFC multigrupos. Este procedimento estatístico teve como propósito detetar possíveis diferenças no modelo de medida da privação material nos dois momentos temporais em estudo. Face aos resultados obtidos, foi possível concluir quanto à invariância dos pesos fatoriais nos diferentes momentos temporais ($\Delta\chi^2= 1.91$, com $\Delta df = 3$).

Neste estudo foram utilizadas como variáveis caracterizadoras dos agregados: a existência de crianças no agregado (categoria de referência: não existirem crianças no agregado), a dimensão do agregado e a área de urbanização do local onde o agregado reside (muito povoada, intermédia e pouco povoada). Para esta última variável e com o objectivo de a utilizar como variável explicativa nos modelos em estudo, foram criadas duas variáveis *dummy*, sendo a categoria de referência área de urbanização pouco povoada. A variável

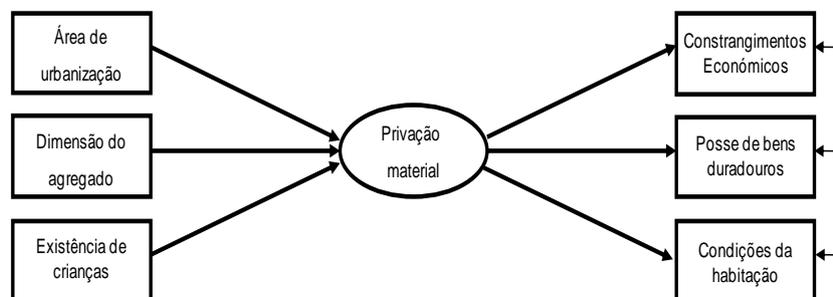


Figura 3.4: Diagrama de um modelo com equações estruturais com a privação material medida por 3 indicadores e 3 variáveis explicativas

dimensão do agregado foi considerada como contínua. Foi então estimado o modelo com equações estruturais representado na figura 3.4. Para as variáveis explicativas existência de crianças e dimensão do agregado, os valores obtidos para as estimativas dos parâmetros em 2007 (-0.16 e -0.14) e no ano seguinte (-0.06 e -0.18) permitem concluir que a privação é menor quando existem crianças no agregado e nos agregados de maior dimensão. Em relação à variável área de urbanização, os valores estimados dos parâmetros não permitem conclusões idênticas nos diferentes momentos temporais. Assim, no ano de 2007, áreas muito povoadas e intermédias mostraram menor privação em relação a áreas pouco povoadas (os valores obtidos foram -0.69 e -0.67), e no ano de 2008, as áreas muito povoadas e intermédias mostraram maior privação relativamente a áreas pouco povoadas, tendo-se obtido os valores 0.37 e 0.46 (tabela 3.4).

	Privação Material 2007	Privação Material 2008
Existência de crianças (não)		
sim	-0.16 (-6.34)	-0.06 (-2.64)
Dimensão do agregado	-0.14 (-10.02)	-0.18 (-12.75)
Área de urbanização (pouco povoada)		
muito povoada	-0.69 (-56.25)	0.37 (15.99)
intermédia	-0.67 (-56.56)	0.46 (20.03)

Tabela 3.4: Efeito das variáveis explicativas (valores dos testes t) sobre a privação material medida por três dimensões. A *bold* estão os valores que se mostraram significativos

3.3 Modelos com Equações Estruturais

3.3.1 Modelos com Equações Estruturais com Efeitos Cruzados e Desfasados

A componente estrutural de um modelo com equações estruturais especifica um conjunto de dependências entre as variáveis latentes de interesse no modelo (Bollen [18]). Um modelo com equações estruturais com efeitos cruzados e desfasados pode ser visto como uma aplicação dos conceitos dos modelos autoregressivos aos modelos com equações estruturais com dados longitudinais. Assim, considerem-se duas variáveis medidas nos momentos t e $t-1$. Cada variável no momento t é explicada pelo seu valor no momento $t-1$ (efeito desfasado) e pelo valor da outra variável no momento $t-1$ (efeito cruzado) (Finkel [48]).

A figura 3.5 apresenta um digrama de um modelo com equações estruturais com efeitos cruzados e desfasados com dois processos (y e w). As variáveis y_1, y_2, y_3, y_4 correspondem a quatro medidas repetidas de um processo (representado pela variável y). As variáveis w_1, w_2, w_3, w_4 correspondem às medidas repetidas do outro processo (representado pela variável w), em cada momento temporal. O modelo pode ser representado pelo seguinte sistema de equações, com $t = 2, 3, 4$ e $i = 1, 2, \dots, N$

$$y_{it} = \beta_{y_t y_{t-1}} y_{it-1} + \beta_{y_t w_{t-1}} w_{it-1} + \varepsilon_{y_{it}}, \quad (3.16)$$

$$w_{it} = \beta_{w_t y_{t-1}} y_{it-1} + \beta_{w_t w_{t-1}} w_{it-1} + \varepsilon_{w_{it}}, \quad (3.17)$$

onde $\beta_{y_t w_{t-1}}$ e $\beta_{w_t y_{t-1}}$ representam os efeitos cruzados e $\beta_{y_t y_{t-1}}$ e $\beta_{w_t w_{t-1}}$ representam os efeitos desfasados. Os termos residuais em cada processo (ε_{y_t} e ε_{w_t}) têm média zero, não são auto correlacionados e não são correlacionados com as variáveis observadas. Dado que os dois processos são modelados simultaneamente, em cada momento temporal, ε_{y_t} e ε_{w_t} podem estar correlacionados. Estas covariâncias são representadas por $\psi_{w_t y_t}$ e são parâmetros do modelo (tal como as variâncias, $\psi_{w_t w_t}$ e $\psi_{y_t y_t}$). Uma distribuição normal multivariada é assumida para as variáveis residuais. Com o propósito de estimar o modelo, a estrutura de variância-covariância de y_t e w_t implícita pelo modelo estimado é igualada à estrutura de variância-covariância observada para as quatro medidas repetidas dos dois processos de interesse.

Os efeitos cruzados e desfasados fornecem informação acerca da estrutura de dependência das variáveis de interesse, nos diferentes momentos temporais. Todavia, estas relações apenas refletem a mudança para o conjunto dos indivíduos, não fornecendo qualquer informação sobre mudanças individuais. Para aceder a mudanças individuais devem ser considerados modelos com trajetória latente (McArdle [88]).

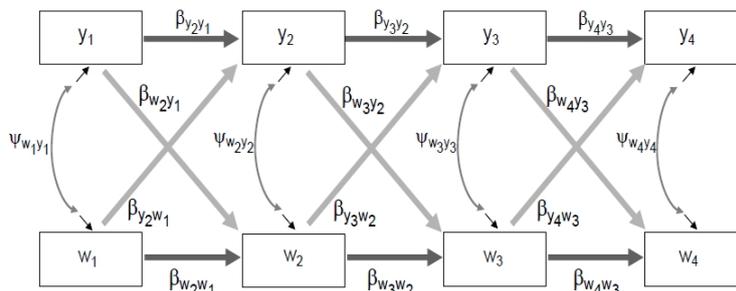


Figura 3.5: Diagrama de um modelo com equações estruturais com efeitos cruzados e desfasados

3.3.2 Resultados do estudo da Privação Material utilizando um Modelo com Equações Estruturais com Efeitos Cruzados e Desfasados

Neste estudo o objetivo é modelar longitudinalmente a evolução da privação material de 2004 a 2007, utilizando um modelo com equações estruturais com efeitos cruzados e desfasados. Por forma, a aferir sobre a privação material, duas diferentes dimensões deste conceito foram consideradas: capacidade financeira e posse de bens duradouros. Os *scores* para cada uma das dimensões foram calculados como somas de itens. Assim, a variável capacidade financeira toma valores entre 0 e 5, enquanto a variável posse de bens duradouros assume valores entre 0 e 4.

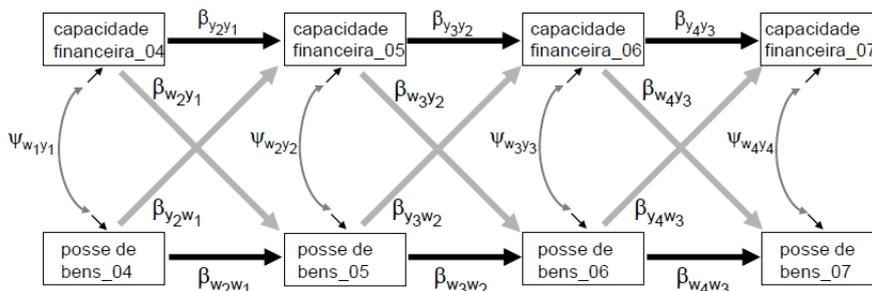


Figura 3.6: Diagrama de um modelo com equações estruturais com efeitos cruzados e desfasados

Os resultados obtidos para o modelo com equações estruturais com efeitos cruzados e desfasados (representados na figura 3.6) mostram que o modelo proposto se ajusta bem aos dados: CFI = 0.992, TLI = 0.992 e WRMR = 0.950. A tabela 3.5 apresenta as estimativas (e os respetivos valores dos testes t) para os coeficientes de regressão numa solução estandardizada (na qual a variância das variáveis é fixa a um para que a magnitude os efeitos possa ser comparável). A maioria dos efeitos cruzados e desfasados mostraram-se significativos, sendo possível concluir que a capacidade financeira e a posse de bens duradouros, em cada momento temporal, é influenciada por si própria e pela outra variável, no momento temporal imediatamente anterior. De salientar o fato de que

	Capacidade Financeira 05 (y_2)	Posse de Bens 05 (w_2)
Capacidade Financeira 04 (y_1)	0.841 (55.980)	0.146 (7.517)
Posse de Bens 04 (w_1)	0.081 (3.901)	0.816 (71.352)
	Capacidade Financeira 06 (y_3)	Posse de Bens 06 (w_3)
Capacidade Financeira 05 (y_2)	0.815 (44.192)	0.026(2.480)
Posse de Bens 05 (w_2)	0.090 (3.476)	0.956 (147.190)
	Capacidade Financeira 07 (y_4)	Posse de Bens 07 (w_4)
Capacidade Financeira 06 (y_3)	0.769 (40.339)	0.062 (3.416)
Posse de Bens 06 (w_3)	0.135 (4.980)	0.939 (86.219)

Tabela 3.5: Estimativas (valores dos testes t), numa solução estandarizada, para os parâmetros de um modelo com equações estruturais com efeitos cruzados e desfasados. A *bold* estão os valores que se mostraram significativos, a um nível de 5%.

os efeitos desfasados são sempre superiores aos efeitos cruzados. O único efeito que não se mostrou significativo é o efeito da capacidade financeira em 2005 na posse de bens em 2006.

3.4 Modelos com Trajetória Latente

3.4.1 Modelos com Trajetória Latente de Primeira Ordem

Os modelos com trajetória latente constituem uma técnica frequentemente utilizada no estudo da mudança usando dados longitudinais, dado que este tipo de modelação, através da estimação de uma trajetória latente ao longo do tempo, permite estudar a mudança quer individualmente, quer para o conjunto dos indivíduos (Preacher et al. [117], Duncan et al. [34]). Esta trajetória que é latente e não observada é estimada a partir da estrutura de médias e de variâncias-covariâncias entre as medidas repetidas das variáveis observadas (y).

O modelo pode ser descrito pela seguinte equação de trajetória, para o elemento $i = 1, \dots, N$ e no momento temporal $t = 1, \dots, T$

$$y_{it} = \alpha_i + \lambda_t \beta_i + \varepsilon_{it}, \quad (3.18)$$

em que, y_{it} é o valor da variável observada y para o elemento i, no momento t, α_i é o intercepto aleatório do elemento i, β_i é o declive aleatório do elemento i, ε_{it} representa o termo residual da trajetória traçada para o elemento i no momento t, e λ_t é uma constante que assume a forma $(t - 1)$ quando é considerada uma trajetória linear. O termo residual ε_t tem distribuição normal com média zero e matriz de variâncias-covariâncias

Θ_ε , isto é, $\varepsilon_t \sim N(\mathbf{0}, \Theta_\varepsilon)$. Os termos residuais não estão correlacionados entre si, isto é, $Cov(\varepsilon_{it}, \varepsilon_{it+s}) = 0$, com $s \neq 0$, conseqüentemente, a matriz Θ_ε é uma matriz diagonal, com elementos θ_{ε_t} na diagonal principal.

Os pesos fatoriais que ligam as variáveis observadas (y_{it}) ao intercepto da trajetória (α_i) são fixos a um por forma a estabelecer o momento inicial. Os pesos fatoriais que ligam as variáveis observadas ao declive da trajetória (β_i) estão habitualmente fixos a 0, 1, 2, 3, ... e refletem a passagem linear do tempo (i.e. $\lambda_t = t - 1$) (figura 3.7).

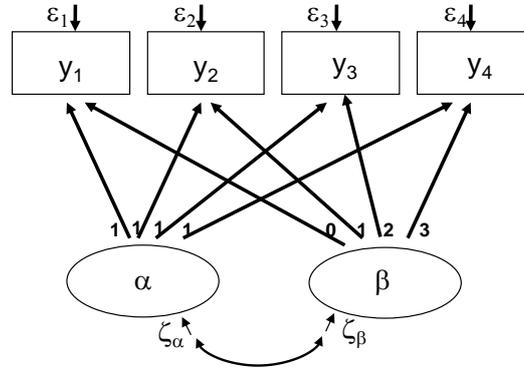


Figura 3.7: Diagrama de um modelo com trajetória latente

As variáveis latentes do modelo α e β são, respetivamente, o interceto e o declive aleatórios e são descritas pelas equações

$$\alpha_i = \mu_\alpha + \zeta_{\alpha_i} \quad (3.19)$$

$$\beta_i = \mu_\beta + \zeta_{\beta_i}, \quad (3.20)$$

em que, μ_α e μ_β representam a média do interceto e do declive, respetivamente. Os interceptos e os declives não estão correlacionados entre elementos e entre si, isto é, $Cov(\alpha_i, \alpha_j) = 0$, $Cov(\beta_i, \beta_j) = 0$ e $Cov(\alpha_i, \beta_j) = 0$ com $i \neq j$. Os termos residuais ζ_{α_i} e ζ_{β_i} que representam a variabilidade no momento inicial e em torno da taxa de crescimento, respetivamente, têm distribuição normal com média zero e não estão correlacionados com ε_{it} . As variâncias dos termos residuais ζ_{α_i} e ζ_{β_i} são dadas por $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, e a covariância por $\psi_{\alpha\beta}$. Assim, $\zeta_{\alpha_i} \sim N(0, \psi_{\alpha\alpha})$, $\zeta_{\beta_i} \sim N(0, \psi_{\beta\beta})$, e ainda, $Cov(\varepsilon_{it}, \zeta_{\alpha_i}) = 0$ e $Cov(\varepsilon_{it}, \zeta_{\beta_i}) = 0$. De salientar que a variância do interceto (α) é equivalente à variância de ζ_{α_i} ou $\psi_{\alpha\alpha}$, e a variância do declive (β) é equivalente à variância de ζ_{β_i} ou $\psi_{\beta\beta}$.

O modelo pode, ainda, ser descrito pela equação reduzida do modelo de trajetória

$$y_{it} = (\mu_\alpha + \lambda_t \mu_\beta) + (\zeta_{\alpha_i} + \lambda_t \zeta_{\beta_i} + \varepsilon_{it}), \quad (3.21)$$

que resulta da substituição das equações 5.2 e 5.3 na equação 3.18, obtendo-se deste modo a trajetória de y_{it} como função do interceto médio, da tendência vezes o declive

médio e de uma parcela de resíduo compósito. O primeiro termo da equação 3.21 diz-se a componente fixa do modelo e representa a estrutura de médias, sendo o segundo termo a componente aleatória que reflete as várias fontes de variabilidade individual.

Os parâmetros de interesse neste tipo de modelo são as médias e as variâncias do intercepto e do declive dessa trajetória, permitindo avaliar a média inicial e a taxa média de mudança para cada elemento e para o conjunto dos elementos, assim como a variabilidade dos elementos em torno da média. Também, a covariância entre os factores latentes (intercepto e declive), que mede a relação entre a média inicial e a taxa média de mudança, é um parâmetro de interesse (Bollen e Curran [21]).

A representação matricial do modelo

A estrutura tradicional dos modelos com equações estruturais pode ser utilizada para representar matricialmente o modelo com trajetória latente. Assim, \mathbf{Y} pode ser escrito como

$$\mathbf{Y} = \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad (3.22)$$

em que, \mathbf{Y} é o vetor ($T \times 1$) das medidas repetidas, $\mathbf{\Lambda}$ é a matriz ($T \times l$) dos pesos fatoriais, $\boldsymbol{\eta}$ é o vetor ($l \times 1$) dos l factores latentes, e $\boldsymbol{\varepsilon}$ é o vetor ($T \times 1$) dos resíduos. Assim, a equação 3.22, para um modelo com trajetória linear ajustado a T medidas repetidas apresenta a seguinte forma

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \dots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \dots & \dots \\ 1 & T-1 \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \dots \\ \varepsilon_{iT} \end{bmatrix}. \quad (3.23)$$

O vetor $\boldsymbol{\eta}$ pode ser definido como

$$\boldsymbol{\eta} = \boldsymbol{\mu}_\eta + \boldsymbol{\zeta}, \quad (3.24)$$

onde $\boldsymbol{\mu}_\eta$ é o vetor ($l \times 1$) das médias dos factores latentes da trajetória e $\boldsymbol{\zeta}$ é o vetor ($l \times 1$) dos resíduos, isto é,

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix} + \begin{bmatrix} \zeta_{\alpha_i} \\ \zeta_{\beta_i} \end{bmatrix}. \quad (3.25)$$

Então a forma reduzida do modelo com trajetória latente pode ser obtida substituindo a equação 3.24 na equação 3.22, resultando

$$\mathbf{Y} = \mathbf{\Lambda}(\boldsymbol{\mu}_\eta + \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}. \quad (3.26)$$

As variâncias e covariâncias implícitas ao modelo na forma reduzida são dadas por

$$\Sigma = \Lambda \Psi \Lambda^T + \Theta_\varepsilon, \quad (3.27)$$

onde Σ é a matriz de variâncias-covariâncias dos y e Θ_ε representa a estrutura de variâncias-covariâncias das parcelas do resíduo para as T medidas repetidas, isto é,

$$\begin{bmatrix} Var(\varepsilon_1) & 0 & 0 & \cdots & 0 \\ 0 & Var(\varepsilon_2) & 0 & \cdots & 0 \\ 0 & 0 & Var(\varepsilon_3) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & Var(\varepsilon_T) \end{bmatrix}. \quad (3.28)$$

De salientar que, apesar de se assumir que os resíduos não estão correlacionados é possível permitir alguma correlação entre os termos residuais das medidas repetidas, isto enquanto o modelo se mantiver identificado.

A matriz Ψ representa a matriz de variâncias-covariância dos resíduos ζ entre os fatores aleatórios do modelo e é dada por

$$\Psi = \begin{bmatrix} \psi_{\alpha\alpha} & \psi_{\alpha\beta} \\ \psi_{\alpha\beta} & \psi_{\beta\beta} \end{bmatrix}. \quad (3.29)$$

A variância de η é igual à variância de ζ (isto é, $\Sigma_{\eta\eta} = \Psi$), se o modelo não contempla a existência de variáveis explicativas da trajetória.

O valor esperado da forma reduzida do modelo de trajetória é

$$E(\mathbf{Y}) = \Lambda \boldsymbol{\mu}_\eta, \quad (3.30)$$

onde $\boldsymbol{\mu}_\eta$ representa o vetor ($l \times 1$) das médias dos fatores e Λ é definida como anteriormente.

A identificação do modelo

Uma questão essencial a ter em consideração antes de proceder à estimação dos parâmetros de interesse do modelo com trajetória latente é a sua identificação. Assim, atendendo a que no capítulo 5 irá ser apresentado um estudo de simulação, a questão da identificação do modelo com trajetória latente é aqui discutida com detalhe.

Num modelo deste tipo são conhecidas as médias, as variâncias e as covariâncias das medidas repetidas das variáveis observadas y , respetivamente, $E(y_{it})$, $Var(y_{it})$ e

$Cov(y_{it}, y_{it-s})$ com $s > 0$ e $t - s \geq 1$. Os parâmetros a estimar (parâmetros desconhecidos) são as médias do intercepto e do declive (μ_α e μ_β , respetivamente), a variância dos resíduos ($Var(\varepsilon_{it})$), a variância do intercepto e do declive e ainda a covariância entre eles ($\psi_{\alpha\alpha}$, $\psi_{\beta\beta}$ e $\psi_{\alpha\beta}$), respetivamente. O modelo diz-se identificado, se for possível expressar os parâmetros desconhecidos do modelo de forma única, como função das médias, variâncias e covariâncias das variáveis observadas.

Considerando a equação 3.18 e assumindo que os parâmetros do modelo se mantêm constantes para o conjunto dos elementos (podendo desaparecer o índice i), obtém-se o valor esperado das variáveis observadas em função dos parâmetros do modelo, dado por

$$E(y_t) = \mu_\alpha + \lambda_t \mu_\beta. \quad (3.31)$$

Sendo o modelo com trajetória latente linear ($\lambda_t = (t - 1)$), e com quatro momentos temporais, é possível escrever

$$E(y_1) = \mu_\alpha \quad (3.32)$$

$$E(y_2) = \mu_\alpha + \mu_\beta \quad (3.33)$$

$$E(y_3) = \mu_\alpha + 2\mu_\beta \quad (3.34)$$

$$E(y_4) = \mu_\alpha + 3\mu_\beta. \quad (3.35)$$

A partir destas equações, a identificação das médias dos fatores latentes (intercepto e declive) podem ser estabelecidas

$$\mu_\alpha = E(y_1) \quad (3.36)$$

$$\mu_\beta = E(y_2) - E(y_1). \quad (3.37)$$

A variância das variáveis observadas é dada por

$$Var(y_t) = Var(\alpha) + \lambda_t^2 Var(\beta) + 2\lambda_t Cov(\alpha, \beta) + Var(\varepsilon_t) \quad (3.38)$$

isto é,

$$Var(y_t) = \psi_{\alpha\alpha} + \lambda_t^2 \psi_{\beta\beta} + 2\lambda_t \psi_{\alpha\beta} + \theta_{\varepsilon_t} \quad (3.39)$$

ou de forma mais detalhada,

$$Var(y_1) = \psi_{\alpha\alpha} + \theta_{\varepsilon_1} \quad (3.40)$$

$$Var(y_2) = \psi_{\alpha\alpha} + \psi_{\beta\beta} + 2\psi_{\alpha\beta} + \theta_{\varepsilon_2} \quad (3.41)$$

$$Var(y_3) = \psi_{\alpha\alpha} + 4\psi_{\beta\beta} + 4\psi_{\alpha\beta} + \theta_{\varepsilon_3} \quad (3.42)$$

$$Var(y_4) = \psi_{\alpha\alpha} + 9\psi_{\beta\beta} + 6\psi_{\alpha\beta} + \theta_{\varepsilon_4}. \quad (3.43)$$

A covariância é dada, de forma genérica, por

$$Cov(y_t, y_{t-s}) = \psi_{\alpha\alpha} + \lambda_t \lambda_{t-s} \psi_{\beta\beta} + (\lambda_t + \lambda_{t-s}) \psi_{\alpha\beta} \quad (3.44)$$

para $s \neq 0$. Isto é,

$$Cov(y_2, y_1) = \psi_{\alpha\alpha} + \psi_{\alpha\beta} \quad (3.45)$$

$$Cov(y_3, y_1) = \psi_{\alpha\alpha} + 2\psi_{\alpha\beta} \quad (3.46)$$

$$Cov(y_4, y_1) = \psi_{\alpha\alpha} + 3\psi_{\alpha\beta} \quad (3.47)$$

$$Cov(y_3, y_2) = \psi_{\alpha\alpha} + 2\psi_{\beta\beta} + 3\psi_{\alpha\beta} \quad (3.48)$$

$$Cov(y_4, y_2) = \psi_{\alpha\alpha} + 3\psi_{\beta\beta} + 4\psi_{\alpha\beta} \quad (3.49)$$

$$Cov(y_4, y_3) = \psi_{\alpha\alpha} + 6\psi_{\beta\beta} + 5\psi_{\alpha\beta}. \quad (3.50)$$

Assim, as variâncias do intercepto e do declive e a covariância entre eles podem ser identificadas a partir das equações acima descritas, do seguinte modo

$$\psi_{\alpha\beta} = Cov(y_1, y_3) - Cov(y_1, y_2) \quad (3.51)$$

$$\psi_{\alpha\alpha} = 2Cov(y_1, y_2) - Cov(y_1, y_3) \quad (3.52)$$

$$\psi_{\beta\beta} = [Cov(y_2, y_3) - Cov(y_1, y_4)]/2. \quad (3.53)$$

A seguinte expressão permite identificar a variância do termo residual

$$\theta_{\varepsilon_t} = Var(y_t) - \psi_{\alpha\alpha} - \lambda_t^2 \psi_{\beta\beta} - 2\lambda_t \psi_{\alpha\beta}. \quad (3.54)$$

De referir que não é possível a identificação e conseqüentemente a estimação de um modelo com trajetória latente só com dois momentos temporais, uma vez que, existem num modelo deste tipo 5 parâmetros conhecidos e 7 desconhecidos. Considerando um modelo com três momentos temporais, o número de parâmetros conhecidos é 9 (3 médias, 3 variâncias e 3 covariâncias), enquanto que o número de parâmetros desconhecidos é 8 (médias e variâncias do intercepto e declive, a covariância entre intercepto e declive e as 3 variâncias do

erro), este pode ser estimado com 1 grau de liberdade. No modelo com quatro momentos temporais, tal como representado na figura 3.7, existem para as variáveis observadas, 4 médias, 4 variâncias e 6 covariâncias, num total de 14 parâmetros conhecidos, existindo 9 parâmetros para os quais é necessária identificação, a média e a variância do intercepto e do declive, a covariância entre o intercepto e o declive e as variâncias dos termos residuais (um em cada momento temporal, t). Deste modo, o modelo diz-se sobre identificado, sendo 5 o número de graus de liberdade na estimação deste. Assim, verifica-se que a sobre identificação do modelo aumenta com o número de momentos temporais deste.

A existência de variáveis ordinais

As equações apresentadas para o modelo com trajetória latente assumem que as variáveis observadas são contínuas e normalmente distribuídas. Assim, se as variáveis observadas são ordinais, como acontece neste estudo, um modelo auxiliar de *thresholds* (τ) ligando a variável ordinal observada (y_{it}) a uma variável latente métrica (y_{it}^*) que se supõe subjacente, deve ser definido do seguinte modo

$$y_{it} = c \quad \text{quando} \quad \tau_{c-1} < y_{it}^* \leq \tau_c, \quad (3.55)$$

onde $c = 1, 2, \dots, C$ é o número total de categorias ordenadas, τ_{c-1} e τ_c são os *thresholds* inferior e superior da categoria c , com $\tau_0 = -\infty$ e $\tau_C = +\infty$. Os valores dos $c - 1$ *thresholds* são ordenados de forma crescente.

Consequentemente, a equação do modelo com trajetória latente com variáveis manifestas ordinais toma a seguinte forma

$$y_{it}^* = \alpha_i + \lambda_t \beta_i + \varepsilon_{it}, \quad (3.56)$$

substituindo y_{it} por y_{it}^* na equação 3.18 (ver figura 3.8 para a representação de um modelo deste tipo).

Este procedimento, que relaciona y_{it} e y_{it}^* , permite calcular correlações, ditas correlações policóricas, que são calculadas a partir das variáveis contínuas subjacentes às variáveis ordinais observadas, face aos *thresholds* definidos.

De acordo com Bollen e Curran [21], quando as variáveis são do tipo ordinal com um número elevado de categorias e apresentam uma distribuição aproximada à normal, tratar estas variáveis como métricas não é uma solução completamente desadequada. No entanto, será sempre desejável ter em consideração a verdadeira natureza das variáveis.

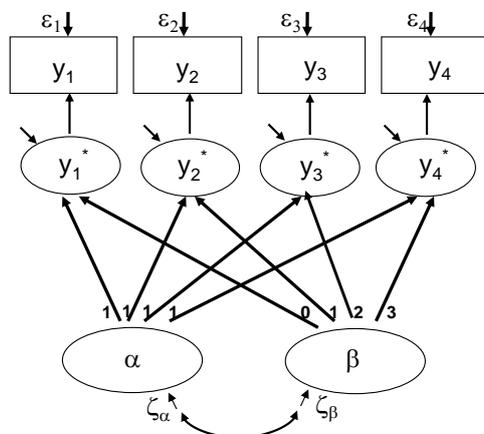


Figura 3.8: Diagrama de um modelo com trajetória latente com variáveis y_t ordinais, sendo y_{it}^* as variáveis latentes contínuas subjacentes

Modelos com Trajetória Latente Bivariada

Por vezes o objetivo é modelar conjuntamente trajetórias para as medidas repetidas de duas (ou mais) variáveis, tornando-se necessário considerar modelos com trajetória latente multivariada. A figura 3.9 apresenta o diagrama de um modelo com trajetória latente bivariada, cujo objetivo é modelar conjuntamente a trajetória de crescimento de y e w . As variáveis latentes α^y e β^y representam o intercepto aleatório e o declive aleatório para o processo y e são medidos pelas medidas repetidas das variáveis observadas y . Analogamente para o processo w .

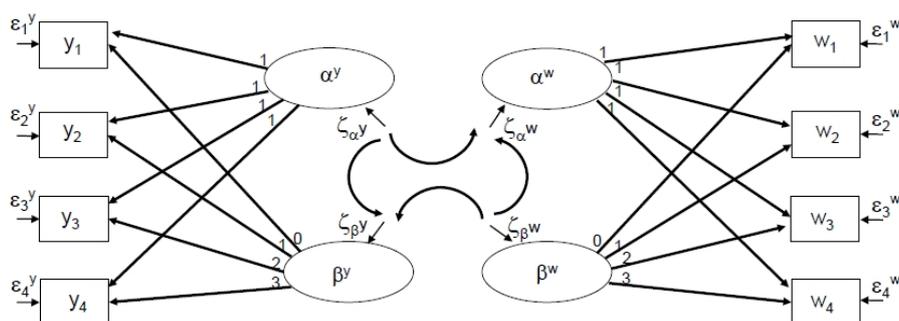


Figura 3.9: Diagrama de um modelo com trajetória latente bivariada para dois processos y e w

De modo geral, para o processo y o valor para o elemento $i = 1, \dots, N$ no momento $t = 1, \dots, T$ é dado por

$$y_{it} = \alpha_i^y + \lambda_t \beta_i^y + \varepsilon_{it}^y, \quad (3.57)$$

e para o processo w o valor para o elemento i no momento t é dado por

$$w_{it} = \alpha_i^w + \lambda_t \beta_i^w + \varepsilon_{it}^w, \quad (3.58)$$

onde $\lambda_t = t - 1$ no caso de ser considerada uma trajetória linear (em consequência os pesos fatoriais associados com o declive tomam os valores 0, 1, 2 e 3). As variâncias dos termos residuais em cada processo são parâmetros no modelo e fazem parte da seguinte matriz

$$\Theta_{\varepsilon\varepsilon} = \begin{bmatrix} \Theta_{\varepsilon_y \varepsilon_y} & \Theta_{\varepsilon_y \varepsilon_w} \\ \Theta_{\varepsilon_w \varepsilon_y} & \Theta_{\varepsilon_w \varepsilon_w} \end{bmatrix} \quad (3.59)$$

Esta matriz é uma matriz diagonal se os termos residuais de cada processo, y e w , não se encontram correlacionados.

Os parâmetros α_i^y e α_i^w , os interceitos aleatórios para o elemento i no processo y ou no processo w , são dados por

$$\alpha_i^y = \mu_\alpha^y + \zeta_{\alpha_i}^y \quad (3.60)$$

e

$$\alpha_i^w = \mu_\alpha^w + \zeta_{\alpha_i}^w, \quad (3.61)$$

e os parâmetros β_i^y e β_i^w , os declives aleatórios nos processos y e w , para os elementos i são dados por

$$\beta_i^y = \mu_\beta^y + \zeta_{\beta_i}^y \quad (3.62)$$

e

$$\beta_i^w = \mu_\beta^w + \zeta_{\beta_i}^w, \quad (3.63)$$

onde, μ_α^* e μ_β^* são as médias do intercepto e do declive (em cada processo, y ou w). $\zeta_{\alpha_i}^*$ e $\zeta_{\beta_i}^*$ são termos da perturbação, com média zero e não correlacionados com ε_{it} (em cada processo, y ou w), e representam a variabilidade entre indivíduos em relação à média

global.

A estrutura de variâncias-covariâncias entre o intercepto e o declive nos dois processos é dada pela seguinte matriz

$$\Psi_{yw} = \begin{bmatrix} \psi_{\alpha_y \alpha_y} & & & & \\ \psi_{\alpha_y \beta_y} & \psi_{\beta_y \beta_y} & & & \\ \psi_{\alpha_y \alpha_w} & \psi_{\alpha_w \beta_y} & \psi_{\alpha_w \alpha_w} & & \\ \psi_{\alpha_y \beta_w} & \psi_{\beta_w \beta_y} & \psi_{\alpha_w \beta_w} & \psi_{\beta_w \beta_w} & \end{bmatrix} \quad (3.64)$$

em que, $\psi_{\alpha_y \alpha_y}$ representa a variância do intercepto, $\psi_{\alpha_y \beta_y}$ é a covariância entre o intercepto e o declive e $\psi_{\beta_y \beta_y}$ é a variância do declive, no processo y . Os parâmetros para o processo w são representados de forma análoga mas têm índice w . $\psi_{\alpha_y \alpha_w}$, $\psi_{\beta_w \beta_y}$, $\psi_{\alpha_w \beta_y}$ e $\psi_{\alpha_y \beta_w}$ representam a covariância entre os interceptos, entre os declives e entre os declives e interceptos dos dois processos, respetivamente. Na especificação do modelo são apenas consideradas as relações de interesse para a investigação em curso.

As equações apresentadas assumem que as variáveis observadas são contínuas e normalmente distribuídas. Para variáveis observadas ordinais assume-se o procedimento subjacente às equações 3.55 e 3.56, anteriormente apresentado.

3.4.2 Resultados do estudo de Privação Material utilizando um Modelo com Trajetória Latente Bivariada

Neste estudo o objetivo é modelar longitudinalmente a evolução da privação material de 2004 a 2007, utilizando um modelo com trajetória latente bivariada. Por forma, a aferir sobre a privação material, duas diferentes dimensões deste conceito foram consideradas: capacidade financeira e posse de bens duradouros. Os *scores* para cada uma das dimensões foram calculados como somas de itens. Assim, a variável capacidade financeira toma valores entre 0 e 5, enquanto a variável posse de bens duradouros assume valores entre 0 e 4.

Os resultados da estimação de um modelo com trajetória latente bivariada (figura 3.10) mostram um ajustamento adequado do modelo, tendo sido obtidos os seguintes valores: CFI=0.991, TLI=0.999 e WRMR=0.568. A tabela 3.6 apresenta as estimativas (e os correspondentes valores dos testes t dentro de parênteses) para as médias, variâncias e covariâncias dos parâmetros aleatórios do modelo.

Assim, da tabela 3.6 é possível concluir que em ambos os processos, capacidade financeira e posse de bens, a média do declive não é significativa, sugerindo que a privação material média não varia significativamente entre 2004 e 2007. As variâncias estimadas para o intercepto e para o declive são ambas significativas, permitindo concluir que os agregados variam, na sua trajetória de mudança ao longo do tempo, bem como no momento inicial (ano de 2004). A covariância entre os interceptos dos dois processos é

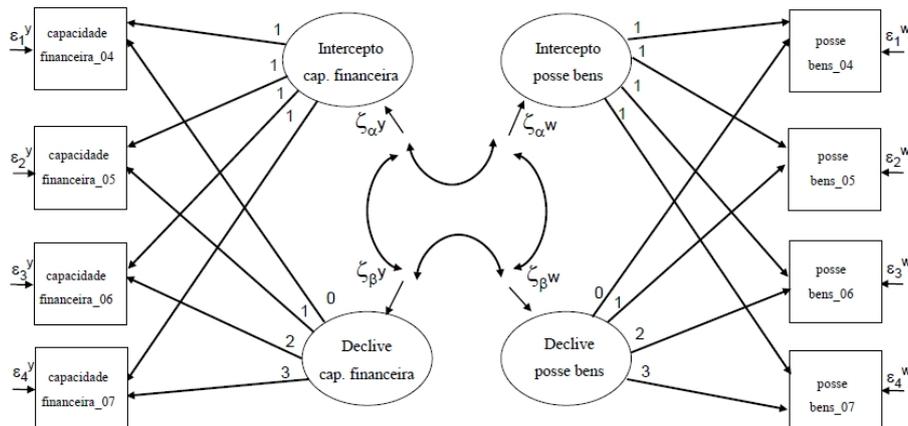


Figura 3.10: Diagrama de um modelo com trajetória latente com dois processos: capacidade financeira e posse de bens duradouros

	Capacidade Financeira FC	Posse de Bens PD
Média do declive	-0.012(-1.367)	0(-0.025)
Variância do intercepto	0.902 (61.733)	0.968 (205.156)
Variância do declive	0.029 (7.905)	0.005 (4.124)
Cov.InterceptoCF/InterceptoPB		0.550 (22.521)
Cov.DecliveCF/DeclivePB		-0.004(-1.282)
Cov.InterceptoCF/DecliveCF		-0.057 (-6.834)
Cov.InterceptoPB/DeclivePB		-0.008(-1.107)

Tabela 3.6: Estimativas (valores do teste t) para os parâmetros aleatórios do modelo com trajetória latente bivariada. A *bold* estão os valores que se mostraram significativos, a um nível de 5%.

significativa, mostrando que o nível inicial de capacidade financeira está associado com o nível inicial de posse de bens. A covariância entre o intercepto da capacidade financeira e o declive da capacidade financeira é também significativa e negativa, indicando que níveis mais elevados de capacidade financeira em 2004 estão relacionados a taxas médias de mudança da capacidade financeira mais lentas ao longo do período de 2004 a 2007.

3.4.3 Modelos com Trajetória Latente de Segunda Ordem

Sempre que o objetivo não é modelar a trajetória das variáveis observadas, mas sim a trajetória de uma variável latente medida por múltiplos indicadores, deve ser considerado um modelo com trajetória latente de segunda ordem. Neste modelo (figura 3.11) existe uma variável latente η medida, em diferentes momentos temporais, por um conjunto de vários indicadores y_j . Habitualmente, existem os mesmos indicadores j , e a mesma variável latente em diferentes momentos temporais, constituindo cada uma destas variáveis um modelo de medida da análise fatorial confirmatória. Os valores dos parâmetros podem

ser considerados iguais em momentos temporais diferentes, caso que é designado por invariância do modelo de medida ao longo do tempo.

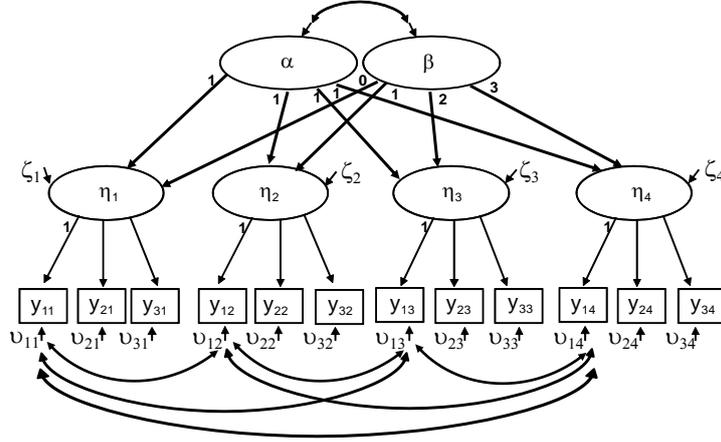


Figura 3.11: Diagrama de um modelo com trajetória latente de segunda ordem, em que a variável latente η é medida por três indicadores, em quatro momentos temporais. Os termos residuais das medidas repetidas do primeiro indicador encontram-se correlacionados

Na componente de medida do modelo, o valor do indicador j , para o elemento i , no momento t , é dada por

$$y_{ijt} = \nu_{jt} + \Lambda_{jt}\eta_{it} + v_{ijt}, \quad (3.65)$$

em que, ν_{jt} é o intercepto do indicador j no momento t , Λ_{jt} corresponde ao peso fatorial do indicador j no momento t , η_{it} é a variável latente para o elemento i no momento t e v_{ijt} é a perturbação do elemento i , no momento t , para o indicador j . O termo residual v_{ijt} tem média zero e não está correlacionado com os outros termos da equação 3.65 (η_{it} , α_i , β_i e λ_t). Além disso, é assumido que este tem uma distribuição normal e a estrutura da matriz de variâncias-covariâncias depende dos termos que se permitem correlacionar. Aliás, de acordo com Bollen e Curran [21] é usual permitir a correlação entre os termos residuais da mesma variável observada em diferentes momentos temporais, sempre que são consideradas no modelo em análise medidas repetidas de uma variável ao longo do tempo. Por forma a não sobrecarregar o diagrama, na figura 3.11 apenas estão representadas as correlações associadas aos termos residuais das medidas repetidas do primeiro indicador, como exemplo.

A equação da trajetória da variável latente η_t é dada por

$$\eta_{it} = \alpha_i + \beta_i\lambda_t + \zeta_{it}, \quad (3.66)$$

em que, α_i representa o intercepto para o elemento i , β_i representa o declive para o ele-

mento i , λ_t define a trajetória em função dos momentos temporais (ver secção anterior) e ζ_{it} representa a perturbação do elemento i , no momento t , que tem média zero, variância Ψ_t e que não está correlacionado com os outros termos da equação 3.66 (α_i , β_i e λ_t). De notar que, λ_t e Λ_{jt} são parâmetros diferentes.

Tal como anteriormente, o intercepto e o declive (os efeitos aleatórios do modelo) são dados pelas equações seguintes

$$\alpha_i = \mu_\alpha + \zeta_{\alpha_i}, \quad (3.67)$$

$$\beta_i = \mu_\beta + \zeta_{\beta_i}, \quad (3.68)$$

em que, μ_α e μ_β são a média do intercepto e do declive para todos os elementos. O modelo assume que os termos residuais ζ_{α_i} e ζ_{β_i} têm distribuição normal com média zero e variância dada por $\Psi_{\alpha\alpha}$ e $\Psi_{\beta\beta}$, respetivamente. Estes termos residuais não estão correlacionados com ζ_{it} e λ_t .

Analogamente ao modelo com trajetória latente de primeira ordem, os parâmetros de interesse são as médias e as variâncias dos efeitos aleatórios, bem como a covariância entre estes efeitos. Além destes, também a estrutura dos pesos fatoriais do modelo de medida (que relaciona a variável latente com os indicadores de medida) é considerada importante. De salientar ainda que a média do intercepto deixa de ser um parâmetro de interesse no modelo, sendo colocada a zero.

No caso de as variáveis observadas serem ordinais assume-se o procedimento subjacente às equações 3.55 e 3.56, anteriormente apresentado.

Modelos com Trajetória Latente de Segunda Ordem Condicionada

Nos modelos com trajetória latente não condicionada é apenas tomado em consideração o efeito do tempo como variável explicativa, sendo o objetivo descrever a trajetória. Todavia, quando se pretende considerar no estudo possíveis variáveis explicativas da trajetória do modelo, este diz-se um modelo com trajetória latente condicionada. Estas variáveis podem ser consideradas como invariantes ou variantes no tempo, sendo a sua introdução no modelo realizada de forma diferente. Por outro lado, o modelo com trajetória latente considerado pode ser de primeira ou segunda. A figura 3.12 apresenta um modelo com trajetória latente de segunda ordem com uma variável explicativa invariante no tempo x e uma variável variante no tempo w_t .

Assim, se a variável que se pensa ser um possível determinante da trajetória se mantém constante ao longo do tempo (variável invariante no tempo) então é estimado o seu efeito sobre o intercepto e o declive da trajetória. O intercepto e o declive têm, agora, por equação, respetivamente

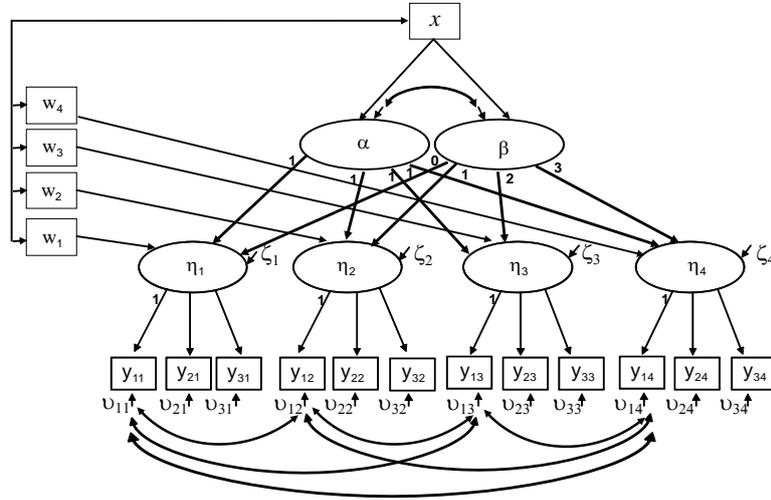


Figura 3.12: Diagrama de um modelo com trajetória latente de segunda ordem condicionada, com variável explicativa invariante (x) e variável explicativa variante no tempo (w_t)

$$\alpha_i = \mu_\alpha + \gamma_\alpha x_i + \zeta_{\alpha_i}, \quad (3.69)$$

$$\beta_i = \mu_\beta + \gamma_\beta x_i + \zeta_{\beta_i}, \quad (3.70)$$

em que, x é a variável explicativa invariante no tempo e γ_α e γ_β os seus coeficientes de regressão. Os restantes termos têm o significado já apresentado anteriormente.

Se a variável explicativa não se mantém constante com a passagem do tempo, então o interesse é agora o efeito das suas medidas repetidas nas medidas repetidas da variável latente η para a qual se modela a trajetória, isto é, a equação da trajetória da variável latente η_t , para o indivíduo i , é dada por

$$\eta_{it} = \alpha_i + \lambda_t \beta_i + \gamma_t w_{it} + \zeta_{it}, \quad (3.71)$$

em que, w_t é a variável explicativa variante no tempo e γ_t são os coeficientes de regressão. Para o significado dos restantes termos ver modelo anterior.

De salientar que o modelo com variáveis explicativas invariantes é mais parcimonioso que o modelo com variáveis explicativas variantes no tempo, sendo os parâmetros de interesse no primeiro caso os coeficientes de regressão dos efeitos aleatórios sobre a variável x (γ_α e γ_β), e no segundo caso os coeficientes de regressão das medidas repetidas de η_t sobre as medidas repetidas da variável explicativa w_t (os γ_t). Todavia, considerar uma variável explicativa como invariante no tempo, sendo ela variante, resulta em problemas de especificação do modelo e um provável enviesamento nos resultados (Bollen e Curran [21]).

3.4.4 Resultados do estudo da Privação Material utilizando um Modelo com Trajetória Latente de Segunda Ordem com Trajetórias Condicionadas

O objetivo deste estudo é modelar longitudinalmente o conceito de privação material recorrendo a um modelo com trajetória latente de segunda ordem. A variável latente privação material é considerada medida por três indicadores (constrangimentos económicos, posse de bens duradouros e condições da habitação), em que cada um deles corresponde a uma soma de itens (*score*), em quatro momentos temporais. Deste modo, a variável constrangimentos económicos toma valores numa escala de 0 a 5 itens em privação, a posse de bens toma valores entre 0 a 5 itens em falta e a variável condições da habitação numa escala de 0 a 3 itens em falta. Estas variáveis foram consideradas como ordinais, tendo sido definido um modelo de *thresholds*, em que variáveis latentes contínuas estão subjacentes às variáveis observadas ordinais. Uma vez descrita a trajetória da privação material no período de 2007 a 2010, são consideradas como variáveis explicativas da privação material e da sua trajetória, o rendimento do agregado, a dimensão do agregado e a área de urbanização.

Numa primeira etapa, um modelo de análise fatorial confirmatória (AFC) é considerado com o propósito de estimar, em cada momento temporal, 2007 a 2010, um modelo de medida para a privação material, ver figura C.1, em anexo. As três variáveis que correspondem às diferentes dimensões da privação material foram consideradas indicadores de cada modelo de medida. Na tabela C.1, em anexo, pode ser observado que os pesos fatoriais, em todos os momentos temporais, para os constrangimentos económicos e posse de bens tomam valores superiores a 0.6, valor recomendado como aceitável na literatura. Quanto aos pesos fatoriais para as condições da habitação assumem valores no limite do aceitável, entre 0.501 e 0.533, exceto para o ano de 2010 em que o valor obtido é de 0.355, bastante inferior a 0.6. De salientar ainda que os constrangimentos económicos no ano de 2010 apresentam um peso fatorial bastante elevado, 0.844.

Numa segunda etapa um modelo com trajetória latente de segunda ordem, figura 3.13, é estimado e as medidas de ajustamento obtidas mostram um ajustamento modelo-dados razoável. Os valores obtidos foram: CFI=TLI=0.998, RMSEA=0.022 e WRMR=0.689. Neste modelo foi permitido correlacionar as medidas repetidas da mesma variável, nos diferentes momentos temporais.

A tabela 3.7 apresenta as estimativas para a média, variâncias e covariância entre intercepto e declive da curva obtidas para o modelo com trajetória latente de segunda ordem que foi considerado para a privação material. De sublinhar que num modelo com trajetória latente de segunda ordem no qual se pretende modelar a trajetória de uma variável latente medida por múltiplos indicadores, a média do intercepto toma o valor zero (porque a média de η_1 é nula) e não é um parâmetro do modelo.

A média do declive é significativa e negativa, sugerindo que a privação material das

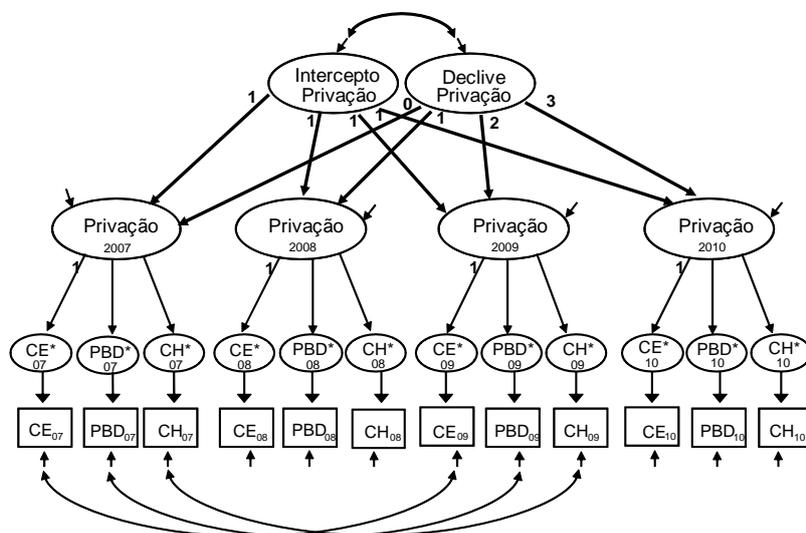


Figura 3.13: Diagrama de um modelo com trajetória latente de segunda ordem da privação material medida por 3 dimensões consideradas como variáveis ordinais

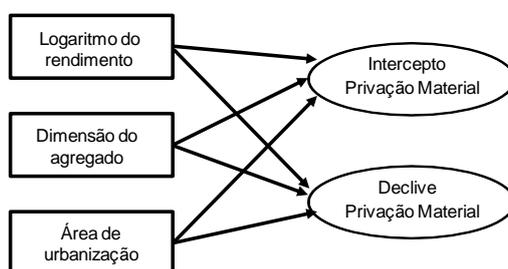


Figura 3.14: Diagrama de um modelo com trajetória latente de segunda ordem condicionada para a privação material. As variáveis explicativas são consideradas como não variantes no tempo

famílias diminui entre 2007 e 2010. A variância estimada para o intercepto é significativa, logo as famílias diferem na privação material que experimentam no momento inicial, ano de 2007. A variância estimada do declive não se mostrou significativa, o que quer dizer que não existem diferenças na taxa média de mudança da privação material para o conjunto das famílias. O valor estimado para a covariância entre o intercepto e o declive não se mostrou significativo, não sendo possível dizer que existe relação entre a privação material no ano de 2007 e a taxa média de mudança de 2007 a 2010.

De seguida foram consideradas variáveis explicativas da trajetória de mudança da privação material, tais como, (logaritmo do) rendimento, dimensão do agregado e área de urbanização. Como a variável área de urbanização é definida em três categorias, foram criadas duas variáveis *dummy*, sendo a categoria de referência área pouco povoada. Pro-

	Privação Material	
	intercepto	declive
Média	—	-0.018 (-2.073)
Variância	0.589 (8.607)	0.007 (1.300)
Covariância	-0.018 (-1.572)	

Tabela 3.7: Estimativas (valores dos testes t) da média, variâncias e covariância do intercepto e declive para um modelo com trajetória latente de segunda ordem para a privação material. A *bold* estão os valores que se mostraram significativos

cedimento idêntico foi considerado para a variável dimensão do agregado, para a qual foi calculada uma variável *dummy* com duas categorias, três ou menos indivíduos (categoria de referência) e mais de três indivíduos. A variável logaritmo do rendimento foi considerada como contínua. O logaritmo da variável rendimento é considerado por forma a atenuar a heterogeneidade dos dados. Com o objetivo de estudar o efeito destas três variáveis na trajetória latente de segunda ordem para a privação material estas foram introduzidas no modelo como invariantes no tempo (ver figura 3.14), sendo apresentados na tabela 3.8 os valores estimados para os coeficientes de regressão dos efeitos aleatórios nas 3 variáveis explicativas.

A análise desta tabela permite observar que o impacto do (logaritmo do) rendimento, da dimensão do agregado e da área de urbanização do local de residência do agregado, no intercepto da trajetória latente da privação material se mostrou significativo. Assim, é possível dizer que agregados com rendimentos mais elevados experimentam uma menor privação material em 2007, isto porque o valor estimado é negativo. Por outro lado, um coeficiente de regressão estimado positivo para a variável dimensão do agregado, permite dizer que agregados com mais de três indivíduos têm maior privação em 2007, face a agregados com três ou menos indivíduos. Do mesmo modo, agregados que no ano de 2007 residem em áreas muito povoadas ou intermédias (grandes ou médias cidades) experimentam maior privação material face a agregados que habitam em áreas pouco povoadas (aldeias). Quanto ao efeito das três covariáveis na taxa média de mudança da privação material, ao longo do período entre 2007 a 2010, este não se mostrou significativo.

Alternativamente, foi considerado um outro modelo em que a variável (logaritmo do) rendimento foi modelada como variante no tempo, figura 3.15, com o objetivo de explicar a variabilidade observada na privação material, no período de 2007 a 2010. As variáveis área de urbanização e dimensão do agregado foram introduzidas no modelo como invariantes, porque apresentam valores análogos ao longo do período em questão (2007 a 2010).

A análise da tabela 3.9 permite ver que o logaritmo do rendimento tem um impacto significativo e negativo na privação material em cada um dos anos, 2007, 2008 e 2010, significando que valores mais elevados de rendimento implicam menor privação material nestes três anos. A dimensão do agregado e a área de urbanização apresentam um impacto positivo e significativo no intercepto da trajetória da privação material, embora no declive da trajetória esse efeito não se tenha mostrado significativo. Deste modo, pode

	Privação Material	
	intercepto	declive
Logaritmo do rendimento	-0.831 (-17.802)	0.014(0.469)
Dimensão do agregado (três ou menos indivíduos) mais de três indivíduos	0.355 (4.541)	0.037(1.377)
Área de urbanização (pouco povoada) muito povoada intermédia	0.339 (4.343) 0.190 (2.393)	-0.025(-0.992) 0.006(0.217)

Tabela 3.8: Estimativas (valores dos testes t) dos coeficientes de regressão das variáveis explicativas do intercepto e do declive para um modelo com trajetória latente de segunda ordem condicionada. A *bold* estão os valores que se mostraram significativos

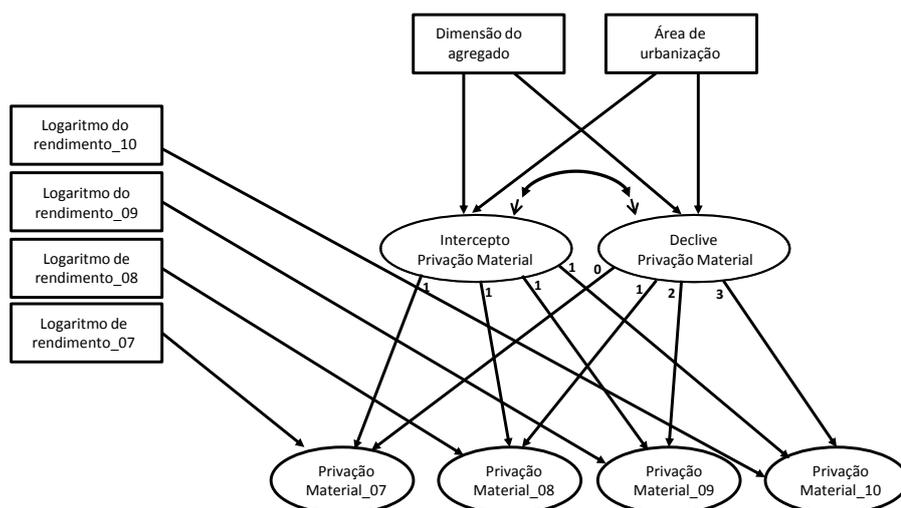


Figura 3.15: Diagrama de um modelo com trajetória latente de segunda ordem condicionada para a privação material. A variável explicativa da trajetória, logaritmo do rendimento, é considerada como variante no tempo. As variáveis dimensão do agregado e área de urbanização são consideradas invariantes

concluir-se que agregados de maior dimensão (mais de três elementos) apresentam níveis de privação mais elevados em 2007, relativamente a agregados menores (três ou menos elementos). Agregados residentes em áreas muito povoadas ou intermédias apresentam, no ano de 2007, uma privação material superior face a agregados que habitam em áreas pouco povoadas.

A comparação dos resultados obtidos na modelação da trajetória latente de segunda

CAPÍTULO 3. MODELAÇÕES COM DADOS COMPLETOS

	Privação Material			
	2007	2008	2009	2010
Logaritmo do rendimento	-0.526 (-6.288)	-0.293 (-2.875)	-0.006(-0.058)	-0.642 (-5.383)
	intercepto		declive	
Dimensão do agregado (três ou menos indivíduos) mais de três indivíduos	0.394 (4.962)		0.046(1.363)	
Área de urbanização (pouco povoada)	0.342 (4.369)		-0.027(-1.057)	
muito povoada	0.188 (2.392)		0.001(0.031)	
intermédia				

Tabela 3.9: Estimativas (valores do teste t) dos coeficientes de regressão da variável explicativa (logaritmo do) rendimento do fator latente privação material, e dos coeficientes de regressão das variáveis explicativas dimensão do agregado e área de urbanização do intercepto e do declive, para um modelo com trajetória latente de segunda ordem condicionada. A *bold* estão os valores que se mostraram significativos

ordem da privação material considerando a variável explicativa (logaritmo do) rendimento como invariante ou variante no tempo permite conclusões distintas, no primeiro caso é possível dizer que agregados com rendimento mais elevado experimentam um nível de privação material mais baixo no ano de 2007, não sendo possível concluir que o (logaritmo do) rendimento influencie a taxa média de mudança da privação material, enquanto que no segundo caso se conclui que agregados com (logaritmo do) rendimento mais elevado apresentam valores de privação material mais baixos nos anos de 2007, 2008 e 2010. De referir ainda que, o modelo no qual a variável explicativa é introduzida como invariante no tempo é mais parcimonioso.

Capítulo 4

Modelações considerando a existência de dados omissos

Alguns autores descreveram a existência de dados omissos como um dos problemas estatísticos e de planeamento mais importantes na investigação (Azar [7]), tendo sido frequente durante vários anos a exclusão das observações com não resposta na modelação de dados conforme referido em Peugh e Enders [115]. Este procedimento é resultado da inexistência, até há alguns anos atrás, de procedimentos específicos para lidar com a problemática das não respostas por parte dos pacotes estatísticos disponíveis no mercado.

A existência de dados omissos é um problema bastante comum em inquéritos, sendo praticamente impossível que este problema não ocorra se o inquérito em estudo é um painel, principalmente devido ao abandono de participantes. Por outro lado, o painel ICOR devido ao seu esquema de rotatividade, só permite uma análise longitudinal com dados completos, considerando apenas 25% dos agregados (1/4 da amostra). Neste contexto, a utilização de abordagens estatísticas para lidar com dados omissos apresenta-se de grande importância.

4.1 Omissões por desenho do estudo

Dados omissos ou incompletos representam um problema bastante comum em estudos nos mais variados domínios, constatando-se que em dados longitudinais a existência de dados omissos constitui a regra e não a exceção, Graham [53].

As omissões que podem ser usualmente encontradas são de diferentes tipos e a sua classificação depende do tipo de dados em presença. Assim, num estudo transversal é usual designar as omissões por *item non response* ou *unit non response*. Num caso de *item non response* em determinado momento temporal um indivíduo não responde a uma das questões de interesse no inquérito; se não responde a todas está-se perante *unit non response*. Em estudos longitudinais é usual fazer a distinção entre dados omissos intermitentes ou *dropout* (também designado por atrito). Ocorre *dropout* quando a partir

forma	A	B	C	D	E	F	G
1	✓	-	-	-	-	-	-
2	-	✓	-	-	-	-	-
3	-	-	✓	-	-	-	-
4	-	-	-	✓	-	-	-
5	-	-	-	-	✓	-	-
6	-	-	-	-	-	✓	-
7	-	-	-	-	-	-	✓

Tabela 4.1: *Multiple matrix design* (✓ = questão colocada; - = questão não colocada; A a G dizem respeito a diferentes grupos de questões)

de determinado momento temporal um indivíduo deixar de responder às questões do inquérito. Caso contrário as omissões dizem-se intermitentes. Todavia, as omissões podem ainda resultar do desenho do estudo: um *planned missing design*, ou podem ainda resultar da entrada tardia do indivíduo para o estudo (caso em que foi selecionado para entrar no inquérito em determinado momento temporal, mas só responde num momento temporal posterior).

Num *planned missing design* os dados omissos ocorrem de acordo com a vontade do investigador, sendo que o objectivo da utilização de um desenho deste tipo prende-se com o propósito de aumentar a qualidade dos dados evitando o esforço de inquirição (Graham et al. [55], Graham et al. [56], McArdle [87]). Todavia, alguns investigadores encaram este tipo de omissão com algum ceticismo, manifestando relutância na implementação desta estratégia. Esta atitude deriva da presunção de que os dados em falta podem influenciar de forma negativa os resultados da análise (Enders [41]).

Os investigadores deparam-se frequentemente com dificuldades no número de perguntas de um questionário que possam ser expetáveis de serem respondidas pelos inquiridos, o que se agrava quando o estudo é longitudinal, caso em que os inquiridos respondem a um grande número de perguntas em várias ocasiões. Esta questão pode ser ultrapassada limitando o número de variáveis ou introduzindo um *planned missing design*. A utilização de um *planned missing design* permite que as perguntas de um questionário sejam distribuídas em diferentes grupos e a cada inquirido seja apresentado um subconjunto desses grupos. Esta estratégia permite recolher dados do conjunto total de perguntas do questionário enquanto se reduz o esforço de inquirição.

Uma revisão da literatura permite encontrar um conjunto alargado de aplicações de vários tipos de *planned missing design*. Na forma mais elementar surgem os *matrix sampling design* (Schoemaker [146] e Thayer [154]), os quais na sua versão mais simples (*multiple matrix design*) consistem na alocação de diferentes variáveis a diferentes respondentes, tal como ilustrado na tabela 4.1.

Basicamente, neste tipo de planeamento existem subgrupos de variáveis criados aleato-

riamente, os quais também de forma aleatória são alocados a subgrupos de unidades. O principal interesse dos investigadores na utilização deste planeamento prende-se com a redução do tempo exigido aos inquiridos na resposta aos inquéritos, procurando assim aumentar a taxa de resposta.

De acordo com Raghunathan e Grizzle [120], os estudos utilizando este tipo de planeamento têm-se limitado à estimação de médias populacionais, dado que a afetação aleatória de subgrupos de variáveis a subgrupos de indivíduos induz frequentemente problemas de estimação, os quais são devidos à não identificação e à limitação de dados para análise multivariada.

Em 1994, McArdle [87] sugere o *fractional block design*, consistindo numa variação do *matrix sampling design*, no qual, para além da existência de informação para todas as variáveis, existe ainda informação para alguns pares de variáveis. Assim, mais do que permitir a estimação da média populacional para todas as variáveis, é ainda possível ao investigador estimar correlações para alguns dos pares de variáveis, Allison [2], Muthén et al. [104].

McArdle [87] no seu estudo sugere o ajustamento de modelos de análise factorial (uma análise multigrupos) para lidar com este desenho e acomodar a existência de dados omissos. Johnson [73] sugere o *balanced incomplete blocks spiral design*, abordagem que assegura a estimação da correlação entre todos os pares de variáveis, bem como da média de todas as variáveis. Este planeamento apresentado por Johnson [73] foi aplicado ao questionário NAEP (*National Assessment of Educational Progress*), um questionário levado a cabo nos Estados Unidos com o objetivo de medir o conhecimento e aptidões dos estudantes em várias áreas do conhecimento, como por exemplo, a matemática ou a leitura, que permite que o mesmo número de indivíduos responda a cada conjunto de itens e a cada par de itens o mesmo número de vezes.

Os planeamentos anteriores constituem exemplos clássicos de dados omissos por desenho do estudo, mas que não necessitam obrigatoriamente de técnicas de análise específicas para este tipo de dados (abordagens para lidar com as omissões existentes), devido às questões que os estudos que os utilizaram se propuseram responder.

Uma variante do *matrix sampling design* proposta por Graham et al. [55], [56] e que tem sido essencialmente usada na área da psicologia, é o *3-form design*, conforme apresentado na tabela 4.2. Este tipo de *design* teve igualmente na base do seu desenvolvimento o objetivo de limitar o tempo necessário para que os inquiridos respondessem aos inquéritos, permitindo alargar o número de questões colocadas. Este tipo de planeamento permite estimar todas as correlações, bem como todas as médias e variâncias. As variáveis no conjunto X são essenciais ao problema em estudo, conseqüentemente são colocadas a todos os inquiridos.

Ainda em Graham et al. [55] é mostrado que a utilização de métodos contemporâneos

Forma	X	A	B	C
1	✓	✓	✓	-
2	✓	✓	-	✓
3	✓	-	✓	✓

Tabela 4.2: *3-form design* (✓ = questão colocada; - = questão não colocada)

Forma	X	A	B	C	D	E
1	✓	✓	✓	-	-	-
2	✓	✓	-	✓	-	-
3	✓	✓	-	-	✓	-
4	✓	✓	-	-	-	✓
5	✓	-	✓	✓	-	-
6	✓	-	✓	-	✓	-
7	✓	-	✓	-	-	✓
8	✓	-	-	✓	✓	-
9	✓	-	-	✓	-	✓
10	✓	-	-	-	✓	✓
11	✓	✓	✓	✓	✓	✓

Tabela 4.3: *Split questionnaire design* (✓ = questão colocada; - = questão não colocada)

para lidar com dados omissos tais como, métodos da máxima verossimilhança e imputação múltipla, possibilitam ao investigador analisar os dados sem excluir os casos incompletos, uma vez que estas técnicas permitem obter estimativas não enviesadas dos parâmetros e erros padrão razoáveis.

Na área do marketing Raghunathan e Grizzle [120] sugerem uma variação do *matrix sampling design* designada por *split questionnaire design*. O *split questionnaire design*, consiste em 11 formas diferentes, uma das quais contendo a resposta a todas as variáveis do inquérito. No *split questionnaire design* (tabela 4.3), para além de uma componente central com questões que são consideradas importantes (por exemplo, questões sociodemográficas), é colocado a cada entrevistado de igual modo apenas uma selecção de outras variáveis. Este método embora baseado no *matrix sampling design*, imputa os dados em falta com o propósito de obter um conjunto de dados completo. Esta abordagem exige apenas que as combinações de variáveis a avaliar conjuntamente tenham origem na mesma subamostra (para evitar problemas de estimação devido à não-identificação). Assim, dependendo do *split design*, qualquer análise que se pretenda pode ser realizada mantendo o tamanho original da amostra.

Estudos como o de Kamakura e Wedel [78], Rässler et al. [123], Thomas et al. [156], entre outros, sugerem o uso de técnicas de imputação múltipla para lidar com a problemática das omissões nos *split questionnaire design*.

cohort	9	10	11	12	13	14	15	16
9	✓	✓	✓	✓	-	-	-	-
11	-	-	✓	✓	✓	✓	-	-
13	-	-	-	-	✓	✓	✓	✓

Tabela 4.4: *Cohort-sequential design* (✓ = questão colocada; - = questão não colocada)

Embora o *split questionnaire design* seja uma técnica muito utilizada na área do marketing existem outras que embora menos populares são igualmente utilizadas e que de algum modo estão relacionadas, tais como, *data fusion*, *subsampling* e *time sampling design* e para as quais também a utilização de procedimentos de imputação múltipla tem sido considerados (Bertchold e Jeannin [17], Grover e Vriens [59], Kamakura e Wedel [78], Vriens et al. [163]).

Por outro lado, em processos de desenvolvimento que envolvem mudanças relacionadas com a idade, é usual a utilização de um *planned missing data design* denominado por *cohort-sequential design* (Nesselroad e Bates [108]).

A ideia subjacente a um *cohort-sequential design* é combinar algumas bases de dados longitudinais de reduzida dimensão numa única base de dados longitudinal. Isto é feito obtendo amostras de diferentes *cohorts* (grupos) de idade no primeiro momento temporal e seguindo cada *cohort* (grupo) ao longo do mesmo período. A tabela 4.4 apresenta um exemplo de *cohort-sequential design* utilizado num estudo sobre o consumo de álcool durante 4 anos, Duncan et al. [35], [37]. Neste estudo são selecionados inicialmente indivíduos de 9, 11 e 13 anos, os quais são seguidos em quatro momentos temporais. Note-se que cada *cohort* da idade apresenta quatro momentos temporais omissos intencionalmente, por exemplo, o da idade de 9 anos tem dados omissos nas idades 13, 14, 15 e 16. O método da máxima verosimilhança para dados omissos permite combinar dados de múltiplos *cohorts* num simples grupo de dados, de forma a permitir examinar a mudança ao longo de um período que excede o número de momentos temporais em que os dados são observados.

Este tipo de planeamento apresenta uma limitação, ao existirem pares de variáveis que são totalmente omissos, como por exemplo, idades 9 e 13, o que embora não comprometa o ajustamento de um modelo com trajetória latente (para o estudo da mudança), condiciona a estimação de correlações. Apesar de apresentar esta limitação o *cohort-sequential design* tem sido muito utilizado em trabalhos na área da psicologia, nomeadamente para examinar mudanças relacionadas com a idade, McArdle e Anderson [89], McArdle e Hamagami [91], Duncan et al. [36]. A abordagem estatística seguida para lidar com o problema das omissões no desenho do estudo consiste no método da máxima verosimilhança (*full information maximum likelihood*).

A análise destes diferentes planeamentos que apresentam desenhos com padrões de omissão, que em alguns casos são bastante semelhantes ao do painel rotativo ICOR e que têm sido utilizados em domínios como a psicologia e o marketing, permitem concluir que o

método da máxima verossimilhança (*full information maximum likelihood*) e a imputação múltipla são as abordagens estatísticas adequadas no tratamento das omissões resultantes da rotatividade do painel. De salientar que, o tratamento do painel rotativo como um *cohort-sequential design* não parece ser adequado, uma vez que este tipo de planejamento considera uma mudança de escala, muito embora o desenho do estudo seja semelhante nos dois casos. Num *cohort-sequential design* o objetivo é analisar a mudança para o conjunto de idades dos indivíduos presentes no estudo (dos 9 aos 16 anos), e não analisar a mudança para os indivíduos ao longo do período temporal em que os dados foram recolhidos (4 momentos temporais).

4.2 Mecanismo de omissão dos dados

De acordo com Roth [129] existem dois fatores que devem ser considerados quando se escolhe uma técnica para lidar com dados omissos, a quantidade de não respostas e o mecanismo de omissão de dados. Quanto maior a percentagem de omissões existente nos dados em análise, mais importante é a escolha de uma técnica apropriada para minimizar o enviesamento que pode ocorrer nos resultados, sendo que uma percentagem de não respostas entre 15% e 20%, pode trazer implicações na estimação dos parâmetros do modelo em estudo.

O mecanismo de omissão de dados é um pressuposto que influencia o comportamento das diferentes técnicas para lidar com dados omissos (Allison [3], Enders [41], Little e Rubin [84], Schaffer e Graham [143]), devendo por essa razão ser tido em consideração sempre que se realiza uma análise dos dados.

De acordo com Rubin [131] o mecanismo de omissão de dados descreve relações entre as variáveis observadas e a probabilidade de não resposta. Seja $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ um vetor de respostas, com uma determinada distribuição e \mathbf{R}_i um vetor ($n_i \times 1$) de indicadores de resposta $\mathbf{R}_i = (R_{i1}, \dots, R_{in_i})^T$, com $R_{ij} = 1$ se Y_{ij} é observado e $R_{ij} = 0$ se Y_{ij} é omissos. A distribuição de \mathbf{R} pode estar relacionada com \mathbf{Y} , assim, admitimos um modelo probabilístico para \mathbf{R} , designado por $P(\mathbf{R}|\mathbf{Y}, \Phi)$, que depende de \mathbf{Y} assim como de parâmetros desconhecidos Φ . De acordo com o valor de \mathbf{R}_i , o conjunto de respostas, \mathbf{Y}_i , pode ser particionado em duas componentes, $\mathbf{Y}_i = (\mathbf{Y}_{iobs}, \mathbf{Y}_{imiss})$, correspondendo às respostas observadas e às omissões, respetivamente. A forma como \mathbf{R} está relacionado com \mathbf{Y} permite distinguir três diferentes mecanismos de omissão

- *missing completely at random* (MCAR), se a probabilidade de omissão de uma variável não está relacionada nem com outras variáveis, nem com o valor possível dessa mesma variável, isto é,

$$P(\mathbf{R}|\mathbf{Y}, \Phi) = P(\mathbf{R}|\Phi); \tag{4.1}$$

- *missing at random* (MAR), quando a probabilidade de omissão de uma variável está relacionada com uma outra variável presente na análise, isto é,

$$P(\mathbf{R}|\mathbf{Y}, \Phi) = P(\mathbf{R}|\mathbf{Y}_{obs}, \Phi), \quad (4.2)$$

este mecanismo pressupõe que a omissão está relacionada apenas a valores medidos nos dados, mas não a valores não medidos;

- *missing not at random* (MNAR), se a probabilidade de omissão de uma variável depende do valor que essa mesma variável assumiria, bem como de outra variável, isto é,

$$P(\mathbf{R}|\mathbf{Y}, \Phi) = P(\mathbf{R}|\mathbf{Y}_{obs}, \mathbf{Y}_{miss}, \Phi), \quad (4.3)$$

este também é designado como omissão informativa ou não ignorável.

No caso de dados multivariados, em que mais do que uma variável apresenta não respostas, \mathbf{R} será uma matriz cuja dimensão será igual à da matriz das observações, isto se todas as variáveis apresentarem valores omissos.

Ao contrário do mecanismo MAR, em que a omissão pode ser explicada através de outras variáveis presentes no estudo, um mecanismo MNAR surge devido a um padrão de omissão explicado pelas variáveis que estão ausentes ou então de valores de variáveis que não foram observadas.

Conseqüentemente, os mecanismos MCAR e MAR dizem-se ignoráveis, pois não é necessário estimar os parâmetros que estabelecem a propensão para a não resposta. Por outro lado, o mecanismo MNAR diz-se não ignorável devido à falta de aleatoriedade da não resposta e ignorar esse mecanismo pode levar a sérios viesamentos nos valores estimados. Assim, compreender qual o mecanismo de omissão de dados presente no estudo é um passo indispensável para fazer inferências corretas.

Num mecanismo MCAR é assumido que a probabilidade de omissão na variável Y não está relacionada com outras variáveis observadas e com os próprios valores de Y . Por outras palavras, a não resposta não é sistemática e os dados observados podem ser vistos como uma subamostra aleatória dos hipotéticos dados completos. Por exemplo, quando num estudo sobre métodos de ensino, uma criança a meio do estudo muda para outra escola, o mecanismo pode ser considerado MCAR se a razão para a mudança não está relacionada com outras variáveis do estudo (condições socioeconómicas, problemas disciplinares ou outras variáveis relacionadas com o estudo). Todavia, Graham e Donaldson [54] acrescentam que o mecanismo de omissão de dados pode estar associado a outras variáveis que não estão em análise.

Embora, seja possível pensar que o ICOR sendo um *planned missing design* tem um mecanismo de omissão de dados MCAR (Schaffer e Graham [143]), dado que todos os anos 25% dos agregados que compõem a “amostra mãe” são substituídos por outros da

mesma área, os metodologistas frequentemente argumentam que é um pressuposto muito limitado e que é impossível de ser verificado na prática (Muthén et al. [104], Raghunathan [119]). Assim, as omissões existentes no ICOR, devido ao desenho do inquérito, têm um mecanismo de omissão que pode ser considerado como MAR. Todavia, existem neste inquérito, outras omissões para as quais não é possível assumir este mecanismo, devendo o seu mecanismo de omissão ser considerado MNAR. Por outro lado, de acordo com Enders [42], não existe um teste para o mecanismo MAR, conseqüentemente é impossível descartar completamente que o mecanismo de omissão dos dados em análise seja MNAR.

4.3 Modelos para lidar com omissões MCAR ou MAR

Quando se está perante uma base de dados que apresenta omissões é importante perceber qual o mecanismo que gera a não resposta, dependendo da utilização de uma abordagem estatística adequada a obtenção de estimativas não enviesadas dos parâmetros do modelo. As abordagens estatísticas *full information maximum likelihood* e imputação múltipla revelam-se adequadas quando o mecanismo de omissão de dados é MCAR ou MAR (Allison [3], Baraldi e Enders [10], Schaffer e Graham [143], Schaffer e Olsen [144]). Estas técnicas são consideradas mais vantajosas comparativamente a procedimentos tradicionais que podem ser encontrados na literatura, tais como, técnicas de eliminação das observações omissas (*listwise* e *pairwise deletion*) ou de técnicas que substituem o valor omisso por um único valor (por exemplo, substituição pela média), as quais frequentemente produzem estimativas enviesadas dos parâmetros (Baraldi e Enders [10]).

Verificando-se a existência de observações omissas a estimação de um processo de mudança recorrendo aos modelos com trajetória latente pode ser levada a efeito pelo método *full information maximum likelihood* (FIML). Este método assenta no pressuposto de que a covariância entre variáveis pode ser usada para inferir valores prováveis para as observações omissas (Schaffer e Graham [143]), e será apresentado com maior detalhe no ponto seguinte.

De acordo com Enders [40] existem três diferentes procedimentos de máxima verosimilhança para lidar com dados omissos: i) o algoritmo EM (*expectation-maximization*); ii) uma abordagem multigrupos; e iii) FIML. Por outro lado, num estudo sobre modelações com variáveis latentes considerando a existência de dados omissos, Rovine [130] classifica estes métodos em dois tipos. O primeiro tipo consiste em estimar a matriz de dados completa (algoritmo EM), e de seguida utilizar esta matriz na modelação de uma trajetória latente (Dempster et al. [32]). O segundo tipo diz-se de estimação direta de parâmetros, podendo ser utilizada recorrendo a uma abordagem multigrupos no *framework* dos modelos de equações estruturais (Allison [2]) ou em alternativa, *full information maximum likelihood* (Finkbeiner [49]). Concetualmente, FIML é um procedimento análogo à abordagem multigrupos, excetuando o facto de considerar uma função de máxima verosimilhança para cada observação, ao invés de considerar uma para cada um dos diferentes grupos, sendo que estes grupos correspondem a distintos padrões de omissão.

Embora autores como Lee [83], Muthén et al. [104], Werts et al. [164] tenham considerado uma abordagem multigrupos no âmbito dos modelos com equações estruturais, foi a técnica apresentada por Allison [2] que se tornou bastante popular e é várias vezes referenciada na literatura da especialidade. Enders [40] compara esta técnica a *pairwise deletion*, uma vez que calcula uma função de verosimilhança para cada grupo de não respostas.

Outra das mais recentes técnicas para lidar com dados omissos e permitir a estimação de um modelo com trajetória latente é a imputação múltipla. A imputação múltipla foi proposta por Little e Rubin [84] e tem recebido grande destaque na literatura. Ao contrário do processo tradicional de imputar com um simples valor a não resposta, este método consiste em substituir cada valor omissos por um conjunto de m valores plausíveis, gerando, desta forma, m conjuntos de dados completos. As estimativas das análises desses conjuntos de dados completos são combinadas tendo em conta a incerteza associada com o processo de imputação. A grande vantagem do método de imputação múltipla é a possibilidade de utilizar abordagens estatísticas na modelação dos dados como estando perante dados completos (Newman [109]), contrariamente ao *full information maximum likelihood* que usa toda a informação disponível para estimar os parâmetros e os erros padrão (Schaffer e Graham [143]). Todavia, estas duas abordagens assentam no pressuposto de que o mecanismo de omissão de dados é ignorável e que os dados têm distribuição normal multivariada, embora a imputação múltipla não seja tão sensível quanto uma abordagem de FIML à não verificação deste pressuposto (Allison [3]).

4.3.1 *Full Information Maximum Likelihood*

Verificando-se a existência de observações omissas a estimação de um processo de mudança usando um modelo com trajetória latente pode ser levada a efeito pelo *full information maximum likelihood* (FIML). Neste método é utilizada toda a informação disponível (todos os dados observados) para estimar os parâmetros do modelo e os seus erros padrão, sendo que uns casos contribuem com mais informação do que outros (Arbuckle [5], Enders [41]).

Neste sentido, se os dados estão completos a função de verosimilhança a maximizar é para o elemento i

$$\log L_i = -\frac{k}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{Y}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}), \quad (4.4)$$

onde, k é o número de variáveis, \mathbf{Y}_i é o vetor das observações para o elemento i , $\boldsymbol{\mu}$ é o vetor das médias populacionais e Σ é a matriz das variâncias-covariâncias.

No caso de existirem dados omissos, a função de verosimilhança para o elemento i passa a ser dada pela seguinte expressão

$$\log L_i = -\frac{k_i}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_i| - \frac{1}{2} (\mathbf{Y}_i - \boldsymbol{\mu}_i)^T \Sigma_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i), \quad (4.5)$$

onde, k_i representa, para o elemento i , o número de variáveis com valores completos e $\boldsymbol{\mu}_i$ e Σ_i são o vetor das médias populacionais e a matriz das variâncias-covariâncias, respectivamente, calculados apenas com os dados disponíveis. De salientar que o cálculo da função $\log L_i$ para a observação i depende apenas das variáveis e dos parâmetros para os quais esse elemento tem dados completos (Enders [41]).

A função de verosimilhança final corresponde à soma de N funções de verosimilhança, para os N elementos, sendo dada por

$$\log L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^N \log L_i. \quad (4.6)$$

Quando o mecanismo de omissão dos dados é *missing completely at random* (MCAR) ou *missing at random* (MAR) e os dados têm uma distribuição normal multivariada, o método FIML produz estimativas dos parâmetros, erros padrão e testes estatísticos que são consistentes e eficientes (Peters e Enders [114]).

4.3.2 Resultados do estudo da Privação Material utilizando *Full Information Maximum Likelihood*

Neste trabalho é proposta a modelação através de um modelo com trajetória latente de um indicador de privação material para os anos de 2006 a 2009. Ao longo destes quatro momentos temporais em análise permaneceram no painel um conjunto de 845 agregados, para os quais existe informação completa e que se passa a designar por “caso completo”. Todavia, existe ainda um outro conjunto de 848 agregados que participaram no estudo entre 2006 e 2008 e que não estiveram em 2009 devido ao esquema de rotatividade do painel. Ao conjunto destes dois grupos constituídos por um total de 1693 agregados designa-se por “caso com todas as observações”, no qual existem aproximadamente 50% de omissões na quarto momento temporal (2009).

A utilização de procedimentos de *full information maximum likelihood* permite estimar um modelo com trajetória latente (figura 4.1) para modelar a trajetória do *score* de privação material entre 2006 e 2009, considerando a totalidade dos 1693 agregados, “caso com todas as observações”. Também para os 845 agregados que constituem o “caso completo” é modelada a trajetória do *score* de privação material, no período considerado, recorrendo a um modelo com trajetória latente. O *score* de privação material é calculado como uma soma de itens para cada um dos quatro momentos temporais considerados. Este *score* toma valores entre 0 e 10, indicando o valor 0 que o agregado não se encontra privado e 10 que não consegue assegurar nenhum dos itens considerados.

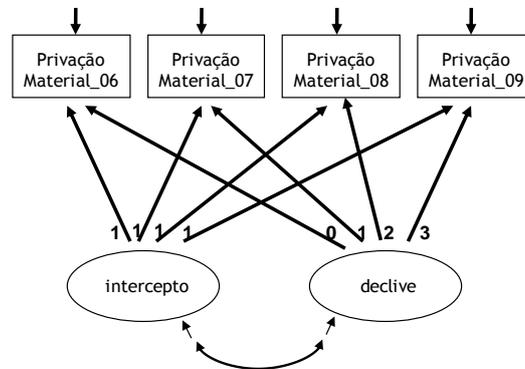


Figura 4.1: Diagrama de um modelo com trajetória latente

Na tabela 4.5 são apresentadas as medidas de qualidade do ajustamento usuais, para o “caso completo” e para o “caso com todas as observações”. Os valores obtidos para as várias medidas consideradas permitem concluir que existe um ajustamento razoável modelo-dados, para ambas as subamostras consideradas. De salientar que o valor do RMSEA para o “caso completo”, se encontra no limite do razoável, enquanto que para o “caso com todas as observações” apresenta um valor claramente superior a 0.08, valor recomendado na literatura da especialidade.

	“caso completo” $N = 845$	“caso com todas as observações” $N = 1693$
RMSEA	0.079	0.125
CFI	0.990	0.968
TLI	0.988	0.961
SRMR	0.025	0.046

Tabela 4.5: Medidas da qualidade do ajustamento (“caso completo” e “caso com todas as observações”)

	“caso completo” $N = 845$	“caso com todas as observações” $N = 1693$
Média do intercepto (μ_α)	2.514	2.500
Média do declive (μ_β)	-0.025	-0.106
Variância do intercepto ($\psi_{\alpha\alpha}$)	3.028	3.014
Variância do declive ($\psi_{\beta\beta}$)	0.123	0.168
Covariância inter/decl ($\psi_{\alpha\beta}$)	-0.293	-0.415

Tabela 4.6: Estimativas dos parâmetros obtidas no ajustamento de um modelo com trajetória latente (“caso completo” e “caso com todas as observações”). A *bold* estão os valores que se mostraram significativos

As estimativas dos parâmetros do modelo com trajetória latente são apresentadas, para o “caso completo” e para “o caso com todas as observações”, na tabela 4.6. Para o conjunto dos 845 agregados que permaneceram no estudo de 2006 a 2009 é possível concluir que em 2006 o valor médio da privação é de 2.514 itens. A taxa média de mudança na privação entre 2006 e 2009, embora negativa, não é significativa. Nem todos os agregados têm o mesmo nível de privação material em 2006 e nem todos mudam da mesma forma. Níveis mais elevados de privação em 2006 estão associados com uma taxa média de mudança mais lenta.

Para o conjunto dos 1 693 agregados considerados, que constituem o “caso com todas as observações”, é possível concluir que em 2006 o valor médio de privação é de 2.500 itens. A taxa média de mudança na privação entre 2006 e 2009 diminui, sendo a média do declive (-0.106) negativa e significativa. Nem todos os agregados têm o mesmo nível de privação material em 2006 e nem todos mudam da mesma forma. Níveis mais elevados de privação em 2006 estão associados com uma taxa média de mudança mais lenta. Embora, a estimativa da média do declive para o caso com todas as observações se tenha mostrado significativa, no caso completo isso não se verifica, permitindo conclusões distintas sobre a taxa média de mudança da privação material entre 2006 e 2009, para as diferentes subamostras consideradas. Assim, são utilizados procedimentos de Monte Carlo para perceber qual das opções metodológicas deve o analista considerar.

Num estudo de simulação utilizando um procedimento de Monte Carlo, os dados são gerados de uma população com determinados valores dos parâmetros, Mooney [98], Smith [149] e Rubinstein [133]. Um grande número de amostras são geradas e um modelo é estimado para cada uma dessas amostras. Os valores dos parâmetros e os erros padrão são médias desses mesmos valores obtidos para as várias amostras geradas. Os seguintes critérios podem ser utilizados na avaliação dos valores estimados: enviesamento relativo das estimativas dos parâmetros; enviesamento relativo do erro padrão; erro quadrático médio; *coverage*; e a potência do teste. Para maior detalhe sobre a realização de um estudo de simulação de Monte Carlo utilizando o *software* estatístico Mplus deve ser consultado o capítulo 5.

Assumindo como valores dos parâmetros populacionais os valores dos parâmetros que foram estimados pelo ajustamento de um modelo com trajetória latente para os dados que constituíam o “caso completo”, são geradas 1 000 amostras cada uma com um padrão de 10%, 20%, 30%, 40% e 50% de omissões no último momento temporal, ano de 2009. Na realidade pretende-se recriar uma amostra que tenha um desenho semelhante à amostra com todos os 1 693 agregados (figura 2.3) e ainda ver o que acontece se diminuir a percentagem de observações omissas nesse momento temporal.

As tabelas 4.7 a 4.11 apresentam os resultados para a geração de 1 000 amostras com 845 observações cada e com um padrão de omissão de 10%, 20%, 30%, 40% e 50% em 2009, sendo possível a partir destas calcular o enviesamento relativo verificado na estimação dos parâmetros e cujos valores são apresentados na tabela 4.12.

CAPÍTULO 4. MODELAÇÕES CONSIDERANDO A EXISTÊNCIA DE DADOS
OMISSOS

De acordo com Muthén e Muthén [106], o enviesamento relativo na estimação de cada parâmetro, θ , pode ser calculado utilizando o valor considerado como parâmetro populacional e a média das estimativas dos parâmetros obtidos nas várias amostras geradas, da seguinte forma

$$B_R(\hat{\theta}) = \frac{E(\hat{\theta}) - \theta}{\theta}. \quad (4.7)$$

Parâmetros	“população”	média	desvio padrão	média erros padrão	proporção de intervalos de confiança a 95%	proporção de réplicas com coeficiente significativo
μ_α	2.514	2.5159	0.0639	0.0627	0.938	1.000
μ_β	-0.025	-0.0256	0.0165	0.0164	0.952	0.351
$\psi_{\alpha\alpha}$	3.028	3.0323	0.1617	0.1661	0.956	1.000
$\psi_{\beta\beta}$	0.123	0.1218	0.0146	0.0150	0.958	1.000
$\psi_{\alpha\beta}$	-0.293	-0.2928	0.0366	0.0375	0.958	1.000

Tabela 4.7: Resultados obtidos no estudo de simulação gerando 1 000 amostras com 10% de omissões no ano de 2009

Parâmetros	“população”	média	desvio padrão	média erros padrão	proporção de intervalos de confiança a 95%	proporção de réplicas com coeficiente significativo
μ_α	2.514	2.5160	0.0639	0.0628	0.938	1.000
μ_β	-0.025	-0.0257	0.0168	0.0168	0.948	0.346
$\psi_{\alpha\alpha}$	3.028	3.0321	0.1619	0.1665	0.956	1.000
$\psi_{\beta\beta}$	0.123	0.1217	0.0153	0.0157	0.954	1.000
$\psi_{\alpha\beta}$	-0.293	-0.2927	0.0378	0.0386	0.956	1.000

Tabela 4.8: Resultados obtidos no estudo de simulação gerando 1 000 amostras com 20% de omissões no ano de 2009

Parâmetros	“população”	média	desvio padrão	média erros padrão	proporção de intervalos de confiança a 95%	proporção de réplicas com coeficiente significativo
μ_α	2.514	2.5159	0.0639	0.0628	0.939	1.000
μ_β	-0.025	-0.0256	0.0172	0.0173	0.948	0.328
$\psi_{\alpha\alpha}$	3.028	3.0323	0.1626	0.1669	0.956	1.000
$\psi_{\beta\beta}$	0.123	0.1218	0.0160	0.0165	0.951	1.000
$\psi_{\alpha\beta}$	-0.293	-0.2928	0.0391	0.0398	0.952	1.000

Tabela 4.9: Resultados obtidos no estudo de simulação gerando 1 000 amostras com 30% de omissões no ano de 2009

CAPÍTULO 4. MODELAÇÕES CONSIDERANDO A EXISTÊNCIA DE DADOS
OMISSOS

Parâmetros	“população”	média	desvio padrão	média erros padrão	proporção de intervalos de confiança a 95%	proporção de réplicas com coeficiente significativo
μ_α	2.514	2.5160	0.0640	0.0628	0.943	1.000
μ_β	-0.025	-0.0257	0.0176	0.0177	0.948	0.316
$\psi_{\alpha\alpha}$	3.028	3.0323	0.1627	0.1673	0.957	1.000
$\psi_{\beta\beta}$	0.123	0.1219	0.0169	0.0175	0.951	1.000
$\psi_{\alpha\beta}$	-0.293	-0.2927	0.0400	0.0412	0.958	1.000

Tabela 4.10: Resultados obtidos no estudo de simulação gerando 1 000 amostras com 40% de omissões no ano de 2009

Parâmetros	“população”	média	desvio padrão	média erros padrão	proporção de intervalos de confiança a 95%	proporção de réplicas com coeficiente significativo
μ_α	2.514	2.5160	0.0640	0.0628	0.938	1.000
μ_β	-0.025	-0.0258	0.0181	0.0183	0.945	0.288
$\psi_{\alpha\alpha}$	3.028	3.0326	0.1635	0.1679	0.955	1.000
$\psi_{\beta\beta}$	0.123	0.1220	0.0182	0.0187	0.950	1.000
$\psi_{\alpha\beta}$	-0.293	-0.2929	0.0415	0.0429	0.961	1.000

Tabela 4.11: Resultados obtidos no estudo de simulação gerando 1 000 amostras com 50% de omissões no ano de 2009

Enviesamento nos Parâmetros	50% omissões	40% omissões	30% omissões	20% omissões	10% omissões
μ_α	0.0008	0.0008	0.0008	0.0008	0.0008
μ_β	0.0320	0.0280	0.0240	0.0280	0.0240
$\psi_{\alpha\alpha}$	0.0015	0.0014	0.0014	0.0014	0.0014
$\psi_{\beta\beta}$	-0.0081	-0.0089	-0.0098	-0.0106	-0.0098
$\psi_{\alpha\beta}$	-0.0003	-0.0010	-0.0007	-0.0010	-0.0007

Tabela 4.12: Enviesamento relativo na estimativa média dos parâmetros quando são geradas 1 000 amostras com 50%, 40%, 30%, 20% e 10% de omissões no ano de 2009

Da análise das tabelas 4.7 a 4.11 é possível concluir que a estimativa média para cada um dos parâmetros de interesse no modelo com trajetória latente apresenta valores análogos, qualquer que seja a percentagem de omissões considerada. Todavia, a média dos erros padrão aumenta à medida que aumenta a percentagem de omissões. Na tabela 4.12 é possível observar que o parâmetro que apresenta maior enviesamento, em qualquer uma das situações consideradas, é a média do declive. Este é o parâmetro para o qual, as conclusões se mostraram contraditórias, quando comparados os resultados do caso completo, com os resultados obtidos para todas as observações.

4.3.3 Imputação Múltipla

Outra das técnicas contemporâneas para lidar com dados omissos, permitindo a estimação de um modelo com trajetória latente é a imputação múltipla. A imputação múltipla consiste em atribuir valores às observações que se encontram omissas, ao contrário do método da máxima verosimilhança que não necessita de imputar ou substituir os valores omissos, usando, alternativamente, toda a informação disponível para estimar os parâmetros e os erros padrão (Schaffer e Graham [143]). Todavia, esta abordagem, tal como o método da máxima verosimilhança, assenta no pressuposto de que o mecanismo de omissão de dados é ignorável e de que os dados seguem distribuição normal multivariada.

Uma abordagem com imputação múltipla consiste em três etapas: uma primeira etapa de imputação, de seguida a etapa da análise e por fim uma etapa em que os resultados obtidos na etapa anterior são combinados (ver figura 4.2). Na fase da imputação são geradas várias réplicas dos dados (m), cada uma das quais contendo um conjunto único de valores imputados. Na fase da análise, para cada uma das (m) réplicas de dados geradas no passo anterior, isto é, para cada um dos conjuntos de dados já completos, é estimado o modelo de interesse (no âmbito do presente trabalho um modelo com trajetória latente). Na última fase, as estimativas dos parâmetros e os erros padrão obtidos (m) no passo anterior são combinadas num único valor.

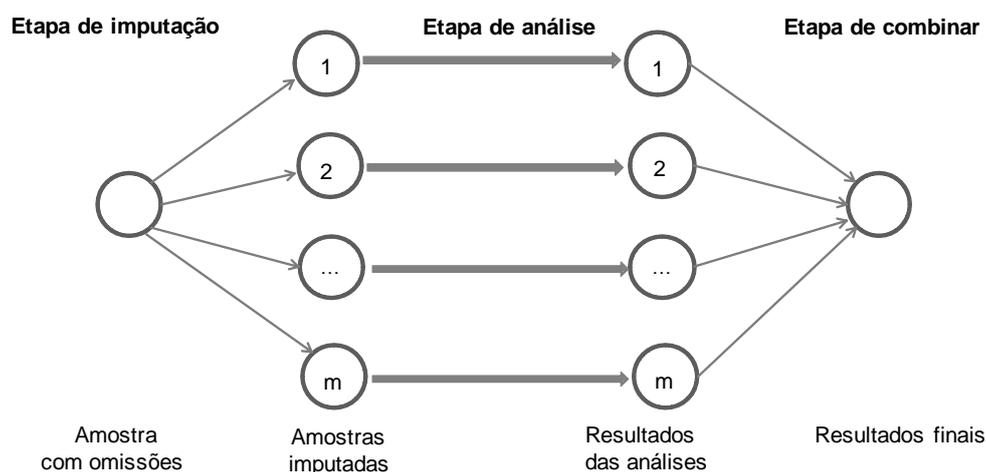


Figura 4.2: Diagrama das três etapas da imputação múltipla

A imputação múltipla, ao contrário do método da máxima verosimilhança, é bastante exigente do ponto de vista computacional, em particular na fase da imputação. Um número considerável de algoritmos diferentes passíveis de utilização nesta primeira fase do método podem ser encontrados na literatura (Lavori et al. [82], Raghunathan et al. [121], Rosenbaum e Rubin [128], Schaffer [142], Tanner e Wong [153]). Todavia, o algoritmo

mais utilizado e que se encontra implementado em grande número de pacotes estatísticos é o *Data Augmentation* (Schaffer [142], Tanner e Wong [153]). Este algoritmo que consiste num processo iterativo em dois passos, é bastante semelhante ao algoritmo EM proposto por Dempster et al. [32]. Os passos E e M determinísticos são substituídos pelos passos, I (passo de imputação) e P (passo de simulação da distribuição *a posteriori*) estocásticos, conforme descrito a seguir.

- **Passo I - Imputação**

No passo I é utilizada uma estimativa do vetor das médias e da matriz de covariâncias para construir um conjunto de equações de regressão que predizem as variáveis incompletas a partir das variáveis observadas. Com este objetivo são constituídos grupos de observações que partilham o mesmo padrão de omissão, sendo que para cada um destes existe uma equação de regressão (ou conjunto de equações de regressão). Esta equação (ou equações) tem um termo residual que se assume com distribuição normal. Quando duas ou mais variáveis se encontram omissas, a distribuição dos resíduos assume-se normal multivariada.

Assim, o passo I pode ser descrito como:

$$\mathbf{Y}_h^* \sim P(\mathbf{Y}_{miss} | \mathbf{Y}_{obs}, \boldsymbol{\theta}_{h-1}^*) \quad (4.8)$$

em que, \mathbf{Y}_h^* representa os valores imputados no passo I na iteração h , \mathbf{Y}_{miss} corresponde aos dados omissos, \mathbf{Y}_{obs} corresponde aos dados observados e $\boldsymbol{\theta}_{h-1}^*$ é o vetor das médias e a matriz das variâncias-covariâncias do passo P anterior (ou seja, iteração $h - 1$). A primeira iteração do passo I utiliza um vetor de médias e uma matriz de covariâncias obtidos pelo método da máxima verosimilhança.

- **Passo P - Simulação da distribuição *a posteriori***

O passo P começa por usar os valores imputados do passo I anterior para estimar o vetor das médias e a matriz de variâncias-covariâncias. De seguida o algoritmo gera um novo conjunto de valores dos parâmetros adicionando um termo residual aleatório a cada elemento $\hat{\boldsymbol{\mu}}$ e $\hat{\boldsymbol{\Sigma}}$, ou seja, um novo vetor de médias e uma nova matriz de variâncias-covariâncias é gerada a partir das respetivas distribuições *a posteriori*. A distribuição *a posteriori* da matriz de covariância é dada por

$$P(\boldsymbol{\Sigma} | \hat{\boldsymbol{\mu}}, \mathbf{Y}) \sim W^{-1}(N - 1, \hat{\boldsymbol{\Lambda}}) \quad (4.9)$$

em que, $\hat{\boldsymbol{\mu}}$ é o vetor das médias amostrais, \mathbf{Y} é a matriz dos dados preenchida do passo I anterior, W^{-1} representa a inversa da distribuição de Wishart, $N - 1$ os graus de liberdade e $\hat{\boldsymbol{\Lambda}}$ é a matriz das somas de quadrados e dos produtos cruzados dos valores amostrais. Utilizando procedimentos de Monte Carlo é gerada a partir da distribuição *a posteriori* uma nova matriz de variâncias-covariâncias, $\boldsymbol{\Sigma}^*$.

O algoritmo usa um procedimento análogo para criar um novo conjunto de médias, sendo a distribuição *a posteriori* do vetor das médias dado por

$$P(\hat{\boldsymbol{\mu}}|\mathbf{Y}, \boldsymbol{\Sigma}) \sim MN(\hat{\boldsymbol{\mu}}, N^{-1}\boldsymbol{\Sigma}^*) \quad (4.10)$$

onde, MN denota uma distribuição normal multivariada, $\hat{\boldsymbol{\mu}}$ é um vetor de médias amostrais e $\boldsymbol{\Sigma}^*$ é a matriz de variâncias-covariâncias simulada. Procedimentos de Monte Carlo geram um novo conjunto de médias a partir da distribuição *a posteriori* e que são designadas por $\hat{\boldsymbol{\mu}}^*$.

Em resumo, o passo P simula o vetor das médias e a matriz de variâncias-covariância da população *a posteriori*, dado a amostra completa, isto é,

$$\boldsymbol{\theta}_h^* \sim P(\boldsymbol{\theta}|\mathbf{Y}_{obs}, \mathbf{Y}_h^*) \quad (4.11)$$

em que, $\boldsymbol{\theta}_h^*$ são os valores dos parâmetros simulados do passo P, \mathbf{Y}_{obs} são os dados observados e \mathbf{Y}_h^* contem os valores imputados do passo I imediatamente anterior.

O objetivo do passo P é gerar estimativas alternativas do vetor da média e da matriz de variâncias-covariâncias para utilizar no passo I seguinte, que irá produzir um novo conjunto de coeficientes de regressão e conseqüentemente um conjunto diferente de imputações. Deste modo, a repetição destes dois passos, I e P, um grande número de vezes cria uma cadeia de Markov com várias cópias de dados, cada uma delas contendo estimativas únicas dos valores omissos, isto é, $(\mathbf{Y}_1^*, \boldsymbol{\theta}_1^*), (\mathbf{Y}_2^*, \boldsymbol{\theta}_2^*), (\mathbf{Y}_3^*, \boldsymbol{\theta}_3^*), \dots, (\mathbf{Y}_h^*, \boldsymbol{\theta}_h^*)$. Esta sucessão converge em distribuição para $P(\mathbf{Y}_{miss}, \boldsymbol{\theta}|\mathbf{Y}_{obs})$. Assumindo que as iterações convergem para uma distribuição estacionária, o objetivo é simular uma extração aproximadamente independente dos valores omissos dessa distribuição.

De salientar que, devido ao procedimento descrito, as amostras preenchidas obtidas numa dada iteração h estão fortemente correlacionadas com as obtidas na iteração $h - 1$, havendo assim a necessidade de ter o cuidado de selecionar para a etapa da análise um número m de amostras tomadas em iterações suficientemente distantes na sequência, por exemplo $\mathbf{Y}_h^*, \mathbf{Y}_{h+a}^*, \mathbf{Y}_{h+2a}^*, \dots$, para um valor grande de a , que podem ser tomadas como imputações de \mathbf{Y}_{miss} . Por outro lado, é necessário perder algumas amostras obtidas nas primeiras iterações do algoritmo até guardar a primeira amostra com valores imputados. Determinar esse número de iterações pode ser complexo, sendo necessário estudar a estabilidade dos parâmetros obtidos no passo P. Os gráficos da função de autocorrelação que são disponibilizados nos pacotes estatísticos que têm implementado procedimentos de imputação múltipla, como, por exemplo o Mplus, são ferramentas indispensáveis nessa análise.

O número de amostras geradas na fase de imputação para posterior análise tem sido alvo de discussão na literatura da área. De acordo com Schaffer [142], o número de

amostras geradas deve ser entre 3 a 5. Todavia abordagens mais recentes sugerem a utilização de pelo menos 20 amostras imputadas (Graham et al. [57]).

Depois de geradas as imputações, cada uma das (m) amostras completas são analisadas com a abordagem estatística desejada, neste caso um modelo com trajetória latente (figura 3.7) é estimado para cada uma das amostras completas. Resultando deste processo um conjunto de estimativas para os parâmetros e erros padrão (em número m), que são combinados num único resultado na terceira etapa utilizando as fórmulas de Rubin.

De acordo com Rubin [132] para qualquer parâmetro deve ser calculada uma média das estimativas obtidas para as várias amostras (m),

$$\bar{\theta} = \frac{1}{m} \sum_{r=1}^m \hat{\theta}_r. \quad (4.12)$$

Todavia, combinar os erros padrão é um pouco mais complexo, uma vez que são consideradas duas fontes de variabilidade amostral: a variância dentro das imputações que é dada por

$$V_W = \frac{1}{m} \sum_{r=1}^m SE_r^2, \quad (4.13)$$

em que SE_r^2 é o quadrado do erro padrão (i.e., variância amostral) para a amostra r , com $r = 1, \dots, m$, e a variância entre imputações dada por

$$V_B = \frac{1}{m-1} \sum_{r=1}^m (\hat{\theta}_r - \bar{\theta})^2. \quad (4.14)$$

Assim, a variância total, que é uma variância combinada, é dada pela seguinte expressão

$$V_T = V_W + V_B + \frac{V_B}{m} \quad (4.15)$$

e conseqüentemente o erro padrão é calculado como $SE = \sqrt{V_T}$.

4.3.4 Resultados do estudo utilizando *Full Information Maximum Likelihood* e Imputação Múltipla

Neste estudo foram consideradas sete diferentes subamostras da amostra inicial que corresponde à base de dados longitudinal de 2006 a 2009, assim como a subamostra que corresponde ao caso completo - agregados que se mantiveram no inquérito durante o período considerado (ver figura 4.3). Todas estas subamostras, à exceção da correspondente ao caso completo, apresentam diferentes percentagens de omissões em distintos momentos temporais, conforme tabela 4.13.

CAPÍTULO 4. MODELAÇÕES CONSIDERANDO A EXISTÊNCIA DE DADOS OMISSOS

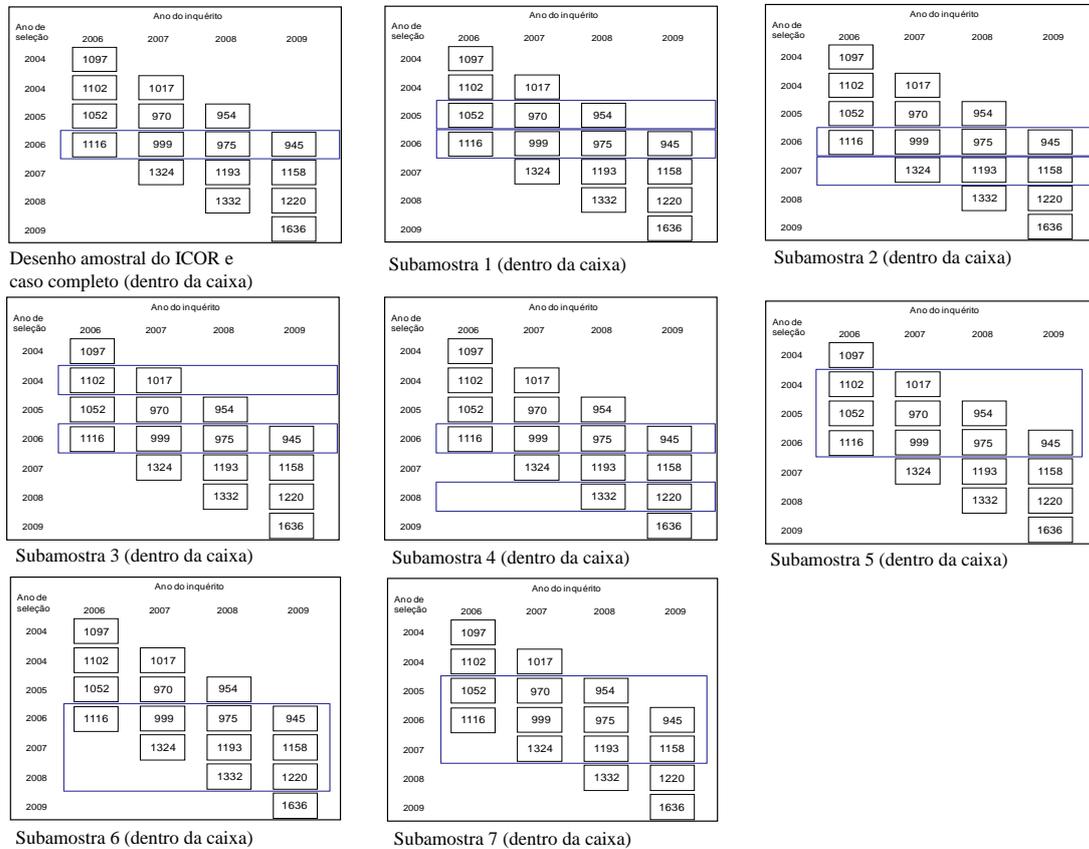


Figura 4.3: Desenho amostral do ICOR e diferentes subamostras consideradas no estudo

Ano	Subamostra						
	1	2	3	4	5	6	7
2006	0%	56%	0%	59%	0%	73%	38%
2007	0%	0%	0%	59%	0%	39%	0%
2008	0%	0%	50%	0%	33%	0%	0%
2009	52%	0%	50%	0%	68%	0%	32%
N	1 762	1 929	1 702	2 061	2 619	3 145	2 846

Tabela 4.13: Percentagem de observações omissas nas diferentes subamostras e a respetiva dimensão

Na modelação longitudinal dos dados é considerado um *score* de privação material que corresponde a uma soma de itens, com valores numa escala de 0 a 9 itens em privação, sendo que 0 itens em privação significa que o agregado não está privado de nenhum destes itens considerados e 9 itens em privação identifica os agregados que estão totalmente privados. Este *score* foi calculado em cada um dos quatro momentos temporais em estudo.

A utilização do software estatístico Mplus permitiu, recorrendo à modelação com um LGCM, proceder à estimação da trajetória de um *score* de privação material para os

anos de 2006 a 2009 (figura 4.1), utilizando a subamostra designada por dados completos. As medidas de qualidade de ajustamento obtidas para este modelo são apresentadas na tabela 4.14 e permitem concluir que é aceitável o ajustamento modelo-dados, embora o valor do RMSEA=0.075 esteja no limite do aceitável (0.08) (Schumacker e Lomax [147]).

A utilização de procedimentos de FIML permitem estimar um modelo com trajetória latente para o *score* de privação material entre 2006 e 2009, para cada uma das restantes subamostras nas quais existem observações omissas e cujas medidas de qualidade do ajustamento são apresentadas na tabela 4.14. Ainda nesta tabela, são apresentadas as diferentes medidas de qualidade de ajustamento quando a abordagem estatística para lidar com as omissões presentes nas diferentes subamostras é a imputação múltipla. No sentido de assegurar uma abordagem “mais tradicional” e uma abordagem “mais contemporânea” foi considerada a imputação de 5 e 20 valores diferentes. A qualidade do ajustamento modelo-dados é aceitável independentemente da subamostra considerada, assim como da abordagem estatística utilizada para lidar com as omissões presentes nas diferentes subamostras, ou ainda, do número de imputações realizadas. De salientar que nas subamostras 2, 4, e 6, o RMSEA apresenta valores no limite do aceitável (0.08), quando é utilizada uma abordagem de imputação múltipla, independentemente do número de imputações geradas. Quando comparadas as medidas de qualidade de ajustamento em cada uma das subamostras consideradas, também é possível observar uma superioridade em termos de ajustamento sempre que é considerada uma abordagem com *full information maximum likelihood* para lidar com a problemática dos dados omissos, exceto nas subamostras 1 e 6, em que o valor do SRMR é superior na abordagem FIML em relação à imputação múltipla com 5 ou 20 imputações.

A tabela 4.15 apresenta as estimativas dos parâmetros obtidas no ajustamento de um modelo com trajetória latente às diferentes subamostras. Para o modelo com dados completos é possível concluir que em 2006 o valor médio da privação para o conjunto dos agregados era de 2.355 itens. O valor da estimativa da média do declive não se mostrou significativo, não permitindo concluir que houve uma diminuição da taxa média de privação dos agregados ao longo do tempo. No entanto, é possível verificar que nem todos os agregados têm o mesmo nível médio de privação material em 2006 e nem todos mudam da mesma forma entre 2006 e 2009 (as variâncias do intercepto e do declive são significativas). Ainda para o modelo com dados completos é possível concluir que níveis mais elevados de privação em 2006 estão associados com uma taxa média de mudança mais lenta. Conclusões idênticas podem ser retiradas para as subamostras 1 e 3, embora o valor médio da privação material em 2006 varie entre 2.339 e 2.346 itens, na subamostra 1, e varie entre 2.357 e 2.430 itens na subamostra 3, conforme a abordagem estatística utilizada na modelação com dados omissos.

Nas restantes subamostras (2, 4, 5, 6, e 7) não é possível de forma inequívoca concluir que existiu uma diminuição da privação material para os agregados ao longo do período considerado, uma vez que as subamostras analisadas permitem conclusões diferentes, sendo que em algumas das subamostras foram obtidas estimativas da média do

declive que se mostraram significativas e noutras subamostras o mesmo não se verificou. Refira-se, no entanto, que para todas as subamostras consideradas foi possível concluir quanto à significância das estimativas da variância do interceto, variância do declive e covariância entre interceto e declive.

Diferentes conjuntos de estimativas são obtidas para os parâmetros do modelo nas diferentes subamostras, mesmo naquelas consideradas comparáveis, tais como a subamostra 1 e 2, a subamostra 3 e 4, e a subamostra 5 e 6, sendo questionável qual das duas opções metodológicas deve o analista tomar.

Assim, com o objetivo de aferir quanto ao enviesamento cometido na estimação dos diferentes parâmetros do modelo, face à existência de dados omissos, foram utilizados uma vez mais procedimentos de Monte Carlo. Neste estudo de simulação foram assumidos como verdadeiros os valores dos parâmetros estimados pelo LGCM para os dados completos. Foram geradas 10 000 amostras de dimensão 845 e com um padrão de omissão de 50% de não respostas no último momento temporal, ano de 2009, isto de forma a gerar amostras com um desenho amostral semelhante ao da subamostra 1. Para cada uma das diferentes amostras geradas foram estimados por FIML os parâmetros do LGCM, sendo o valor médio dessas estimativas apresentado na terceira coluna da tabela 4.15.

A análise da penúltima coluna da tabela 4.16 permite concluir que a percentagem de réplicas nas quais um intervalo a 95% contém o verdadeiro parâmetro é próxima de 95%, para todos os parâmetros de interesse no ajustamento de um modelo com trajetória latente. De salientar ainda que a percentagem de réplicas nas quais é possível rejeitar a hipótese de que o verdadeiro parâmetro é igual a zero, com um nível de significância de 5%, é de 100% quando considerados todos os parâmetros do modelo, exceto a média do declive (ver última coluna da tabela 4.16. Para este caso, apenas em 7.9% das amostras geradas esta estimativa se mostrou significativa. Assim, a significância do valor estimado do parâmetro média do declive nem sempre se verifica, dependendo da subamostra considerada (com ou sem omissões).

As tabelas D.1 a D.6, em anexo, apresentam os resultados obtidos quando, recorrendo a procedimentos de Monte Carlo, são geradas para cada padrão de omissão idêntico às subamostras 2, 3, 4, 5, 6 e 7, 10 000 amostras com 845 observações. Da análise destas seis diferentes tabelas é possível verificar que para a média do declive a percentagem de amostras nas quais o parâmetro se mostrou significativo varia entre 6.6% e 7.7%. Deste modo, são reiteradas as conclusões anteriores, quanto à significância da média do declive.

Abordagem estatística	Dados completos	Subamostra						
		1	2	3	4	5	6	7
AIC	1)FIML	18982.475	20852.972	16169.390	18783.401	24735.098	29215.445	29420.387
	2)IM(20)	21506.641	23802.119	21051.042	25595.255	32091.548	39047.957	34835.463
	3)IM(5)	21504.080	23846.545	21036.530	25661.816	32013.606	38962.005	34909.650
BIC	1)FIML	19031.743	20903.055	16218.346	18834.080	24787.933	29269.927	29473.970
	2)IM(20)	21555.908	23852.202	21099.998	25645.934	32115.787	39102.439	34889.046
	3)IM(5)	21553.347	23896.628	21085.486	25712.494	32066.440	39016.487	34963.233
RMSEA	1)FIML	0.056	0.065	0.055	0.064	0.047	0.058	0.058
	2)IM(20)	0.069	0.085	0.074	0.081	0.070	0.087	0.076
	3)IM(5)	0.071	0.083	0.074	0.087	0.072	0.083	0.076
CFI	1)FIML	0.994	0.991	0.993	0.988	0.995	0.991	0.992
	2)IM(20)	0.990	0.984	0.990	0.987	0.991	0.984	0.987
	3)IM(5)	0.989	0.985	0.990	0.985	0.990	0.986	0.988
TLI	1)FIML	0.993	0.989	0.992	0.986	0.994	0.989	0.991
	2)IM(20)	0.988	0.981	0.988	0.985	0.989	0.981	0.985
	3)IM(5)	0.987	0.982	0.988	0.982	0.988	0.983	0.985
SRMR	1)FIML	0.022	0.021	0.024	0.026	0.022	0.024	0.018
	2)IM(20)	0.021	0.022	0.022	0.026	0.023	0.023	0.019
	3)IM(5)	0.020	0.021	0.024	0.027	0.024	0.022	0.019

Tabela 4.14: Medidas de qualidade do ajustamento para as diferentes subamostras consideradas. Na imputação múltipla foi considerada a imputação de 5 e 20 valores distintos, abordagens IM(5) e IM(20), respetivamente

Estimativas dos parâmetros	Abordagem estatística	Dados completos $N = 845$	Subamostra						
			1 $N = 1762$	2 $N = 1929$	3 $N = 1702$	4 $N = 2061$	5 $N = 2619$	6 $N = 3145$	7 $N = 2846$
Média do intercepto	1)FIML	2.355	2.339	2.411	2.430	2.402	2.388	2.468	2.377
	2)IM(20)		2.346	2.357	2.357	2.381	2.398	2.402	2.357
	3)IM(5)		2.346	2.358	2.429	2.393	2.399	2.394	2.360
Média do declive	1)FIML	-0.008	0.003	-0.032	-0.010	-0.030	-0.001	-0.045	-0.017
	2)IM(20)		-0.011	-0.009	-0.020	-0.012	-0.019	-0.016	-0.014
	3)IM(5)		-0.012	-0.010	-0.018	-0.008	-0.019	-0.013	-0.015
Variância do intercepto	1)FIML	2.441	2.397	2.476	2.630	2.487	2.538	2.534	2.461
	2)IM(20)		2.383	2.434	2.613	2.496	2.506	2.511	2.423
	3)IM(5)		2.386	2.419	2.621	2.514	2.503	2.492	2.415
Variância do declive	1)FIML	0.089	0.092	0.103	0.093	0.087	0.095	0.102	0.109
	2)IM(20)		0.088	0.095	0.093	0.085	0.085	0.089	0.096
	3)IM(5)		0.086	0.096	0.097	0.091	0.084	0.089	0.095
Covariância intercepto/declive	1)FIML	-0.196	-0.203	-0.231	-0.212	-0.182	-0.214	-0.224	-0.239
	2)IM(20)		-0.190	-0.214	-0.213	-0.182	-0.197	-0.205	-0.215
	3)IM(5)		-0.193	-0.212	-0.224	-0.192	-0.193	-0.201	-0.213

Tabela 4.15: Estimativas dos parâmetros obtidas no ajustamento de um modelo com trajetória latente a diferentes subamostras. A *bold* estão os valores que se mostraram significativos

Parâmetros	“população”	média	desvio padrão	média erros padrão	proporção de intervalos de confiança a 95%	proporção de réplicas com coeficiente significativo
μ_α	2.355	2.3555	0.0570	0.0573	0.949	1.000
μ_β	-0.008	-0.0081	0.0171	0.0170	0.948	0.079
$\psi_{\alpha\alpha}$	2.441	2.4387	0.1389	0.1395	0.951	1.000
$\psi_{\beta\beta}$	0.089	0.0888	0.0161	0.0160	0.951	1.000
$\psi_{\alpha\beta}$	-0.196	-0.1956	0.0359	0.0359	0.949	1.000

Tabela 4.16: Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 1

4.4 Modelos para lidar com omissões MNAR

Apesar de que assumir um mecanismo de omissão de dados MAR é usualmente aceitável, existem situações em que este pressuposto não se verifica. Em tais situações, os valores obtidos para as estimativas dos parâmetros do modelo em estudo são enviesados, se as abordagens estatísticas utilizadas se baseiam num mecanismo MAR. Por outro lado, a não existência de testes que confirmem este mecanismo de omissão de dados, não permite descartar completamente a existência de um mecanismo MNAR.

De acordo com Enders [41], Muthén et al. [102] e Reinecke [125] existem duas diferentes abordagens para lidar com dados que apresentam não respostas com mecanismo MNAR, os *selection models* e o *pattern mixture model*, assumindo a fatorização da distribuição conjunta do conjunto de respostas, \mathbf{Y} , e os indicadores de omissão, \mathbf{R} , de forma diferente em cada uma destas abordagens. De modo simplificado, a fatorização da distribuição conjunta de \mathbf{Y} e \mathbf{R} pode ser escrita como

$$[\mathbf{Y}, \mathbf{R}] = [\mathbf{R}|\mathbf{Y}][\mathbf{Y}], \quad (4.16)$$

numa abordagem com os *selection model* e

$$[\mathbf{Y}, \mathbf{R}] = [\mathbf{Y}|\mathbf{R}][\mathbf{R}], \quad (4.17)$$

numa abordagem com os *pattern mixture model*.

Num *pattern mixture model* é assumido que as observações que apresentem o mesmo padrão de omissão têm uma distribuição comum, sendo assim possível considerar para cada subgrupo de observações, com o mesmo padrão de omissões, a mesma trajetória latente. Por fim, para o conjunto de todas as observações é calculada uma trajetória média, sendo as estimativas dos parâmetros que definem essa trajetória, médias ponderadas das diferentes estimativas obtidas para esses parâmetros, em cada um dos subgrupos. Schaffer e Graham [143] salientam a proximidade deste procedimento com uma abordagem multi-grupos do método da máxima verossimilhança.

Uma abordagem estatística alternativa para lidar com dados cujo mecanismo de não resposta é MNAR são os *selection models* (Demirtas [30], Enders [41], Molenberghs e Fitzmaurice [95] e Muthén et al. [102]). Esta técnica apareceu pela primeira vez na literatura aplicada à econometria, no âmbito da análise de regressão como método de correção do enviesamento na presença de omissões MNAR (Amemiya [4], Heckman [65] e Heckman [66]). Todavia, o seu desenvolvimento e aplicação têm-se verificado em estudos na área da medicina, nos quais é frequente a ocorrência de omissões do tipo *dropout*.

Os *selection models* para dados longitudinais combinam um modelo com trajetória latente com equações adicionais, usualmente regressões logísticas, que explicam as omissões. Enders [41] e Enders [42] referem duas diferentes abordagens dos *selection models* para dados longitudinais, a abordagem de Wu e Carroll [170] e a abordagem de Diggle e Kenward [33]. A primeira técnica usa os fatores aleatórios do modelo (intercepto e declive da trajetória latente) para explicarem a probabilidade de omissão, enquanto que a segunda técnica sugerida por Diggle e Kenward [33] considera que a probabilidade de omissão é explicada pelas variáveis observadas.

De acordo com vários autores, Albert e Follmann [1], Demirtas [30], Enders [42], a escolha do modelo adequado pode depender do tipo de problema que se pretende estudar. Assim, o modelo de Diggle e Kenward é considerado útil em situações nas quais é plausível assumir que a omissão é dependente do valor que seria observado. Por exemplo, num estudo sobre drogas percebe-se que indivíduos que usem drogas num momento anterior ao teste de deteção possam não responder, uma vez que o resultado seria positivo. Por outro lado, o modelo sugerido por Wu e Carroll adequa-se particularmente a situações em que o abandono do estudo pode estar relacionado com a trajetória observada. Por exemplo, num estudo sobre a qualidade de vida em doentes sujeitos a um novo tratamento contra o cancro, o estado de saúde pode agravar-se de tal forma que conduza ao abandono.

Vários autores sugerem a comparação dos resultados obtidos na aplicação destas duas metodologias, *selection model* e *pattern mixture model*, com o objetivo de responder a determinada questão científica, como sejam, a evolução da privação material ou a evolução de uma determinada doença, mas também para obter mais conhecimento em cada uma destas modelações, (Enders [42], Hedeker e Gibbons [67], Molenberghs et al. [97], Muthén et al. [102], Verbeke et al. [161]).

De acordo com Enders [42] a codificação dos indicadores de omissão pode ser feita de diferentes formas de acordo com o tipo de omissão que se verifica nos dados. Num estudo em que apenas existem omissões por atrito (*dropouts*) a codificação utilizada deve ser

$$R_t = \begin{cases} 0 & \text{valor observado no momento } t \\ 1 & \text{atrito no momento } t \\ 9999 & \text{atrito num momento anterior} \end{cases} \quad (4.18)$$

correspondendo 9999 ao código de omissão.

A ideia subjacente a esta codificação é a de que as equações das regressões logísticas quantificam a influência das variáveis explicativas na probabilidade condicional de abandono no momento t , sabendo que os agregados ainda estavam presentes no estudos no instante $t - 1$.

Se apenas existem não respostas intermitentes, a codificação dos indicadores de omissão deve ser a seguinte

$$R_t = \begin{cases} 0 & \text{valor observado no momento } t \\ 1 & \text{valor omissos no momento } t \end{cases} \quad (4.19)$$

sendo que, os coeficientes das regressões logísticas refletem a influência das variáveis explicativas na probabilidade de omissão no momento t , independentemente da presença dos agregados no momento $t - 1$.

Se existirem os dois tipos de omissão *dropout* e não respostas intermitentes, a codificação mais utilizada é a seguinte

$$R_t = \begin{cases} 0 & \text{valor observado ou intermitente no momento } t \\ 1 & \text{atrito no momento } t \\ 9999 & \text{atrito num momento anterior} \end{cases} \quad (4.20)$$

em que as não respostas intermitentes tomam a mesma codificação que os valores observados, sugerindo assim que as omissões intermitentes seguem um mecanismo de omissão ignorável. Consequentemente, os coeficientes das regressões logísticas representam a influência das explicativas na probabilidade de atrito no momento t , sabendo que o agregado ainda se mantém no estudo.

Enquanto que no procedimento de codificação anterior os dois diferentes tipos de omissão são tratados de igual forma, Enders [42] sugere uma outra codificação que tem subjacente o conceito de regressão logística multinomial

$$R_t = \begin{cases} 0 & \text{valor observado no momento } t \\ 1 & \text{atrito no momento } t \\ 2 & \text{não resposta intermitente no momento } t \\ 9999 & \text{atrito num momento anterior} \end{cases} \quad (4.21)$$

Esta estratégia de codificação implica a existência de coeficientes de regressão distintos para os diferentes tipos de omissão.

O tipo de omissão que é mais comum de encontrar em dados longitudinais é o *dropout* e com menor frequência as não respostas intermitentes. Todavia, a base de dados ICOR apresenta, além destas últimas, outro tipo de omissão, as entradas tardias. Uma estratégia de codificação proposta nesta tese é a de considerar o indicador de omissão definido da seguinte forma

$$R_j = \begin{cases} 1 & \text{padrão de omissão } j \\ 0 & \text{caso contrário} \end{cases} \quad (4.22)$$

Deste modo, existirá um número de indicadores de omissão R_j igual ao número de padrões de omissão menos um. A codificação aqui proposta para o indicador de omissão R_j pode ser utilizada para comparar os diferentes padrões de omissão com o caso completo, que se obtém quando todos os R_j são zero (categoria de referência).

4.4.1 *Pattern Mixture Models*

Numa abordagem *pattern mixture model* a decomposição da distribuição conjunta de Y e R é dada por

$$P(Y_i, R_i | \boldsymbol{\theta}, \boldsymbol{\Phi}) = P(Y_i | R_i, \boldsymbol{\theta}) P(R_i | \boldsymbol{\Phi}), \quad (4.23)$$

onde Y_i é o valor da variável para a observação i , R_i é o correspondente indicador de omissão, $\boldsymbol{\theta}$ é um conjunto de parâmetros que descrevem a distribuição de Y (os parâmetros do modelo com trajetória latente), $\boldsymbol{\Phi}$ os parâmetros que descrevem a propensão para as omissões em Y (um conjunto de coeficientes de regressões logísticas que predizem R), $P(Y_i | R_i, \boldsymbol{\theta})$ é a distribuição condicional de Y , dado um particular valor de R e $P(R_i | \boldsymbol{\Phi})$ é a distribuição marginal de R . Esta decomposição implica um modelo em duas partes onde a distribuição condicional de Y representa os parâmetros do modelo com trajetória latente para um grupo de observações que partilham o mesmo padrão de omissões e onde a distribuição marginal de R descreve a incidência de diferentes padrões de omissão. Assim, numa abordagem com o *pattern mixture model* é assumido que as omissões influenciam a trajetória estimada (Enders [42], Muthén et al. [102]).

Uma abordagem com o *pattern mixture model* pode ser realizada de acordo com duas diferentes parametrizações. Uma das parametrizações recorre a uma estratégia semelhante a uma análise multigrupos (Enders [42]). Assim, o procedimento consiste em dividir a amostra em subgrupos de observações que partilham o mesmo padrão de omissões, sendo estimado para cada um destes subgrupos um modelo com trajetória latente. De seguida são calculadas médias ponderadas das estimativas obtidas em cada um dos subgrupos, com o objetivo de obter estimativas dos parâmetros para o conjunto de todas as observações (Enders [41]). Este procedimento é realizado no âmbito dos modelos de mistura. Na figura 4.4 apresenta-se o diagrama de um modelo de mistura, em que c corresponde aos grupos pré-definidos para os diferentes padrões de omissão (isto é, cada observação tem probabilidade zero ou um de pertencer a determinado subgrupo).

A utilização de um *pattern mixture model*, abordagem multigrupos, pode apresentar problemas de identificação, se, por exemplo, para um determinado subgrupo, apenas existirem observações para um momento temporal, a maioria dos parâmetros de um modelo

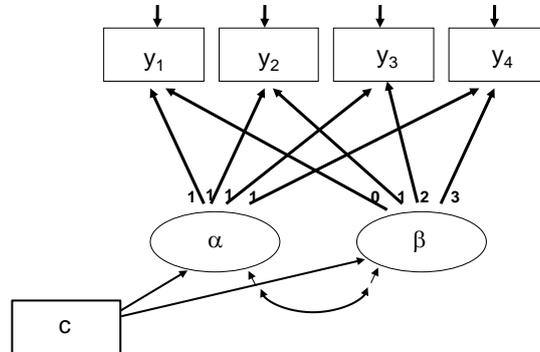


Figura 4.4: Diagrama de um *Pattern Mixture Model* - abordagem multigrupos (c corresponde aos grupos pré-definidos)

com trajetória latente não é estimável. A estimação será apenas possível para o subgrupo que consiste no caso completo, em que o modelo está completamente identificado. Assim, diferentes estratégias podem ser consideradas na especificação de valores para os parâmetros não estimáveis, no sentido de tornar o modelo identificado. Algumas das estratégias mais utilizadas consistem em impor restrições ao modelo utilizando informação acerca dos parâmetros do modelo de uns padrões de omissão (subgrupos) para os outros, tais como, *complete case missing variable restriction*, *neighboring case missing variable restriction* e *available case missing variable restriction* (Demirtas e Schaffer [31], Enders [41], Molenberghs et al. [96], Thijs et al. [155]). De acordo com vários autores, tais como, Albert e Follmann [1] e Ibrahim e Molenberghs [71], entre outros, a introdução deste tipo de restrições num *pattern mixture model*, permite aos investigadores estabelecerem os seus pressupostos de forma explícita, sendo mesmo recomendável uma análise de sensibilidade considerando uma variedade de estratégias de identificação face aos mesmos dados.

Na parametrização dita convencional do *pattern mixture model* é introduzido um indicador de omissão no modelo com trajetória latente como variável explicativa da trajetória, Muthén et al. [102]. O procedimento consiste em considerar indicadores de padrão de omissão, que correspondem a variáveis *dummy* indicando a presença de determinado padrão de omissão (ver equação 4.22 no caso da estratégia de codificação aqui proposta). Estes indicadores são iguais em número aos padrões de omissão menos um (caso completo). Desta forma é possível avaliar o efeito dos diferentes padrões de omissão no comportamento da trajetória face ao caso completo (categoria de referência). A figura 4.5 representa o diagrama de um *pattern mixture model* - parametrização convencional, no qual são considerados três indicadores de padrão de omissão, R_1 , R_2 e R_3 .

4.4.2 Selection Models

Numa abordagem *selection models* a decomposição da distribuição conjunta é dada por

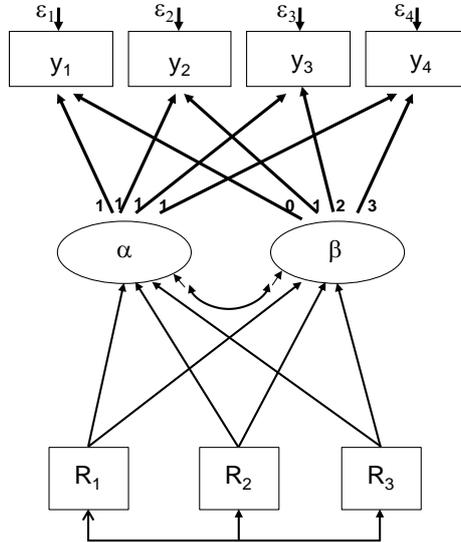


Figura 4.5: Diagrama de um *Pattern Mixture Model* - parametrização convencional

$$P(Y_i, R_i | \boldsymbol{\theta}, \boldsymbol{\Phi}) = P(Y_i | \boldsymbol{\theta}) P(R_i | Y_i, \boldsymbol{\Phi}), \quad (4.24)$$

onde Y_i é o valor da variável para a observação i , R_i é o correspondente indicador de omissão, $\boldsymbol{\theta}$ é um conjunto de parâmetros que descrevem a distribuição de Y (os parâmetros do modelo com trajetória latente), $\boldsymbol{\Phi}$ os parâmetros que descrevem a propensão para as omissões em Y (um conjunto de coeficientes de regressões logísticas que predizem R), $P(Y_i | \boldsymbol{\theta})$ é a distribuição marginal de Y e $P(R_i | Y_i, \boldsymbol{\Phi})$ é a distribuição condicional dos dados omissos, dado Y . Esta factorização implica um modelo constituído por duas partes onde a distribuição marginal corresponde ao modelo com trajetória latente e a distribuição condicional corresponde a um modelo de regressão que usa Y para explicar a probabilidade dos dados omissos. Assim, e ao contrário do que acontece nos *pattern mixture models*, nos *selection models* é assumido que os valores observados ou a trajetória latente influenciam a existência de dados omissos.

Os *selection models* para dados longitudinais combinam um modelo com trajetória latente, com um conjunto de equações de regressão logística (também é usual o recurso a modelos *probit*) que explicam a omissão. A ligação entre as medidas repetidas de uma variável e os indicadores de omissão pode ser feita de duas formas distintas. De acordo com Wu e Carroll [170], as medidas repetidas são indiretamente relacionadas com as probabilidades de resposta através dos efeitos aleatórios do modelo com trajetória latente. Em contraste, a abordagem dos modelos de seleção por Diggle e Kenward [33] relaciona diretamente a probabilidade de omissão dos dados com as medidas repetidas, num determinado momento temporal.

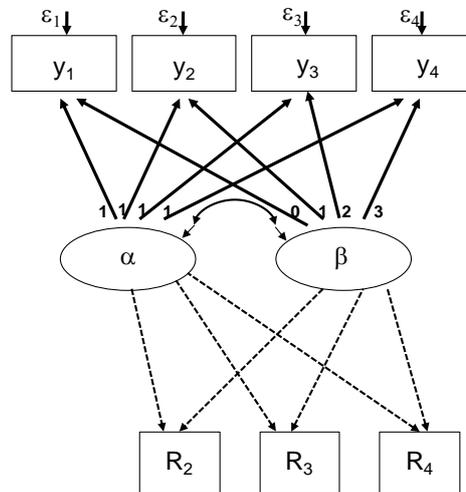


Figura 4.6: Diagrama de um *Selection Model* - abordagem de Wu-Carroll

A figura 4.6 apresenta um modelo de Wu e Carroll, em que R_2, R_3 e R_4 são os indicadores de omissão que indicam se em determinado momento temporal existe observação ou não, isto é, tomam o valor 1 ou 0 se a variável tem valor observado ou não. As setas a tracejado que ligam o intercepto e o declive (os efeitos aleatórios do modelo) com os indicadores de omissão representam regressões logísticas, permitindo deste modo que a probabilidade de omissão dependa de todas as medidas repetidas, incluindo os valores não observados de medidas posteriores. De acordo com Reinecke [125], a detecção de um mecanismo de omissão com este modelo afigura-se complexa.

Num modelo de Diggle e Kenward, embora os R 's continuem a representar indicadores de omissão, a probabilidade de omissão no momento t depende diretamente do valor da variável no momento t e no momento $t - 1$, por exemplo, R_2 depende de Y_2 e de Y_1 , tal como representado na figura 4.7. As setas a tracejado tal como no modelo anterior, representam regressões logísticas. Os valores obtidos para os coeficientes das regressões logísticas consideradas permitem concluir sobre o mecanismo de omissão dos dados. Assim, se não é possível concluir pela significância de uma relação entre Y_t, Y_{t-1} e R_t , as não respostas não estão relacionadas com as observações, o que significa que o mecanismo de omissão de dados é MCAR. Se o efeito da variável Y_{t-1} em R_t é significativo, então o mecanismo de omissão pode ser considerado MAR. Todavia, se o efeito que se mostra significativo é o de Y_t em R_t , um mecanismo de omissão MNAR deve ser considerado plausível, Reinecke [125].

A identificação deste tipo de modelos é feita assumindo diferentes pressupostos de distribuição. Na abordagem de Wu e Carroll, os pressupostos de distribuição são estabelecidos para os efeitos aleatórios, enquanto que na abordagem de Diggle e Kenward

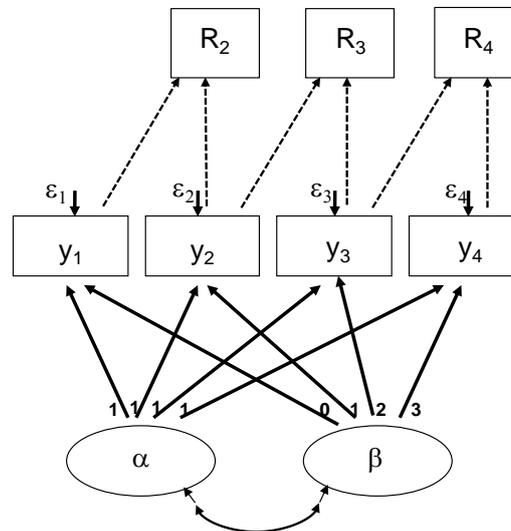


Figura 4.7: Diagrama de um *Selection Model* - abordagem de Diggle-Kenward

estes pressupostos de distribuição são feitos para as medidas repetidas. O procedimento usual, sendo as observações contínuas, é assumir uma distribuição normal multivariada para os interceptos e declives individuais ou para as medidas repetidas. Além disso, o modelo de Wu e Carroll assume que as medidas repetidas e os indicadores de omissão são condicionalmente independentes, dados os efeitos aleatórios.

4.4.3 Resultados do estudo utilizando *Pattern Mixture Models* e *Selection Models*

Este estudo propõe a modelação através de um modelo com trajetória latente de um *score* de privação material para os anos de 2006 a 2009. Do total dos agregados que estão na base de dados entre 2006 a 2009, 6 101 agregados, apenas se mantiveram no painel durante os quatro anos uma pequena fração do total dos agregados, 845 agregados, verificando-se a existência de diferentes padrões de omissão (figura 4.8). Deste modo, o objetivo deste estudo é lidar com os diferentes padrões de omissão que podem ocorrer num painel rotativo como o ICOR, utilizando diferentes abordagens estatísticas para efeitos de estimação do modelo, tendo em consideração o mecanismo de omissão dos dados. Este estudo é apresentado em duas etapas. Numa primeira etapa (correspondente ao estudo 6) foram consideradas as abordagens, *full information maximum likelihood*, e *pattern mixture model* com parametrização multigrupos. Numa segunda etapa (correspondente ao estudo 7) foram consideradas as abordagens, *pattern mixture model* com duas diferentes parametrizações, e *selection model*. De salientar, que as omissões existentes no painel são devido ao desenho do estudo (está-se perante um painel rotativo), mas existem igualmente nesta base de dados outras omissões, tais como, entradas tardias, saídas antes do tempo, entradas tardias e saídas antes de tempo, e não respondas intermitentes.

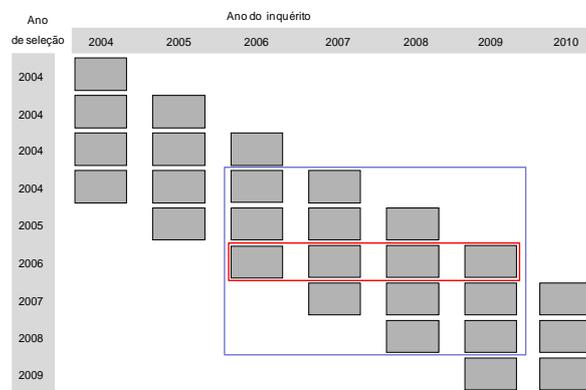


Figura 4.8: ICOR-Desenho do painel, dentro da caixa azul os agregados considerados no estudo, dentro da caixa encarnada o caso completo

Conforme apresentado na tabela 4.17 existem nos dados em estudo 15 padrões diferentes de omissão, sendo que o primeiro corresponde aos 845 agregados que não apresentam omissões, o qual é designado por “caso completo”. Os quatro padrões seguintes são devidos à rotatividade do painel, enquanto que os restantes devem-se a outro tipo de omissões.

2006	2007	2008	2009	N
✓	✓	✓	✓	845
✓	✓	-	-	857
✓	✓	✓	-	917
-	✓	✓	✓	1 084
-	-	✓	✓	1 216
-	-	-	✓	51
✓	-	✓	-	59
✓	-	-	-	531
-	✓	✓	-	137
-	✓	-	-	147
✓	✓	-	✓	25
✓	-	✓	✓	33
-	-	✓	-	159
-	✓	-	✓	40
✓	-	-	✓	0

Tabela 4.17: Diferentes padrões de omissão existentes na base de dados em estudo (✓ = valor observado; - = valor omissão)

A análise descritiva dos dados, apresentada no capítulo 2, figura 2.8, mostra a evolução do número médio de itens em privação para cada padrão de omissão presente na amostra em estudo ao longo dos quatro momentos temporais, sendo esta informação utilizada posteriormente com o objetivo de auxiliar a agrupar os diferentes padrões para uma abordagem com o *pattern mixture model* (parametrização multigrupos).

A modelação longitudinal de um *score* de privação material ao longo de quatro momentos temporais é realizada recorrendo a um modelo com trajetória latente (figura 4.1). Este *score* calculado como uma soma de itens (ver página 87) assume valores entre 0, que significa não estar privado, e 9, correspondendo à privação completa.

São consideradas duas amostras, uma que consiste nos 845 agregados que se mantiveram no estudo nos anos de 2006 a 2009, designada por “caso completo”, e a outra composta por todos os 6 101 agregados, logo com omissões, designada por “caso com todas as observações”. As medidas de qualidade do ajustamento obtidas na modelação com um modelo com trajetória latente permitem concluir que existe um ajustamento razoável modelo-dados (ver tabela 4.18). De salientar o valor do RMSEA de 0.075, para o “caso completo”, valor este que se encontra no limite do aceitável. Para lidar com os dados omissos na amostra para o “caso com todas as observações” foi considerada uma abordagem estatística com o *full information maximum likelihood*.

	“caso completo” <i>N</i> = 845	“caso com todas as observações” <i>N</i> = 6 101
RMSEA	0.075	0.048
CFI	0.991	0.992
TLI	0.989	0.991
SRMR	0.023	0.020

Tabela 4.18: Medidas da qualidade do ajustamento (caso completo e com todas as observações)

As estimativas dos parâmetros do modelo com trajetória latente para as duas diferentes amostras consideradas são apresentadas na tabela 4.19. Para os 845 agregados que constituem o “caso completo”, o nível médio de privação em 2006 é de 2.355 itens, sendo de 2.390 itens para o “caso com todas as observações”. Em ambas as amostras é possível concluir que existem diferenças entre os agregados no nível médio de privação no momento inicial (ano de 2006), bem como na taxa média de mudança da privação material, entre 2006 e 2009. Do mesmo modo, para ambos os casos é possível concluir que a níveis médios de privação mais elevados em 2006 correspondem trajetória médias de mudança menos aceleradas. Todavia, quanto à taxa média de mudança na privação material a conclusão apresenta-se contraditória nas duas amostras, sendo significativa uma diminuição da privação material média ao longo do tempo para o “caso com todas as observações”, enquanto que para os 845 que compõem o “caso completo” a trajetória média de mudança não se mostrou significativa.

No sentido de aferir sobre o enviesamento que se verifica na estimação dos parâmetros de um modelo com trajetória latente nas duas amostras, foram utilizadas técnicas de *bootstrap*, com o objetivo de construir intervalos não paramétricos para os parâmetros média do intercepto e média do declive.

CAPÍTULO 4. MODELAÇÕES CONSIDERANDO A EXISTÊNCIA DE DADOS
OMISSOS

	“caso completo” $N = 845$	“caso com todas as observações” $N = 6101$
Média do intercepto (μ_α)	2.355	2.390
Média do declive (μ_β)	-0.008	-0.022*
Variância do intercepto ($\psi_{\alpha\alpha}$)	2.441	2.580
Variância do declive ($\psi_{\beta\beta}$)	0.089	0.116
Covariância inter/decl ($\psi_{\alpha\beta}$)	-0.196	-0.205

Tabela 4.19: Estimativas dos parâmetros obtidas no ajustamento de um modelo com trajetória latente (“caso completo” e “caso com todas as observações”). A *bold* estão os valores que se mostraram significativos;* 1% nível de significância.

A grande vantagem que as técnicas de *bootstrap* apresentam é a de não necessitarem de pressupostos sobre a distribuição de probabilidade, sendo a ideia base desta técnica de reamostragem, Efron e Tibshirani [39], a de repetidamente extrair amostras de dimensão N com reposição. Dado que as amostras são extraídas com reposição, algumas observações podem aparecer mais do que uma vez numa dada amostra, enquanto que outras não aparecerão.

Assim, seja X_1, X_2, \dots, X_N , uma amostra de dimensão N , com distribuição F de parâmetro θ (amostra original). A técnica de *bootstrap* consiste em reamostrar, com reposição, B amostras $X_1^*, X_2^*, \dots, X_B^*$, independentes de tamanho N . Após a extração das B amostras obtém-se para cada uma destas amostras uma estimativa para o parâmetro de interesse, $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$. A partir da coleção das B estimativas do parâmetro pode-se obter a distribuição de *bootstrap* do estimador $\hat{\theta}$, $F_{\hat{\theta}}$, sendo deste modo possível a construção de intervalos de confiança não paramétricos (figura E.1, em anexo).

Um dos métodos utilizado na construção de intervalos de confiança de *bootstrap* a $100(1 - \alpha)\%$, consiste em considerar os $(\alpha/2)$ - ésimo e $(1 - \alpha/2)$ - ésimo percentis da distribuição empírica de $\hat{\theta}^*$, obtendo-se

$$IC_{bootstrap} = [\hat{\theta}_{(\alpha/2)}^*, \hat{\theta}_{(1-\alpha/2)}^*]. \quad (4.25)$$

Na utilização desta técnica os dados da amostra servem como uma população em miniatura, a partir da qual recorrendo a simulações de Monte Carlo se obtém uma função de distribuição amostral empírica para o estimador de cada parâmetro. A questão de considerar a amostra como uma miniatura da população pode ser vista como uma desvantagem deste tipo de estimação, caso em que a amostra é pequena. Por outro lado, a existência de não respostas é outra das desvantagens. Todavia, das duas amostras consideradas a que contém omissões é de dimensão considerável.

Assim, com o objetivo de construir intervalos de confiança de *bootstrap* para os parâmetros média do intercepto e média do declive foram retiradas 500 amostras (valor recomendado pelo *software* Mplus) com reposição de dimensão 845 cada (a dimensão da amostra designada por “caso completo”), obtendo-se os valores na tabela 4.20. De forma

análoga, foram selecionadas 500 amostras com reposição de dimensão 6 101 observações (dimensão da amostra designada por “caso com todas as observações”), cujos valores são apresentados na tabela 4.21. Da análise das tabelas 4.20 e 4.21 pode verificar-se que os valores estimados para ambos os parâmetros em análise se encontram num intervalo de confiança *bootstrap* a 90%.

	limite inferior 0.5%	limite inferior 2.5%	limite inferior 5%	limite superior 5%	limite superior 2.5%	limite superior 0.5%
Média do intercepto (μ_α)	2.217	2.244	2.264	2.444	2.466	2.490
Média do declive (μ_β)	-0.043	-0.037	-0.030	0.020	0.026	0.033

Tabela 4.20: Caso completo: Intervalos de confiança não paramétricos obtidos por *bootstrap*

	limite inferior 0.5%	limite inferior 2.5%	limite inferior 5%	limite superior 5%	limite superior 2.5%	limite superior 0.5%
Média do intercepto (μ_α)	2.337	2.342	2.354	2.431	2.439	2.457
Média do declive (μ_β)	-0.045	-0.038	-0.036	-0.009	-0.005	0.001

Tabela 4.21: Caso com todas as observações: Intervalos de confiança não paramétricos obtidos por *bootstrap*

De seguida, uma abordagem com o *pattern mixture model*, utilizando uma parametrização multigrupos, foi considerada, uma vez que não é possível assegurar que as omissões existentes nos dados em análise tenham um mecanismo MAR. Deste modo, assume-se a existência de dois diferentes padrões (ou grupos) de omissão: 1) as observações que constituem o “caso completo” mais as observações que apresentam omissões resultantes da rotatividade do painel, e que se admite terem um mecanismo MAR; 2) as observações que apresentam omissões que se supõe terem um mecanismo MNAR. O primeiro padrão designado por “caso completo+MAR” corresponde a 80.6% das observações, correspondendo 18.4% das observações ao segundo padrão designado por “MNAR”. Um modelo com trajetória latente (figura 4.1) foi considerado para cada um destes padrões, apresentando-se na tabela 4.22 as estimativas dos parâmetros dos modelos. Uma trajetória média é calculada para o conjunto dos agregados em estudo, correspondendo as estimativas dos parâmetros desta nova trajetória a uma média ponderada das estimativas dos mesmos parâmetros da trajetória obtida para cada um dos dois diferentes padrões considerados (ver tabela 4.22).

CAPÍTULO 4. MODELAÇÕES CONSIDERANDO A EXISTÊNCIA DE DADOS
OMISSOS

	Média do intercepto	Média do declive
“caso completo+ MAR”	2.432	-0.027*
“MNAR”	2.226	-0.015
Trajectoria média	2.392	-0.025*

Tabela 4.22: Estimativas dos parâmetros obtidas recorrendo ao *Pattern Mixture Model* (parametrização multigrupos - 2 grupos). A *bold* estão os valores que se mostraram significativos; * 1% nível de significância.

As estimativas dos parâmetros para a trajetória média permitem corroborar as conclusões obtidas para a amostra com todas as observações numa abordagem estatística com o *full information maximum likelihood*, dado que o valor estimado para a média do declive se mostrou significativo. Também uma comparação com os intervalos não paramétricos, tabela 4.20 e 4.21, construídos pelas técnicas de *bootstrap* permite concluir que os valores estimados da média do declive se situam num intervalo de *bootstrap* com confiança de 90%, o mesmo acontecendo para a estimativa da média do intercepto da trajetória média. A estimativa da média do intercepto para o padrão “caso completo+MAR” situa-se num intervalo de *bootstrap* a 95% (tabela 4.21), enquanto que a estimativa do mesmo parâmetro para o padrão “MNAR” se encontra num intervalo de *bootstrap* a 99% (tabela 4.20).

De salientar que, a percentagem de observações em cada um dos grupos considerados é bastante diferente, o que vai influenciar o valor obtido para cada um dos parâmetros da trajetória média. Em particular, o valor estimado da média do declive mostra-se significativo na trajetória média, mas o valor estimado deste parâmetro apenas se mostrou significativo no grupo com maior número de observações (“caso completo +MAR”).

De acordo com vários autores, na aplicação de um *pattern mixture model* deve considerar-se diferente número de grupos com padrões de omissão que estão presentes nos dados, no sentido de se proceder a uma análise de sensibilidade. Assim, foi considerado um *pattern mixture model*, com parametrização multigrupos, em que se assumiu a existência de quatro diferentes padrões nos dados: 1) o caso completo; 2) as omissões que correspondem aos agregados que entraram depois de 2006 no painel; 3) as omissões decorrentes do abandono do painel antes de 2009; e 4) outras omissões. A análise da figura 2.8 mostra que esta decomposição é aceitável. Estes quatro grupos representam, respetivamente, 13.9%, 38.5%, 37.8%, e 9.8% das observações. Para cada um destes grupos foi estimada uma trajetória, sendo os valores estimados para os parâmetros destas quatro trajetórias, bem como os valores calculados para a trajetória média apresentados na tabela 4.23.

Da análise desta tabela e analogamente ao caso anterior, é possível concluir que existe uma diminuição da privação material média para o conjunto dos agregados considerados neste painel, entre 2006 e 2009. A análise dos intervalos não paramétricos mostra que as estimativas da média do intercepto e da média do declive se encontram dentro de um

CAPÍTULO 4. MODELAÇÕES CONSIDERANDO A EXISTÊNCIA DE DADOS OMISSOS

	Média do intercepto	Média do declive
“caso completo”	2.346	-0.003
“entradas tardias”	2.584	-0.093
“saídas antes do tempo”	2.379	-0.014
“outras”	2.130	-0.040
Trajectoria média	2.429	-0.027*

Tabela 4.23: Estimativas dos parâmetros obtidas recorrendo ao *Pattern Mixture Model* (parametrização multigrupos - 4 grupos). A *bold* estão os valores que se mostraram significativos; *1% nível de significância.

intervalo de *bootstrap* com 90% de confiança.

De salientar ainda que, apenas para o padrão “entradas tardias” o valor estimado da média do declive se mostrou significativo. Para os outros padrões, embora os valores estimados sejam negativos não se mostraram significativos.

Foi ainda considerada a decomposição dos dados em três diferentes padrões: 1) caso completo; 2) não respostas intermitentes; e 3) outras omissões; correspondendo a 13.9%, 2.5% e 83.6% das observações, respetivamente. Três diferentes restrições sobre a média do declive das trajetórias dos distintos grupos foram consideradas, tais como, igualdade entre dois dos três grupos (tabelas 4.24 a 4.26), no sentido de perceber se o valor estimado para o parâmetro média do declive na trajetória média se continua a mostrar significativo. Todas as aproximações foram consensuais nos resultados, permitindo concluir, tal como nas anteriores análises, que a trajetória média apresenta uma média do declive negativa e significativa com valores que estão dentro dos intervalos de confiança não paramétricos considerados. Todavia, existem alguns grupos em que o valor estimado para a média do declive não se mostrou significativo.

As estimativas da média do intercepto e da média do declive na trajetória média apresentam valores dentro de um intervalo de confiança de *bootstrap* com 90% de confiança.

	Média do intercepto	Média do declive
“caso completo”	2.358	-0.005
“intermitentes”	2.482	-0.037
“outras”	2.088	-0.005
Trajectoria média	2.410	-0.032

Tabela 4.24: Estimativas dos parâmetros obtidas recorrendo ao *Pattern Mixture Model* (3 grupos-restrrição 1). A *bold* estão os valores que se mostraram significativos.

As medidas de qualidade do ajustamento modelo-dados são apresentadas na tabela 4.27, podendo afirmar-se que o melhor ajustamento corresponde a uma abordagem com o

CAPÍTULO 4. MODELAÇÕES CONSIDERANDO A EXISTÊNCIA DE DADOS
OMISSOS

	Média do intercepto	Média do declive
“caso completo”	2.364	-0.008
“intermitentes”	2.423	-0.034
“outras”	2.135	-0.034
Trajetória média	2.408	-0.030

Tabela 4.25: Estimativas dos parâmetros obtidas recorrendo ao *Pattern Mixture Model* (3 grupos-restrição 2). A *bold* estão os valores que se mostraram significativos.

	Média do intercepto	Média do declive
“caso completo”	2.398	-0.027
“intermitentes”	2.414	-0.027
“outras”	2.048	0.019
Trajetória média	2.402	-0.026

Tabela 4.26: Estimativas dos parâmetros obtidas recorrendo ao *Pattern Mixture Model* (3 grupos-restrição 3). A *bold* estão os valores que se mostraram significativos.

pattern mixture model em que são considerados dois padrões: 1) “caso completo+MAR”; e 2) “MNAR”. Todavia, nas diferentes abordagens com o *pattern mixture model* é possível concluir que a média do declive é significativa, ou seja, existe uma diminuição na taxa média de privação material, entre 2006 e 2009, resultado que vem reiterar o obtido com uma abordagem *full information maximum likelihood*.

	AIC	BIC
2 padrões		
(“completo+MAR”/“MNAR”)	55 207.275	55 287.869
3 padrões (restricção 1)		
(“comp”/“intermitentes”/“outras”)	55 615.772	55 689.650
3 padrões (restricção 2)		
(“comp”/“intermitentes”/“outras”)	55 617.047	55 690.925
3 padrões (restricção 3)		
(“comp”/“intermitentes”/“outras”)	55 617.757	55 656.680
4 padrões		
(“comp” / “entradas tardias” / “saídas antes tempo” / “outras”)	64 294.035	64 414.927

Tabela 4.27: Medidas de qualidade do ajustamento utilizando *Pattern Mixture Model*

Numa segunda etapa deste estudo (correspondente ao estudo 7) começou por ser ponderada a utilização de duas abordagens estatísticas diferentes, usualmente utilizadas para lidar com dados cujo mecanismo de omissão é MNAR, os *pattern mixture models* e os *selection models*. Todavia, após uma análise detalhada destas duas classes de modelos não pareceu ser boa prática recorrer à utilização dos *selection models*, pois embora estes modelos sejam muito utilizados na área da medicina e atualmente na área das ciências

sociais, a sua aplicação à temática da privação material não parece ser muito plausível (não faz sentido que um dado agregado deixe de participar no inquérito devido à evolução da sua curva de privação material, ou porque experimentou, em determinado momento temporal, privação material).

Recorda-se (ver equação 4.23 e 4.24) que a ideia subjacente aos *selection models* é de que os valores observados ou a curva de trajetória (dependendo da abordagem considerada) determinam a probabilidade de omissão dos dados. Consequentemente, foi considerada apenas uma abordagem com os *pattern mixture models*. A aplicação dos *pattern mixture models* pode ser feita de acordo com duas diferentes parametrizações, uma aproximação multigrupos (Enders [42]) e uma aproximação convencional (Muthén et al. [102]).

A utilização de um *pattern mixture model*, aproximação multigrupos, já utilizada no estudo 6, é complementada com uma aproximação convencional. Assim, começou por considerar-se que nos dados existiam dois diferentes padrões de omissão, um correspondente ao caso completo mais as observações que apresentam omissões devido à rotatividade, e que se admite terem um mecanismo MAR, e o outro padrão correspondente às observações que apresentam outras omissões, e que supomos terem um mecanismo MNAR. A tabela 4.22 apresenta as estimativas da média do intercepto e do declive para cada um dos grupos, bem como para uma trajetória média do *score* de privação material para a totalidade dos agregados que compõem a amostra. A análise desta tabela permite concluir que as estimativas da média do intercepto e da média do declive na trajetória média são significativas. Para o padrão “caso completo+MAR” a estimativa da média do declive apresenta-se significativa, não se revelando significativa para o padrão “MNAR”, (figura 4.9).

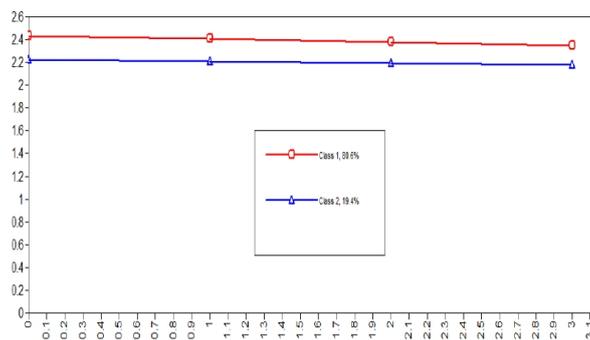


Figura 4.9: Trajetórias estimadas para cada um dos dois padrões considerados: “caso completo+MAR” (a encarnado) e “MNAR” (a azul)

No sentido de confirmar os resultados obtidos procede-se a uma abordagem *pattern mixture model* recorrendo a uma parametrização convencional. Quando se utiliza a parametrização convencional de um *pattern mixture model* os valores obtidos são esti-

mativas dos efeitos dos indicadores de omissão sobre o intercepto e o declive da curva. O indicador de omissão R toma o valor 1, para o padrão “MNAR”, e 0 caso contrário, assim os efeitos representam as diferenças do padrão “MNAR” face ao padrão “caso completo+MAR” (categoria de referência, que se obtém quando $R = 0$). Os valores estimados foram, para o efeito sobre o intercepto 0.206, que se mostrou significativo e para o efeito sobre o declive, 0.012, que não se mostrou significativo. Deste modo, a conclusão quanto à existência de diferenças na evolução da privação material entre os dois padrões de omissão considerados, na abordagem multigrupos, não foi corroborada. Todavia, esta discrepância nos resultados pode ser devida à significância do valor obtido estar no limite.

De seguida, de acordo com uma abordagem com o *pattern mixture model* utilizando uma parametrização multigrupos foram considerados três diferentes padrões nos dados: 1) “caso completo”; 2) dados com omissões devido à rotatividade e cujo mecanismo é MAR; e 3) dados com outras omissões cujo mecanismo é MNAR. Estes três padrões são diferentes dos considerados anteriormente (estudo 6) e não são consideradas restrições. A tabela 4.28 apresenta as estimativas obtidas para os parâmetros de cada uma das trajetórias, assim como para a trajetória média de todas as observações. A média dos valores estimados para a média do declive, para os diferentes padrões considerados, é negativa e significativa, sendo possível concluir que existe uma diminuição da trajetória média de mudança, no período considerado, para o conjunto de todos os agregados, figura 4.10. Todavia, apenas o grupo com padrão “MAR” mostrou ter uma estimativa da média do declive significativa. Os três padrões representavam uma percentagem de 13.9%, 66.7% e 19.4% do total dos agregados, respetivamente.

	Média do intercepto	Média do declive
“caso completo”	2.346	-0.003
“MAR”	2.462	-0.039
“MNAR”	2.226	-0.015
Trajectoria média	2.401	-0.030*

Tabela 4.28: Estimativas dos parâmetros obtidas recorrendo ao *Pattern Mixture Model* (parametrização multigrupos - 3 grupos). A *bold* estão os valores que se mostraram significativos; *1% nível de significância.

O recurso à parametrização convencional permitiu obter os valores apresentados na tabela 4.29. A análise desta tabela permite conclusões idênticas às anteriores, uma vez que apenas para o grupo com padrão MAR o efeito sobre o declive, -0.039, mostrou ser significativo, face ao caso completo (categoria de referência).

A codificação do indicador de padrão de omissão considerada foi

$$R_1 = \begin{cases} 1 & \text{padrão “MAR”} \\ 0 & \text{caso contrário} \end{cases} \quad (4.26)$$

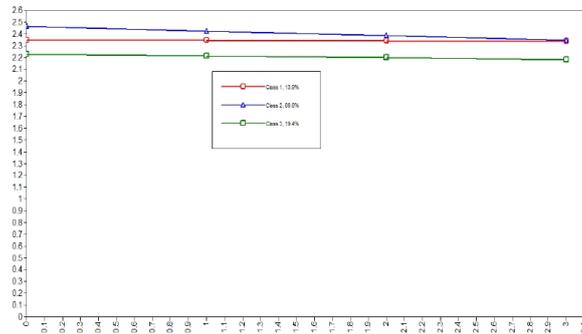


Figura 4.10: Trajetórias estimadas para cada um dos três padrões considerados: “caso completo”(a encarnado), “MAR”(a azul) e “MNAR”(a verde)

	Intercepto	Declive
“MAR”	0.116*	-0.039*
“MNAR”	-0.120	-0.011

Tabela 4.29: Estimativas dos parâmetros obtidas recorrendo ao *Pattern Mixture Model* (parametrização convencional - 3 grupos). A *bold* estão os valores que se mostraram significativos; *1% nível de significância.

$$R_2 = \begin{cases} 1 & \text{padrão “MNAR”} \\ 0 & \text{caso contrário} \end{cases} \quad (4.27)$$

O caso completo verifica-se quando $R_1 = 0$ e $R_2 = 0$.

Quando na aplicação de um *pattern mixture model* com parametrização multigrupos se consideram quatro diferentes padrões nos dados: 1) “caso completo”; 2) “entradas tardias”; 3) “saídas antes do tempo”; e 4) “outras omissões”; (ou seja, quatro grupos), os resultados são os apresentados na tabela 4.23. A análise desta tabela permite concluir que apenas para o grupo das entradas tardias a média do declive da trajetória é significativa, mas tal como nas análises anteriores a trajetória média, para todas as observações, apresenta uma taxa média de mudança significativa e negativa ao longo do período considerado. Na figura 4.11 podem ver-se as quatro trajetórias estimadas.

De seguida, considerando ainda, estes quatro padrões de omissão: 1) “caso completo”; 2) “entradas tardias”; 3) “saídas antes do tempo”; e 4) “outras omissões”; a que correspondem 13.9%, 38.5%, 37.8% e 9.8% das observações, respetivamente, é utilizado um *pattern mixture model* com parametrização convencional.

Recorrendo a uma parametrização convencional a codificação do indicador do padrão de omissão é o seguinte:

$$R_1 = \begin{cases} 1 & \text{padrão “entradas tardias”} \\ 0 & \text{caso contrário} \end{cases} \quad (4.28)$$

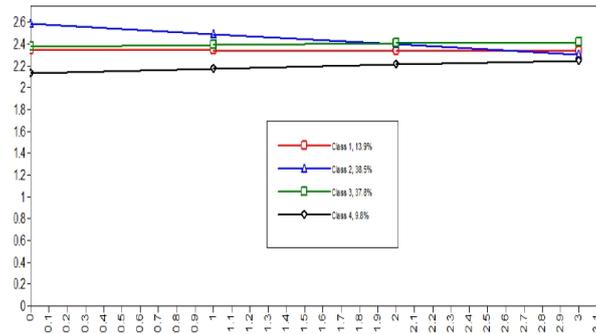


Figura 4.11: Trajetórias estimadas para cada um dos quatro padrões considerados: “caso completo”(a encarnado), “entradas tardias”(a azul), “saídas antes do tempo”(a verde) e “outras”(a preto)

$$R_2 = \begin{cases} 1 & \text{padrão “saídas antes do tempo”} \\ 0 & \text{caso contrário} \end{cases} \quad (4.29)$$

$$R_3 = \begin{cases} 1 & \text{padrão “outras omissões”} \\ 0 & \text{caso contrário} \end{cases} \quad (4.30)$$

Os resultados obtidos, tabela 4.30, são análogos aos obtidos com uma parametrização multigrupos do mesmo modelo, sendo apenas significativo o efeito do padrão “entradas tardias” face ao “caso completo”, -0.090. O “caso completo” é obtido quando $R_1 = 0$ e $R_2 = 0$ e $R_3 = 0$ (categoria de referência).

	Intercepto	Declive
“entradas tardias”	0.238*	-0.090
“saídas antes do tempo”	0.033	0.017
“outras omissões”	-0.215	0.044

Tabela 4.30: Estimativas dos parâmetros obtidas recorrendo ao *Pattern Mixture Model* (parametrização convencional - 4 grupos). A *bold* estão os valores que se mostraram significativos;* 1% nível de significância.

As tabelas 4.31 e 4.32 apresentam as medidas de qualidade do ajustamento para os diferentes modelos considerados. Analisando estas tabelas constata-se que o modelo com dois padrões proporciona melhores resultados numa parametrização multigrupos do *pattern mixture model*. No entanto, as diferentes medidas de qualidade do ajustamento obtidas, para o mesmo modelo numa parametrização convencional, não permitem claramente concluir que o modelo no qual são considerados dois padrões apresenta melhor ajustamento do que o modelo em que se consideram três padrões, uma vez que o valor obtido de AIC é menor, quando são considerados três padrões, face ao valor da mesma medida, quando se consideram dois padrões, embora sejam valores muito próximos. Salienta-se

ainda, que o modelo com três padrões distintos afigura-se mais adequado, uma vez que considera o padrão “caso completo”, que na realidade corresponde aos agregados que estiveram presentes no estudo entre 2006 e 2009.

	AIC	BIC
2 padrões (“completo+MAR”/“MNAR”)	55 207.275	55 287.869
3 padrões (“comp”/“MAR”/“MNAR”)	59 721.465	59 822.208
4 padrões (“comp”/“entradas tardias”/“saidas antes tempo”/“outras”)	64 294.035	64 414.927

Tabela 4.31: Medidas de qualidade do ajustamento utilizando um *Pattern Mixture Model* (parametrização multigrupos)

	AIC	BIC
2 padrões (“completo+MAR”/“MNAR”)	49 206.784	49 280.662
3 padrões (“comp”/“MAR”/“MNAR”)	49 206.264	49 293.575
4 padrões (“comp”/“entradas tardias”/“saidas antes tempo”/“outras”)	49 215.500	49 316.243

Tabela 4.32: Medidas de qualidade do ajustamento utilizando um *Pattern Mixture Model* (parametrização convencional)

Os resultados obtidos recorrendo a uma abordagem com o *pattern mixture model* (com diferentes parametrizações) permitem corroborar as conclusões retiradas com a aplicação de uma abordagem *full information maximum likelihood*, isto é, todas estas diferentes abordagens permitem afirmar que, entre 2006 e 2009, a taxa média de mudança da privação material diminuiu. Neste sentido, é possível afirmar que, é indiferente a utilização de abordagens estatísticas que assentam num mecanismo de omissão MAR ou MNAR para modelar a trajetória de privação material, entre 2006 e 2009, na presença de dados omissos. Todavia, é de salientar que este resultado é contrário ao obtido para os 845 agregados que se mantiveram no estudo entre 2006 e 2009 (caso completo). Para estes não foi possível concluir que existe uma diminuição da taxa média de mudança da privação material, entre 2006 e 2009.

Capítulo 5

Estudo de simulação

5.1 Aspectos introdutórios e principais etapas

Nos estudos apresentados no capítulo anterior, os resultados obtidos na estimação de um modelo com trajetória latente, para os agregados que se mantiveram no estudo de 2006 a 2009 (caso completo) apresentam contradições relativamente aos obtidos para o conjunto de todos os agregados (caso com todas as observações). As abordagens estatísticas utilizadas nestes estudos, para lidar com as omissões existente nos dados, pressupõem que estes são normalmente distribuídos. Todavia, uma análise ao gráfico 2.4 em que estão representados os valores do indicador de privação material, nos momentos temporais de 2006 a 2009, indicia a não presença de uma distribuição normal, sendo visível uma assimetria positiva na distribuição dos dados, podendo esta justificar parcialmente a contradição atrás referida. Do mesmo modo, um insuficiente número de momentos temporais para modelação, ou a dimensão amostral podem ainda ser causas desta contradição. Consequentemente, um estudo de simulação recorrendo a procedimentos de Monte Carlo afigura-se necessário no sentido de analisar o efeito do desvio dos dados face à distribuição normal na estimação de um modelo com trajetória latente, bem como, o efeito da dimensão da amostra e do número de momentos temporais disponíveis para modelação.

Por outro lado, os dados utilizados constituem um painel rotativo, apresentando deste modo omissões que resultam do desenho do estudo. Assim, o efeito da presença de omissões nos dados, resultantes de um *planned missing design*, na estimação de um modelo com trajetória latente é outra questão que se apresenta pertinente e que se pretende analisar neste estudo de simulação.

Numa revisão de literatura sobre estudos de simulação podem ser encontrados vários trabalhos utilizando modelos de análise fatorial confirmatória, tais como, Muthén e Muthén [106] que estuda o efeito da dimensão da amostra e o efeito da existência de não respostas, Bentler e Yuan [15] no qual é estudado o efeito da dimensão da amostra. Encontra-se ainda o trabalho de Fan et al. [46] que analisa o efeito de erros de especificação do modelo e o trabalho de Olsson et al. [112] que estuda o efeito de diferentes métodos de estimação. Por outro lado, Enders [40] analisa o efeito da dimensão da amostra, o efeito das não

respostas e da distribuição dos dados na estimação de um modelo com equações estruturais, com diferentes técnicas para lidar com dados omissos. Ainda, utilizando modelos de equações estruturais, Powell e Schafer [116] apresentam um estudo sobre a robustez da estatística de χ^2 , e Hoogland e Boomsma [68] também apresentam um estudo sobre a robustez de vários métodos de estimação. Tolvanen [157] apresenta um estudo de simulação no âmbito dos modelos de mistura, com o objetivo de avaliar o efeito da dimensão da amostra e número de momentos temporais. Outro estudo visando os modelos de mistura deve-se a Nylund et al. [111]. Todavia, o número de estudos de simulação utilizando modelos com trajetória latente é bastante reduzido, existindo um trabalho de Muthén e Muthén [106] no qual é analisado o efeito da dimensão da amostra na estimação de um modelo de análise fatorial confirmatória e de um modelo com trajetória latente, assim como analisada a potência do teste. Também de Muthén e Asparouhov [101] existe um estudo em que são gerados e analisados, recorrendo aos modelos com trajetória latente, dados com distribuição normal e não normal, com o objetivo de comparar o efeito da utilização de métodos de estimação robustos e não robustos. Deste modo, revela-se de grande interesse o estudo de simulação aqui apresentado.

Quando se planeia um estudo de simulação de Monte Carlo várias decisões devem ser tomadas. Trabalhos como os de Bandalos e Leite [9], Boomsma [22], Paxton et al. [113] pretendem guiar o investigador na elaboração de um estudo deste tipo. De acordo com Paxton et al. [113] são nove as etapas a realizar num estudo de simulação:

- Estabelecer, de forma clara, o objetivo do estudo
- Estabelecer o modelo de interesse
- Definir as questões específicas de investigação, tais como dimensão da amostra
- Definir os valores dos parâmetros na população
- Escolher o software adequado
- Realizar o estudo de simulação
- Guardar os resultados
- Verificar os resultados e encontrar possíveis problemas
- Apresentar os resultados

Os pontos seguintes apresentam a implementação de um estudo de simulação de Monte Carlo, no qual foram cumpridas as etapas estabelecidas como essenciais, pelos autores atrás referidos, para a realização de um estudo deste tipo.

5.1.1 Objetivo do estudo

Este estudo de simulação tem por objetivo analisar o impacto do desvio face à normalidade da distribuição dos dados observados na estimação de um modelo com trajetória latente, assim como analisar o efeito da existência de diferentes padrões de omissão com diferentes percentagens de não resposta, nessa estimação. Do mesmo modo, se pretende avaliar o efeito da existência de omissões resultantes de um *planned missing design* em dados que apresentam distribuição normal, obtidos a partir de modelos com diferentes combinações de parâmetros.

O efeito da dimensão da amostra, do número de momentos temporais, e da fiabilidade que se pretende assegurar, na estimação de um modelo com trajetória latente, constituem ainda questões de interesse neste estudo de simulação. O efeito da dimensão da amostra, de acordo com Paxton et al. [113], é a característica mais estudada quando se realiza um estudo de simulação.

5.1.2 O modelo

O modelo considerado é um modelo com trajetória latente linear com quatro momentos temporais ($T = 4$), características que foram consideradas nos estudos atrás apresentados. Todavia, foram ainda utilizados modelos com trajetória latente com 3 e 5 momentos temporais, com o objetivo de analisar o efeito do número de momentos temporais na estimação de um modelo deste tipo.

Recorde-se que o modelo com trajetória latente é descrito pela seguinte equação de trajetória para y_{it} , com $i = 1, \dots, N$ e $t = 1, \dots, T$

$$y_{it} = \alpha_i + \lambda_t \beta_i + \varepsilon_{it}, \quad (5.1)$$

em que, o intercepto e o declive apresentam as seguintes equações:

$$\alpha_i = \mu_\alpha + \zeta_{\alpha_i} \quad (5.2)$$

$$\beta_i = \mu_\beta + \zeta_{\beta_i} \quad (5.3)$$

Os parâmetros que definem este modelo são as médias do intercepto (μ_α) e do declive (μ_β), as variâncias do intercepto ($\psi_{\alpha\alpha}$) e do declive ($\psi_{\beta\beta}$) e a covariância entre o intercepto e a variância ($\psi_{\alpha\beta}$). Deve ainda ser considerada a variância dos termos residuais ($\theta_{\varepsilon_t} = Var(\varepsilon_t)$).

Como consequência dos pressupostos do modelo sobre a distribuição de probabilidade dos termos residuais, $\varepsilon_{it} \sim N(0, \Theta_\varepsilon)$, $\zeta_{\alpha_i} \sim N(0, \Psi_{\alpha\alpha})$, $\zeta_{\beta_i} \sim N(0, \Psi_{\beta\beta})$, os dados gerados a partir da estrutura imposta por um modelo com trajetória latente têm distribuição normal.

A fiabilidade do indicador y_t , num modelo deste tipo, é dada por

$$Fiabilidade(y_t) = (\psi_{\alpha\alpha} + \lambda_t^2 \psi_{\beta\beta} + 2\lambda_t \psi_{\alpha\beta}) / (\psi_{\alpha\alpha} + \lambda_t^2 \psi_{\beta\beta} + 2\lambda_t \psi_{\alpha\beta} + \theta_{\varepsilon_t}). \quad (5.4)$$

5.1.3 Software utilizado

O estudo de simulação recorrendo a procedimentos de Monte Carlo foi realizado com recurso ao *software* estatístico Mplus, tendo sido programadas várias rotinas no sentido de responder às questões de investigação estabelecidas. Este *software* permite de uma forma integrada, gerar m amostras de dados a partir da estrutura de um modelo com trajetória latente, cujos parâmetros populacionais são definidos à priori pelo investigador, com determinado número de momentos temporais. Posteriormente para cada uma das m amostras geradas, é estimado um modelo com trajetória latente, obtendo-se, deste modo, m estimativas para cada um dos parâmetros do modelo. Se nas amostras geradas existem omissões, uma abordagem *full information maximum likelihood* é utilizada na estimação. Para cada um dos parâmetros do modelo o Mplus disponibiliza a média das estimativas, calculada a partir das m amostras independentes que foram geradas, e o desvio padrão das estimativas, calculado para o conjunto das amostras geradas. Quando o número de amostras m é elevado, este desvio-padrão pode ser considerado como o erro padrão do parâmetro populacional. São ainda disponibilizados o erro padrão médio de estimação, calculado a partir de cada uma das m amostras, a *coverage*, que indica a proporção de amostras para as quais um intervalo a 95% contém o verdadeiro parâmetro, e o erro quadrático médio, uma medida de consistência em média quadrática.

O pacote estatístico Mplus apresenta-se como o *software* mais versátil para estimar modelos com trajetória latente na presença de variáveis que podem ser métricas ou ordinais ou sempre que os dados em análise apresentem omissões. Por estas razões, este foi o *software* utilizado nos vários estudos apresentados neste trabalho. Por outro lado, devido à existência de inúmeras potencialidades, por explorar, que este *software* apresenta para a realização de um estudo de simulação no âmbito dos procedimentos de Monte Carlo, a sua utilização neste estudo de simulação foi o culminar de uma sequência lógica.

O Mplus permite gerar dados com distribuição não normal usando uma mistura de duas subpopulações com distribuição normal, tendo sido esta a razão da programação de algumas rotinas no âmbito dos modelos de mistura. Assim, a não normalidade é obtida gerando dados de duas classes, com diferentes percentagens de observações, sendo que para cada classe é definido um modelo com trajetória latente. Os valores atribuídos aos parâmetros dos dois modelos devem ser distintos, por forma a gerar uma perturbação no conjunto de dados. Em cada uma das classes os dados são gerados com uma distribuição normal. De acordo com Muthén e Asparouhov [101], na classe maioritária, a média do intercepto deve ser um valor baixo, enquanto que na classe minoritária o valor deste parâmetro deve ser alto, tal como a variância do intercepto e a média do declive. Deste modo, as amostras com dados assim gerados apresentam uma assimetria positiva elevada.

Ao contrário de outros pacotes estatísticos para os quais uma distribuição não normal dos dados resulta da definição dos valores de assimetria e achatamento pretendidos para os dados gerados (ver Enders [40]), o Mplus permite gerar dados não normais cuja assimetria e achatamento presente nas amostras resulta da variação de dois fatores: 1) o valor das médias e variâncias do intercepto e declive do modelo com trajetória latente, a partir do

qual os dados são gerados, em cada uma das classes; 2) a percentagem de observações presentes em cada uma das duas classes (maioritária e minoritária) do modelo de mistura considerado.

5.2 Opções metodológicas

5.2.1 Planear as questões de investigação

Neste estudo de simulação o efeito de sete fatores na geração de dados e posterior estimação são considerados, assumindo diferentes conjuntos de valores:

1. Dimensão da amostra: $N = 50, 250, 500$ e $1\,000$

As dimensões de 250, 500 e 1 000 são as mais comuns em estudos de simulação, ver por exemplo Enders [40] ou Paxton et al. [113]. Por outro lado, de acordo com Muthén e Muthén [106], é frequente na área das ciências sociais não ser possível obter amostras de dimensão adequada, a maior parte das vezes as bases de dados têm uma dimensão muito pequena (inferior a 100 casos), pelo que neste estudo se considerou uma dimensão de 50.

2. Número de momentos temporais : $T = 3, 4$ e 5
3. Distribuição dos dados gerados: os dados apresentam uma distribuição normal; os dados apresentam desvios face à distribuição normal
4. Diferentes padrões de omissão resultantes de um *planned missing design* foram considerados para o modelo com trajetória latente com 3, 4 e 5 momentos temporais, pretende-se gerar dados que apresentem padrões de omissão semelhantes aos presentes no painel rotativo ICOR.

Assim, para os modelos com 3 momentos temporais foi considerado o *planned missing design* apresentado na tabela 5.1. Cada um dos 4 padrões que compõem este *planned missing design* representa 1/4 das observações (semelhante ao esquema de rotatividade do ICOR), resultando em diferentes percentagens de omissões em distintos momentos temporais

Y_1	Y_2	Y_3
✓	✓	-
✓	✓	✓
✓	✓	✓
-	✓	✓

Tabela 5.1: *Planned missing design* para modelos com $T = 3$ (✓ = valor observado; - = valor omissão)

Para os modelos com 4 momentos temporais, foi considerado o *planned missing design* com omissões em 1, 2 e 3 momentos temporais, conforme apresentado na

tabela 5.2. Cada um dos 4 padrões que compõem este *planned missing design* representa 1/4 das observações (semelhante ao esquema de rotatividade do ICOR), resultando em diferentes percentagens de omissões em diversos momentos temporais.

Y ₁	Y ₂	Y ₃	Y ₄
✓	✓	✓	-
✓	✓	✓	✓
-	✓	✓	✓
-	-	✓	✓

Tabela 5.2: *Planned missing design* para modelos com $T = 4$ (✓ = valor observado; - = valor omissão)

Para os modelos com 5 momentos temporais foi considerado o *planned missing design* apresentado na tabela 5.3. Cada um dos 4 padrões que compõem este *planned missing design* representa 1/4 das observações (semelhante ao esquema de rotatividade do ICOR), resultando em diferentes percentagens de omissões em diversos momentos temporais.

Y ₁	Y ₂	Y ₃	Y ₄	Y ₅
✓	✓	✓	✓	-
-	✓	✓	✓	✓
-	-	✓	✓	✓
-	-	-	✓	✓

Tabela 5.3: *Planned missing design* para modelos com $T = 5$ (✓ = valor observado; - = valor omissão)

- Utilização do método de estimação de máxima verosimilhança (ML) e um método robusto de máxima verosimilhança (MLR)

O método de estimação de máxima verosimilhança produz boas estimativas dos parâmetros sob o pressuposto da normalidade multivariada, ou até mesmo quando os dados apresentam um desvio na distribuição face à normalidade dos dados, mas pode acontecer uma subestimação dos erros padrão de estimação. Assim, por forma a ultrapassar este problema é utilizada uma versão robusta do estimador de máxima verosimilhança. Por robusta entende-se que esta abordagem calcula erros padrão corrigidos, que são menos influenciados por valores excessivos de achatamento do que os não corrigidos (Bollen e Curran [21]) e ajusta os valores obtidos para a estatística de χ^2 (Muthén e Asparouhov [100]).

- Diferentes modelos com trajetória latente resultantes de diferentes combinações de valores dos parâmetros foram considerados
- Fiabilidade dos indicadores de 0.5 e 0.4

Foi considerada uma fiabilidade dos indicadores de 0.5 porque é um valor realista e frequente em trabalhos na área dos modelos de equações estruturais. Foi ainda

considerado um valor de fiabilidade dos indicadores de 0.4, que pode ser considerado como baixa fiabilidade do indicador.

Quanto ao número de amostras que devem ser geradas, de acordo com Muthén e Asparouhov [100], uma escolha conservadora sugere a geração de 10 000 réplicas. Todavia, em muitos casos apenas a geração de 500 réplicas é suficiente, apresentando ainda a vantagem de uma maior rapidez na condução do estudo de simulação. Powell e Schafer [116], num estudo sobre a robustez da estatística do χ^2 em modelos de equações estruturais, referem um número de réplicas que varia entre 20 e 1 000 com uma média de 200 réplicas. Num estudo de simulação utilizando *item response theory*, Harwell [64] afirma que a escolha do número de réplicas depende do objetivo do estudo, uma vez que existem efeitos que são mais estáveis que outros, sendo necessárias mais réplicas quando se pretende avaliar o erro padrão do que quando o objetivo é medir o enviesamento das estimativas dos parâmetros. Assim, e face às diferentes recomendações, neste estudo foram geradas 1 000 amostras em cada uma das situações consideradas.

5.2.2 Valores dos parâmetros populacionais

Neste estudo os dados são gerados aleatoriamente a partir de um modelo com trajetória latente, em que os valores dos parâmetros que definem o modelo são estabelecidos a priori. Assim, os dados obtidos apresentam distribuição normal com médias, variâncias e covariâncias que dependem dos valores atribuídos aos parâmetros do modelo utilizado na geração de dados (ver equações 3.31, 3.39 e 3.44). Todavia, se o objetivo é obter dados com distribuição não normal, o procedimento consiste em gerar dois grupos de dados, cada um deles com distribuições normal (embora diferente entre si) e considerar uma mistura dos dados dos dois grupos, como será explicado mais à frente neste estudo.

Geração de dados com distribuição normal

Por forma a gerar dados com distribuição normal, foram considerados diferentes modelos com trajetória latente nos quais os parâmetros média e variância do intercepto tomaram os valores 0 e 1, respetivamente, isto é, $\mu_\alpha = 0$ e $\psi_{\alpha\alpha} = 1$. Para a média do declive, foram considerados os valores $\mu_\beta = -4, -2, -1, 0, 1, 2, 4$, e para a variância do declive os valores $\psi_{\beta\beta} = 0.2, 1, 2, 4$. A covariância entre o intercepto e o declive $\psi_{\alpha\beta}$ tomou valores por forma a considerar diferentes graus de correlação entre o intercepto e o declive, tais como, $-0.9, -0.5, 0, 0.5, 0.9$. A variância dos termos residuais foi estabelecida de maneira a assegurar uma fiabilidade de 0.5, em todas as combinações consideradas, num total de 140 modelos diferentes (ver tabela 5.4). Em alguns dos modelos foi ainda considerada uma fiabilidade de 0.4.

De acordo com Muthén e Muthén [106] é usual um rácio de 5 entre os valores das variâncias do intercepto e do declive, assim foram considerados valores no sentido de garantir esta relação e o valor inverso.

Modelo	μ_α	$\psi_{\alpha\alpha}$	μ_β	$\psi_{\beta\beta}$	
A	1	0	1	0	0.2
	2	0	1	0	1
	3	0	1	0	2
	4	0	1	0	4
B	1	0	1	1	0.2
	2	0	1	1	1
	3	0	1	1	2
	4	0	1	1	4
C	1	0	1	2	0.2
	2	0	1	2	1
	3	0	1	2	2
	4	0	1	2	4
D	1	0	1	4	0.2
	2	0	1	4	1
	3	0	1	4	2
	4	0	1	4	4
E	1	0	1	-1	0.2
	2	0	1	-1	1
	3	0	1	-1	2
	4	0	1	-1	4
F	1	0	1	-2	0.2
	2	0	1	-2	1
	3	0	1	-2	2
	4	0	1	-2	4
G	1	0	1	-4	0.2
	2	0	1	-4	1
	3	0	1	-4	2
	4	0	1	-4	4

Tabela 5.4: Parâmetros dos Modelos A1 a G4, para cada um dos modelos $\psi_{\alpha\beta}$ é função da correlação (α, β) , a qual assumiu valores -0.9, -0.5, 0, 0.5, 0.9

De acordo com as equações 3.31, 3.39 e 3.44, por exemplo, para o modelo A1 com correlação 0.9 entre o intercepto e o declive, os valores observados Y_1, Y_2, Y_3 e Y_4 seguem uma distribuição normal com médias, variâncias e covariâncias dadas por

$$E(\mathbf{Y}) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad Cov(\mathbf{Y}) = \begin{bmatrix} 2.5 & 1.4 & 1.8 & 2.2 \\ & 5 & 2.6 & 3.2 \\ & & 8.5 & 4.2 \\ & & & 13 \end{bmatrix}. \quad (5.5)$$

Para o modelo A4 com correlação 0.9 entre o intercepto e o declive obtêm-se os seguintes valores

$$E(\mathbf{Y}) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad Cov(\mathbf{Y}) = \begin{bmatrix} 2.5 & 2.8 & 4.6 & 6.4 \\ & 21.5 & 14.4 & 20.2 \\ & & 60.5 & 34 \\ & & & 119.5 \end{bmatrix}. \quad (5.6)$$

Para o modelo B1 e correlação 0 entre o intercepto e o declive vem

$$E(\mathbf{Y}) = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} \quad Cov(\mathbf{Y}) = \begin{bmatrix} 2 & 1 & 1 & 1 \\ & 2.4 & 1.4 & 1.6 \\ & & 3.6 & 2.2 \\ & & & 5.6 \end{bmatrix}. \quad (5.7)$$

Para o modelo B4 e correlação 0 entre o intercepto e o declive vem

$$E(\mathbf{Y}) = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} \quad Cov(\mathbf{Y}) = \begin{bmatrix} 2 & 1 & 1 & 1 \\ & 10 & 9 & 13 \\ & & 26 & 25 \\ & & & 74 \end{bmatrix}. \quad (5.8)$$

Geração de dados com distribuição não normal (mistura de normais)

Uma abordagem possível para gerar amostras de dados com distribuição não normal consiste em considerar uma mistura de dados provenientes de duas classes, em que cada uma delas tem dados com distribuição normal com diferentes parâmetros caracterizadores, sendo estes dados gerados a partir de um modelo com trajetória latente com diferentes combinações de parâmetros. Este foi o procedimento adotado neste trabalho (por ser o implementado no pacote estatístico utilizado, Mplus). Foram consideradas duas classes, uma minoritária e outra majoritária, sendo a percentagem de observações em cada uma delas de 12% e 88%, respetivamente (Muthén [101]), em que as combinações de valores dos parâmetros do modelo com trajetória latente se encontram na tabela 5.5. Os 5 diferentes modelos apresentados nesta tabela permitem obter dados com diferentes graus de desvio face à normalidade, sendo os valores de assimetria e achatamento univariados resultantes os mais utilizados em estudos aplicados, valores esses que são apresentados na tabela 5.6, (Bandalos e Leite [9], Enders [40], Muthén e Kaplan [103] e Muthén e Muthén [106]). Por outro lado, Bandalos e Leite [9] salientam o facto de que os níveis de assimetria e achatamento a considerar devem ser função do objetivo do estudo. Assim, como os dados utilizados nos estudos 5, 6 e 7, desta tese, o indicador de privação material para os anos de 2006 a 2009 tem uma distribuição com assimetria positiva apenas foram considerados modelos que geram dados que satisfazem esta condição, embora em diferentes graus.

	classe minoritária					classe majoritária				
	μ_α	$\psi_{\alpha\alpha}$	μ_β	$\psi_{\beta\beta}$	$\psi_{\alpha\beta}$	μ_α	$\psi_{\alpha\alpha}$	μ_β	$\psi_{\beta\beta}$	$\psi_{\alpha\beta}$
Modelo H1	2.5	1	0	0.2	0	0	1	0	0.2	0
Modelo H2	2.5	5	4	0.2	0	0	1	0	0.2	0
Modelo H3	15	1	0	0.2	0	0	1	0	0.2	0
Modelo H4	15	5	4	0.2	0	0	1	0	0.2	0
Modelo H5	15	35	4	0.2	0	0	1	0	0.2	0

Tabela 5.5: Parâmetros para os Modelos H1 a H5

Para cada um destes modelos, H1 a H5, o valor assumido para a covariância entre o intercepto e o declive é $\psi_{\alpha\beta} = 0$, em cada classe, de forma a garantir uma correlação

igual a 0 entre o intercepto e o declive. As variâncias dos termos residuais são fixadas a valores que garantem uma fiabilidade de 0.5 na classe maioritária, tendo iguais valores sido considerados na classe minoritária.

Por exemplo, para o modelo H1 recorrendo às equações 3.31, 3.39 e 3.44, verifica-se que os indicadores Y_1 , Y_2 , Y_3 e Y_4 seguem uma distribuição normal com médias, variâncias e covariâncias dadas por

$$E(\mathbf{Y}) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad Cov(\mathbf{Y}) = \begin{bmatrix} 2 & 1.2 & 1.4 & 1.6 \\ & 2.4 & 1.4 & 1.6 \\ & & 3.6 & 2.2 \\ & & & 5.6 \end{bmatrix}, \quad (5.9)$$

na classe maioritária e

$$E(\mathbf{Y}) = \begin{bmatrix} 2.5 \\ 2.5 \\ 2.5 \\ 2.5 \end{bmatrix} \quad Cov(\mathbf{Y}) = \begin{bmatrix} 2 & 1.2 & 1.4 & 1.6 \\ & 2.4 & 1.4 & 1.6 \\ & & 3.6 & 2.2 \\ & & & 5.6 \end{bmatrix}, \quad (5.10)$$

na classe minoritária. Para o modelo H5, os indicadores Y_1 , Y_2 , Y_3 e Y_4 seguem uma distribuição normal com médias, variâncias e covariâncias dadas por

$$E(\mathbf{Y}) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad Cov(\mathbf{Y}) = \begin{bmatrix} 2 & 1.2 & 1.4 & 1.6 \\ & 2.4 & 1.4 & 1.6 \\ & & 3.6 & 2.2 \\ & & & 5.6 \end{bmatrix}, \quad (5.11)$$

na classe maioritária e

$$E(\mathbf{Y}) = \begin{bmatrix} 15 \\ 19 \\ 23 \\ 27 \end{bmatrix} \quad Cov(\mathbf{Y}) = \begin{bmatrix} 36 & 35.2 & 35.4 & 35.6 \\ & 36.4 & 35.4 & 35.6 \\ & & 37.6 & 36.2 \\ & & & 39.6 \end{bmatrix}, \quad (5.12)$$

na classe minoritária.

Assimetria	Achatamento
Fraca: <0.5	Fraco: <0.5
Suave: 0.5 a 1.5	Suave: 0.5 a 2
Moderada: 1.5 a 2.5	Moderado: 2 a 4
Severa: 2.5 a 3.5	Severo: 4 a 7

Tabela 5.6: Níveis de assimetria e achatamento

Geração de dados com omissões

Na análise com dados omissos, os dados são gerados de forma a apresentarem não respostas que configuram um *planned missing design*, com estrutura análoga aos que podem ser encontrados no painel rotativo ICOR. Assim, diferentes *planned missing design*, com 4 padrões de omissão cada, foram considerados, para um modelo com trajetória latente com três, quatro e cinco momentos temporais.

Para o modelo com 3 momentos temporais considerou-se um *planned missing design* com 25% de omissões no 1º e 3º momento temporal, ver tabela 5.1. Para o modelo com 4 momentos temporais o *planned missing design* considerado apresenta 50% de omissões no 1º momento temporal e 25% no 2º e 4º momentos temporais, ver tabela 5.2. Para o modelo com 5 momentos temporais, o *planned missing design* considerado apresenta 75% de omissões no 1º momento temporal, 50% no 2º e 25% no 3º e 5º momentos temporais, ver tabela 5.3.

5.2.3 Critérios de avaliação

Uma revisão de literatura (Bandalos e Leite [9], Enders [40] e Muthén e Muthén [106]) permitiu concluir que os critérios mais utilizados para aferir sobre a qualidade das estimativas obtidas em estudos de simulação, na área dos modelos com equações estruturais são: 1) o enviesamento; 2) o erro quadrático médio; e 3) a *coverage*.

Seja θ um qualquer parâmetro do modelo que se pretende estimar, o primeiro critério de avaliação é o enviesamento que pode ser calculado de diferentes formas. O enviesamento bruto é dado por

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta, \quad (5.13)$$

e o enviesamento relativo é dado por

$$B_R(\hat{\theta}) = \frac{E(\hat{\theta}) - \theta}{\theta}. \quad (5.14)$$

O valor obtido de enviesamento pode ser multiplicado por 100, para obter a percentagem de enviesamento existente na estimação do parâmetro em análise.

Autores como Muthén et al. [104] sugerem que valores de enviesamento relativo inferiores a 0.10 ou 0.15 devem ser negligenciáveis. Todavia, Hoogland e Boosmsma [68] consideram que apenas para valores de enviesamento inferiores a 0.05, isto é, $|B_R(\hat{\theta})| < 0.05$, se pode concluir que não existe enviesamento na estimação de um parâmetro.

O segundo critério é o erro quadrático médio (EQM) do estimador $\hat{\theta}$, dado por

$$\widehat{EQM}(\hat{\theta}) = \frac{1}{m} \sum_{r=1}^m (\hat{\theta} - \theta)^2 = \widehat{Var}(\hat{\theta}) + \hat{B}^2(\hat{\theta}), \quad (5.15)$$

em que

$$\widehat{Var}(\hat{\theta}) = \frac{1}{m} \sum_{r=1}^m (\hat{\theta} - E(\hat{\theta}))^2, \quad (5.16)$$

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta, \quad (5.17)$$

$$E(\hat{\theta}) = \frac{1}{m} \sum_{r=1}^m \hat{\theta}. \quad (5.18)$$

O erro quadrático médio consiste na soma do quadrado do enviesamento mais a variância do parâmetro, conseqüentemente se o estimador pode ser considerado não enviesado, esta é uma medida da eficiência do estimador do parâmetro e um indicador de consistência.

O terceiro critério é a *coverage*, medida que indica a proporção de réplicas (amostras) nas quais um intervalo de confiança a 95% contem o verdadeiro parâmetro e é dado por

$$coverage(\hat{\theta}) = \frac{1}{m} \sum_{r=1}^m IC_{95\%}(\hat{\theta}) \quad (5.19)$$

onde

$$IC_{95\%}(\hat{\theta}) = \begin{cases} 1, & \hat{\theta} - 1.96 \times \widehat{epm}(\hat{\theta}) < \theta < \hat{\theta} - 1.96 \times \widehat{epm}(\hat{\theta}) \\ 0, & \text{caso contrário} \end{cases}$$

em que, $epm(\hat{\theta})$ representa o erro padrão médio do valor estimado $\hat{\theta}$. De acordo com Enders [40], em condições ótimas, os valores de *coverage* deveriam ser iguais ao nível de confiança do intervalo, 0.95 (ou 95%). Por outro lado, Collins et al. [27] sugerem que os valores de *coverage* inferiores a 0.90 (ou 90%) são problemáticos.

5.3 Resultados do estudo de simulação para dados gerados com distribuição normal

O modelo com trajetória latente é examinado sob diferentes condições: 1) dados com distribuição normal sem omissões, com diferentes valores populacionais para a média do declive; 2) dados com distribuição normal sem omissões, com diferentes graus de correlação populacional entre o intercepto e o declive; 3) dados com distribuição normal sem omissões e diferentes valores populacionais da variância do declive; 4) dados com distribuição normal sem omissões considerando diferentes valores de fiabilidade; 5) dados com distribuição normal sem omissões gerados a partir de modelo com diferente número de

momentos temporais e diferentes dimensões amostrais; 6) dados com distribuição normal com omissões, que configuram um *planned missing design*, gerados a partir de modelos com diferentes graus de correlação entre o intercepto e o declive, diferente número de momentos temporais e diferentes dimensões amostrais.

Tal como descrito por outros autores, num estudo de simulação é usual ocorrerem problemas na estimação do modelo em algumas das m amostras geradas, em particular quando estas são de pequena dimensão (Bandalos e Leite [9], Paxton et al. [113], Tolvanen [157]). As causas deste problema podem ser devido à não convergência da estimação, caso em que as amostras que originam problemas são excluídas da análise ou podem ocorrer variâncias negativas, caso em que são mantidas na análise posterior. De acordo com Paxton et al. [113] é esperado ocorrerem variâncias negativas em amostras pequenas devido à usual variação da amostra.

Neste estudo ocorrem problemas de estimação devido à existência de variâncias negativas do intercepto e do declive, em amostras de dimensão 50 e em alguns casos de dimensão 250, com e sem omissões, essencialmente em modelos com correlação entre o intercepto e o declive nula ou positiva, e aumentando com a diminuição do número de momentos temporais do modelo.

5.3.1 Dados gerados com distribuição normal - completos

Foram geradas a partir de um modelo com trajetória latente 1 000 réplicas de dados com distribuição normal com dimensão 50, 250, 500 e 1 000 observações. Para os parâmetros que definem o modelo com trajetória latente foram consideradas as 140 combinações estabelecidas na tabela 5.4, correspondentes aos modelos A a G, cada um deles com diferentes graus de correlação entre o intercepto e o declive, e as variâncias dos termos residuais foram fixadas de modo a assegurar uma fiabilidade de 0.5 em todos os momentos temporais. Para cada um dos modelos A a G foi permitido variar o parâmetro populacional $\psi_{\beta\beta}$. Este procedimento deu origem aos modelos A1, A2, A3 e A4; B1, B2, B3 e B4; C1, C2, C3 e C4; D1, D2, D3 e D4; E1, E2, E3 e E4; F1, F2, F3 e F4; G1, G2, G3 e G4. Posteriormente, foram considerados alguns modelos em que a variância dos termos residuais foi fixada de maneira a garantir uma fiabilidade de 0.4 em cada um dos momentos temporais. Em todas estas rotinas de simulação foi utilizado como método de estimação o método de máxima verosimilhança (ML).

Quando se geram dados a partir dos modelos A1, B1, C1, D1, E1, F1 e G1, as estimativas médias dos parâmetros obtidas na estimação de um modelo com trajetória latente mantêm-se iguais, exceto as estimativas médias do parâmetro média do declive, μ_{β} , o parâmetro que varia entre os diferentes modelos. Todavia, o enviesamento bruto na estimativa média deste parâmetro é igual nos diferentes modelos, e o enviesamento relativo diminui à medida que o seu valor populacional aumenta (em valor absoluto). Os valores obtidos de erro quadrático médio e de *coverage* mantêm-se iguais para todos os parâme-

tros, qualquer que seja a dimensão da amostra e o grau de correlação entre o intercepto e o declive (α e β). Conclusão análoga é possível estabelecer entre os modelos A2, B2, C2, D2, E2, F2 e G2, entre os modelos A3, B3, C3, D3, E3, F3 e G3, e entre os modelos A4, B4, C4, D4, E4, F4 e G4. Consequentemente, apenas serão apresentados os resultados para os modelos A (A1, A2, A3 e A4).

Sempre que o parâmetro populacional é zero, não é possível calcular o enviesamento relativo, sendo apresentado apenas o enviesamento bruto (desvio em relação a zero). Assim, é apresentado para cada parâmetro o enviesamento bruto e sempre que possível o enviesamento relativo dentro de parênteses.

Modelos A com $T = 4$ e correlação(α, β) = -0.9

A tabela 5.7 apresenta os valores de enviesamento na estimativa média de cada um dos parâmetros do modelo com 4 momentos temporais obtidos na estimação do modelo com trajetória latente, a dados gerados com distribuição normal a partir dos modelos A com 4 momentos temporais e correlação entre intercepto e declive igual a -0.9.

A análise desta tabela permite verificar que o enviesamento relativo nas estimativas médias dos parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ diminui à medida que a dimensão da amostra aumenta, e diminui quando se compara o modelo A1 (no qual o parâmetro populacional $\psi_{\beta\beta}=0.2$) com o modelo A2 (com $\psi_{\beta\beta}=1$), mas para os parâmetros $\psi_{\alpha\beta}$ e $\psi_{\alpha\alpha}$ voltam a aumentar quando se consideram os modelos A3 ($\psi_{\beta\beta}=2$) e A4 ($\psi_{\beta\beta}=4$), diminuindo para $\psi_{\beta\beta}$. Para os parâmetros μ_α e μ_β verifica-se que o enviesamento bruto na sua estimativa média é menor em amostras maiores e é maior no modelo A4 ($\psi_{\beta\beta}=4$), isto é, aumenta com o aumento do valor do parâmetro populacional $\psi_{\beta\beta}$ de 0.2 para 4, exceto para o parâmetro μ_α quando $N=50$.

De referir que, os valores de enviesamento relativo e bruto (para μ_α e μ_β) na estimativa média para todos os parâmetros do modelo é sempre inferior a 0.05, valor a partir do qual é referido na literatura como indicativo da existência de enviesamento.

A análise da tabela 5.8 permite observar que o erro quadrático médio (EQM) para estes parâmetros aumenta sempre que o parâmetro populacional variância do declive aumenta de 0.2 para 4, mesmo em amostras de pequena dimensão. O valor do EQM diminui quando a dimensão da amostra aumenta, como descrito na literatura. Para os parâmetros $\psi_{\alpha\beta}$, μ_β e $\psi_{\beta\beta}$ o EQM aumenta com o aumento do valor populacional de $\psi_{\beta\beta}$ de 0.2 para 4, e diminui quando a dimensão da amostra aumenta. O parâmetro que apresenta um EQM menor quando a variância populacional do declive é igual a 0.2, é a variância do intercepto, $\psi_{\alpha\alpha}$, que assume os valores 0.0983, 0.0198, 0.0098 e 0.0049, para amostras de dimensão 50, 250, 500 e 1000, respetivamente. Todavia, este parâmetro, para $\psi_{\beta\beta}=0.2$ e amostras de pequena dimensão (50 observações), apresenta um enviesamento na sua estimativa média superior ao dos restantes parâmetros do modelo, ver tabela 5.7.

Os valores de *coverage* para os diferentes parâmetros (ver tabela 5.9) encontram-se

acima do aceitável, 0.90, qualquer que seja a dimensão da amostra, e o valor populacional de $\psi_{\beta\beta}$ considerado (modelos A1 a A4). De salientar ainda que, para $\psi_{\alpha\beta}$ e $\psi_{\alpha\alpha}$ os valores de *coverage* são baixos, 0.919 e 0.918, respetivamente, quando o valor populacional de $\psi_{\beta\beta}=0.2$ e amostras de pequena dimensão (50 observações). De um modo geral, os parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ apresentam valores menores quando $N=50$. Em amostras de dimensão elevada (1 000 observações), o parâmetro que apresenta valores mais baixos de *coverage* é $\psi_{\alpha\alpha}$.

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.0052 (-0.0130)	0.0038 (-0.0095)	0.0032 (-0.0080)	0.0026 (-0.0065)
	1	0.0077 (-0.0086)	0.0076 (-0.0084)	0.0051 (-0.0057)	0.0044 (-0.0049)
	2	0.0135 (-0.0106)	0.0116 (-0.0091)	0.0076 (-0.0060)	0.0061 (-0.0048)
	4	0.0193 (-0.0107)	0.0170 (-0.0094)	0.0114 (-0.0063)	0.0090 (-0.0050)
μ_{α}	0.2	0.0030	0.0038	0.0014	-0.0005
	1	-0.0053	0.0029	0.0009	-0.0003
	2	-0.0045	0.0035	0.0013	0
	4	-0.0032	0.0040	0.0015	0.0001
μ_{β}	0.2	0.0026	-0.0013	-0.0004	0.0001
	1	0.0079	-0.0021	-0.0006	-0.0006
	2	0.0092	-0.0034	-0.0012	-0.0015
	4	0.0122	-0.0047	-0.0017	-0.0025
$\psi_{\alpha\alpha}$	0.2	-0.0134 (-0.0134)	-0.0079 (-0.0079)	-0.0059 (-0.0059)	-0.0053 (-0.0053)
	1	-0.0082 (-0.0082)	-0.0095 (-0.0095)	-0.0070 (-0.0070)	-0.0062 (-0.0062)
	2	-0.0122 (-0.0122)	-0.0130 (-0.0130)	-0.0104 (-0.0104)	-0.0085 (-0.0085)
	4	-0.0148 (-0.0148)	-0.0178 (-0.0178)	-0.0160 (-0.0160)	-0.0128 (-0.0128)
$\psi_{\beta\beta}$	0.2	-0.0028 (-0.0140)	-0.0019 (-0.0095)	-0.0016 (-0.0080)	-0.0013 (-0.0065)
	1	-0.0122 (-0.0122)	-0.0074 (-0.0074)	-0.0037 (-0.0037)	-0.0034 (-0.0034)
	2	-0.0316 (-0.0158)	-0.0148 (-0.0074)	-0.0059 (-0.0030)	-0.0052 (-0.0026)
	4	-0.0701 (-0.0175)	-0.0226 (-0.0072)	-0.0091 (-0.0023)	-0.0084 (-0.0021)

Tabela 5.7: Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros, dos modelos A, com correlação(α, β) = -0.9

Modelos A com $T = 4$ e correlação(α, β) = -0.5

Da análise da tabela 5.10 é possível observar que o enviesamento relativo na estimativa média de $\psi_{\alpha\beta}$ diminui quando o valor do parâmetro populacional $\psi_{\beta\beta}$ passa de 0.2 para 4 (do modelo A1 para A4), o mesmo acontecendo com a estimativa média de $\psi_{\beta\beta}$ para amostras de dimensão 250 ou mais, e com a estimativa média de $\psi_{\alpha\alpha}$ se $N=50$. Se a dimensão da amostra é de 250 observações ou superior, verifica-se um aumento no enviesamento relativo da estimativa média de $\psi_{\alpha\alpha}$, quando o valor do parâmetro populacional $\psi_{\beta\beta}$ aumenta de 0.2 para 4.

Para o parâmetro μ_{β} existe um aumento no enviesamento bruto da sua estimativa média, quando o valor populacional de $\psi_{\beta\beta}$ aumenta e o parâmetro μ_{α} tende a manter valores semelhantes sendo a dimensão da amostra é de 250 ou mais observações. Em amostras

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\ 000$
$\psi_{\alpha\beta}$	0.2	0.0190	0.0037	0.0019	0.0010
	1	0.0835	0.0164	0.0078	0.0042
	2	0.1910	0.0373	0.0174	0.0094
	4	0.4180	0.0814	0.0377	0.0204
μ_{α}	0.2	0.0271	0.0055	0.0027	0.0014
	1	0.0292	0.0057	0.0028	0.0015
	2	0.0329	0.0065	0.0031	0.0017
	4	0.0354	0.0070	0.0033	0.0018
μ_{β}	0.2	0.0060	0.0012	0.0006	0.0003
	1	0.0281	0.0053	0.0026	0.0014
	2	0.0547	0.0104	0.0051	0.0027
	4	0.1070	0.0201	0.0099	0.0052
$\psi_{\alpha\alpha}$	0.2	0.0983	0.0198	0.0098	0.0049
	1	0.1108	0.0218	0.0105	0.0056
	2	0.1839	0.0361	0.0169	0.0092
	4	0.3715	0.0721	0.0336	0.0183
$\psi_{\beta\beta}$	0.2	0.0047	0.0009	0.0005	0.0002
	1	0.0820	0.0163	0.0078	0.0041
	2	0.3039	0.0602	0.0288	0.0152
	4	1.1172	0.2225	0.1072	0.0562

Tabela 5.8: Erro quadrático médio para os parâmetros dos modelos A com correlação(α, β) = -0.9

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\ 000$
$\psi_{\alpha\beta}$	0.2	0.919	0.935	0.946	0.936
	1	0.925	0.944	0.950	0.944
	2	0.929	0.953	0.959	0.948
	4	0.933	0.955	0.965	0.949
μ_{α}	0.2	0.954	0.955	0.956	0.956
	1	0.951	0.955	0.958	0.954
	2	0.950	0.958	0.965	0.953
	4	0.949	0.956	0.967	0.947
μ_{β}	0.2	0.954	0.953	0.956	0.951
	1	0.950	0.953	0.959	0.945
	2	0.947	0.951	0.959	0.949
	4	0.942	0.952	0.956	0.946
$\psi_{\alpha\alpha}$	0.2	0.918	0.931	0.947	0.933
	1	0.926	0.941	0.947	0.941
	2	0.935	0.945	0.947	0.938
	4	0.942	0.948	0.946	0.938
$\psi_{\beta\beta}$	0.2	0.923	0.950	0.947	0.942
	1	0.920	0.947	0.949	0.948
	2	0.924	0.939	0.954	0.951
	4	0.930	0.939	0.955	0.954

Tabela 5.9: Coverage para os parâmetros dos modelos A, com correlação(α, β) = -0.9

de dimensão mais elevada, $N=1\ 000$, o enviesamento relativo na estimativa média de $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ é menor face aos mesmos valores para $N=500$.

Analisando a tabela 5.10 é ainda possível verificar que o enviesamento na estimativa média dos parâmetros do modelo é inferior a 0.05. O parâmetro que apresenta um maior enviesamento na sua estimativa média é $\psi_{\beta\beta}$ quando a dimensão da amostra é pequena (50 observações) e se a dimensão da amostra é grande (1 000 observações) é $\psi_{\alpha\alpha}$, no modelo A4 ($\psi_{\beta\beta}=4$).

Na tabela 5.11 é possível verificar que o EQM para todos os parâmetros do modelo regista um aumento, com o aumento do valor populacional de $\psi_{\beta\beta}$. Aliás, para o parâmetro $\psi_{\alpha\alpha}$ o erro quadrático médio toma valores superiores, sempre que a variância populacional do declive toma valores baixos (0.2 e 1), face aos restantes parâmetros do modelo. Quando a dimensão da amostra aumenta, o EQM diminui para todos os parâmetros, qualquer que seja o valor populacional de $\psi_{\beta\beta}$.

A análise da tabela 5.12 permite verificar que os valores de *coverage* para todos os parâmetros do modelo se encontram acima do aceitável (0.90), qualquer que seja o valor de $\psi_{\beta\beta}$ e a dimensão da amostra, embora para os parâmetros μ_β , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ estes apresentem valores inferiores quando $N=50$, face aos restantes resultados. O parâmetro $\psi_{\alpha\alpha}$ também apresenta valores baixos de *coverage* se $N=1\ 000$.

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\ 000$
$\psi_{\alpha\beta}$	0.2	0.0026 (-0.0118)	0.0041 (-0.0186)	0.0043 (-0.0195)	0.0033 (-0.0150)
	1	0.0036 (-0.0072)	0.0078 (-0.0156)	0.0072 (-0.0144)	0.0055 (-0.0110)
	2	0.0058 (-0.0082)	0.0110 (-0.0155)	0.0098 (-0.0138)	0.0074 (-0.0104)
	4	0.0083 (-0.0083)	0.0155 (-0.0155)	0.0137 (-0.0137)	0.0102 (-0.0102)
μ_α	0.2	-0.0009	0.0038	0.0015	-0.0004
	1	-0.0029	0.0037	0.0014	-0.0001
	2	-0.0028	0.0039	0.0015	0
	4	-0.0023	0.0041	0.0016	0.0001
μ_β	0.2	0.0055	-0.0008	-0.0002	-0.0003
	1	0.0100	-0.0013	-0.0003	-0.0012
	2	0.0125	-0.0018	-0.0005	-0.0021
	4	0.0163	-0.0024	-0.0008	-0.0032
$\psi_{\alpha\alpha}$	0.2	-0.0119 (-0.0119)	-0.0097 (-0.0097)	-0.0077 (-0.0077)	-0.0069 (-0.0069)
	1	-0.0089 (-0.0089)	-0.0132 (-0.0132)	-0.0113 (-0.0113)	-0.0099 (-0.0099)
	2	-0.0084 (-0.0084)	-0.0164 (-0.0164)	-0.0150 (-0.0150)	-0.0126 (-0.0126)
	4	-0.0070 (-0.0070)	-0.0211 (-0.0211)	-0.0208 (-0.0208)	-0.0171 (-0.0171)
$\psi_{\beta\beta}$	0.2	-0.0042 (-0.0210)	-0.0024 (-0.0120)	-0.0019 (-0.0095)	-0.0017 (-0.0085)
	1	-0.0210 (-0.0210)	-0.0088 (-0.0088)	-0.0042 (-0.0042)	-0.0040 (-0.0040)
	2	-0.0468 (-0.0234)	-0.0168 (-0.0084)	-0.0063 (-0.0032)	-0.0062 (-0.0031)
	4	-0.0996 (-0.0249)	-0.0321 (-0.0080)	-0.0094 (-0.0024)	-0.0098 (-0.0025)

Tabela 5.10: Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros, dos modelos A, com correlação(α, β) = -0.5

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\ 000$
$\psi_{\alpha\beta}$	0.2	0.0277	0.0053	0.0027	0.0014
	1	0.1039	0.0205	0.0099	0.0053
	2	0.2127	0.0417	0.0199	0.0107
	4	0.4415	0.0857	0.0406	0.0218
μ_{α}	0.2	0.0308	0.0062	0.0030	0.0016
	1	0.0333	0.0066	0.0031	0.0017
	2	0.0349	0.0069	0.0033	0.0018
	4	0.0361	0.0071	0.0034	0.0018
μ_{β}	0.2	0.0088	0.0017	0.0009	0.0004
	1	0.0327	0.0062	0.0031	0.0016
	2	0.0614	0.0115	0.0058	0.0029
	4	0.1180	0.0220	0.0112	0.0056
$\psi_{\alpha\alpha}$	0.2	0.1479	0.0286	0.0146	0.0074
	1	0.2203	0.0428	0.0212	0.0113
	2	0.3403	0.0659	0.0322	0.0173
	4	0.6131	0.1179	0.0573	0.0307
$\psi_{\beta\beta}$	0.2	0.0108	0.0020	0.0011	0.0005
	1	0.1160	0.0219	0.0109	0.0057
	2	0.3621	0.0722	0.0359	0.0189
	4	1.2641	0.2539	0.1266	0.0664

Tabela 5.11: Erro quadrático médio para os parâmetros dos modelos A com correlação(α, β) = -0.5

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\ 000$
$\psi_{\alpha\beta}$	0.2	0.929	0.950	0.945	0.942
	1	0.934	0.946	0.952	0.946
	2	0.936	0.945	0.956	0.945
	4	0.938	0.947	0.963	0.947
μ_{α}	0.2	0.959	0.955	0.961	0.959
	1	0.954	0.958	0.965	0.954
	2	0.953	0.956	0.966	0.950
	4	0.953	0.951	0.967	0.950
μ_{β}	0.2	0.940	0.953	0.956	0.945
	1	0.936	0.943	0.951	0.949
	2	0.935	0.948	0.950	0.928
	4	0.928	0.943	0.947	0.931
$\psi_{\alpha\alpha}$	0.2	0.925	0.941	0.937	0.937
	1	0.927	0.943	0.950	0.928
	2	0.934	0.939	0.946	0.928
	4	0.941	0.938	0.947	0.931
$\psi_{\beta\beta}$	0.2	0.926	0.939	0.936	0.942
	1	0.928	0.939	0.938	0.946
	2	0.932	0.947	0.945	0.945
	4	0.933	0.947	0.948	0.941

Tabela 5.12: Coverage para os parâmetros dos modelos A, com correlação(α, β) = -0.5

Modelos A com $T = 4$ e correlação(α, β) = 0

A tabela 5.13 permite verificar que quando o valor do parâmetro populacional $\psi_{\beta\beta}$ no modelo aumenta de 0.2 para 4, o enviesamento bruto na estimativa média do parâmetros $\psi_{\alpha\beta}$ aumenta, enquanto que para $\psi_{\beta\beta}$ o enviesamento relativo na sua estimativa média diminui, qualquer que seja a dimensão da amostra. Para $\psi_{\alpha\alpha}$ existe um aumento para amostras de dimensão 250 ou mais observações. O enviesamento bruto na estimativa média de μ_α e μ_β não parece ser influenciado pelo valor populacional $\psi_{\beta\beta}$ (0.2 a 4), em especial, quando a dimensão da amostra é de 250 observações ou mais. Para amostras de dimensão elevada (1 000 observações) o parâmetro que apresenta maior enviesamento é $\psi_{\alpha\alpha}$, (-0.0195) quando o parâmetro $\psi_{\beta\beta}=4$ (modelo A4), e para amostras de pequena dimensão (50 observações) o parâmetro que apresenta maior enviesamento é $\psi_{\beta\beta}$, no modelo A1 ($\psi_{\beta\beta}=0.2$). Todavia, os enviesamentos brutos e relativos obtidos nas estimativas médias de todos os parâmetros do modelo são inferiores a 0.05, valor abaixo do qual o enviesamento pode ser considerado negligenciável.

Os valores do EQM para μ_α diminuem quando a dimensão da amostra aumenta de 50 para 1000, e tendem a manter-se estáveis se o parâmetro populacional $\psi_{\beta\beta}$ aumenta de 0.2 para 4 (tabela 5.14), especialmente para amostras de dimensão superior a 250 observações. A análise da tabela 5.14 permite ainda verificar que para os parâmetros $\psi_{\alpha\beta}$, μ_β , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ o erro quadrático médio diminui com o aumento de $\psi_{\beta\beta}$ de 0.2 até 4. De salientar que, o parâmetro que apresenta o erro quadrático médio maior é a variância do intercepto, $\psi_{\alpha\alpha}$, sempre que o valor populacional de $\psi_{\beta\beta}$ é baixo (0.2, 1 e 2).

Na tabela 5.15 apresentam-se os valores de *coverage*, todos superiores a 0.90, valor considerado no limite do aceitável na literatura, embora os menores valores sejam para os parâmetros μ_β , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ quando $N=50$, e se $N=1\ 000$, para $\psi_{\alpha\alpha}$.

Modelos A com $T = 4$ e correlação(α, β) = 0.5

Na tabela 5.16 observa-se que o enviesamento relativo na estimativa média dos parâmetros $\psi_{\alpha\beta}$ e $\psi_{\beta\beta}$ diminui com o aumento do valor do parâmetro populacional $\psi_{\beta\beta}$ de 0.2 para 4, enquanto que o enviesamento relativo na estimativa média de $\psi_{\alpha\alpha}$ aumenta, para amostras com 250 ou mais observações. Para o parâmetro μ_α o enviesamento bruto na sua estimativa média mantêm-se constante quando aumenta o valor populacional de $\psi_{\beta\beta}$ (modelos A1 a A4), sendo que μ_β apresenta um aumento no enviesamento obtido.

Para amostras de dimensão elevada (1 000 observações) o parâmetro que apresenta um maior valor de enviesamento relativo é $\psi_{\alpha\alpha}$ (quando $\psi_{\beta\beta}=4$), -0.0190, e para amostras de pequena dimensão (50 observações) é $\psi_{\beta\beta}$ (quando $\psi_{\beta\beta}=0.2$), -0.0430. Todavia, todos os valores de enviesamento obtidos são inferiores a 0.05.

A análise da tabela 5.17 permite concluir que o EQM diminui quando aumenta a dimensão da amostra e aumenta quando aumenta o valor populacional de $\psi_{\beta\beta}$. Em particular, os valores obtidos para μ_α mantêm-se constantes independentemente da variação

CAPÍTULO 5. ESTUDO DE SIMULAÇÃO

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.0006	0.0038	0.0050	0.0036
	1	-0.0020	0.0065	0.0083	0.0060
	2	-0.0035	0.0087	0.0110	0.0079
	4	-0.0057	0.0120	0.0150	0.0107
μ_{α}	0.2	-0.0011	0.0040	0.0016	-0.0002
	1	-0.0020	0.0040	0.0016	0
	2	-0.0020	0.0041	0.0016	0.0001
	4	-0.0018	0.0042	0.0017	0.0002
μ_{β}	0.2	0.0070	-0.0004	0	-0.0005
	1	0.0118	-0.0001	0.0002	-0.0015
	2	0.0150	0	0.0002	-0.0023
	4	0.0194	0.0001	0.0003	-0.0033
$\psi_{\alpha\alpha}$	0.2	-0.0111 (-0.0111)	-0.0104 (-0.0104)	-0.0087 (-0.0087)	-0.0077 (-0.0077)
	1	-0.0052 (-0.0052)	-0.0140 (-0.0140)	-0.0132 (-0.0132)	-0.0116 (-0.0116)
	2	-0.0011 (-0.0011)	-0.0170 (-0.0170)	-0.0170 (-0.0170)	-0.0148 (-0.0148)
	4	0.0050 (0.0050)	-0.0212 (-0.0212)	-0.0227 (-0.0227)	-0.0195 (-0.0195)
$\psi_{\beta\beta}$	0.2	-0.0066 (-0.0330)	-0.0028 (-0.0140)	-0.0020 (-0.0100)	-0.0018 (-0.0090)
	1	-0.0271 (-0.0271)	-0.0093 (-0.0093)	-0.0038 (-0.0038)	-0.0044 (-0.0044)
	2	-0.0550 (-0.0275)	-0.0171 (-0.0086)	-0.0055 (-0.0028)	-0.0069 (-0.0035)
	4	-0.1122 (-0.0281)	-0.0323 (-0.0081)	-0.0080 (-0.0020)	-0.0111 (-0.0028)

Tabela 5.13: Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros, dos modelos A, com correlação(α, β) = 0

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.0380	0.0074	0.0038	0.0020
	1	0.1285	0.0254	0.0127	0.0066
	2	0.2449	0.0484	0.0240	0.0124
	4	0.4813	0.0947	0.0468	0.0243
μ_{α}	0.2	0.0326	0.0065	0.0031	0.0017
	1	0.0347	0.0069	0.0033	0.0018
	2	0.0357	0.0071	0.0033	0.0018
	4	0.0364	0.0072	0.0034	0.0018
μ_{β}	0.2	0.0113	0.0022	0.0011	0.0005
	1	0.0368	0.0070	0.0036	0.0018
	2	0.0669	0.0126	0.0065	0.0032
	4	0.1256	0.0236	0.0123	0.0059
$\psi_{\alpha\alpha}$	0.2	0.1981	0.0375	0.0196	0.0099
	1	0.3431	0.0647	0.0334	0.0172
	2	0.5205	0.0985	0.0504	0.0262
	4	0.8754	0.1660	0.0845	0.0440
$\psi_{\beta\beta}$	0.2	0.0195	0.0037	0.0019	0.0009
	1	0.1490	0.0289	0.0148	0.0076
	2	0.4433	0.0877	0.0446	0.0232
	4	1.4514	0.2920	0.1478	0.0776

Tabela 5.14: Erro quadrático médio para os parâmetros dos modelos A com correlação(α, β) = 0

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.941	0.943	0.943	0.945
	1	0.946	0.938	0.947	0.942
	2	0.940	0.944	0.950	0.947
	4	0.940	0.946	0.955	0.944
μ_{α}	0.2	0.961	0.957	0.962	0.955
	1	0.958	0.955	0.966	0.954
	2	0.954	0.956	0.964	0.951
	4	0.951	0.952	0.967	0.951
μ_{β}	0.2	0.933	0.946	0.947	0.939
	1	0.927	0.947	0.944	0.942
	2	0.925	0.949	0.944	0.950
	4	0.919	0.946	0.937	0.946
$\psi_{\alpha\alpha}$	0.2	0.926	0.950	0.936	0.938
	1	0.929	0.947	0.944	0.925
	2	0.932	0.943	0.943	0.932
	4	0.935	0.939	0.946	0.935
$\psi_{\beta\beta}$	0.2	0.929	0.939	0.942	0.941
	1	0.934	0.946	0.946	0.952
	2	0.924	0.942	0.948	0.948
	4	0.923	0.950	0.950	0.946

Tabela 5.15: *Coverage* para os parâmetros dos modelos A, com correlação(α, β) = 0

do parâmetro populacional $\psi_{\beta\beta}$, especialmente se a dimensão da amostra é igual ou maior a 250 observações. Se o valor populacional de $\psi_{\beta\beta}$ é baixo, diga-se 0.2, 1 e 2, então o parâmetro $\psi_{\alpha\alpha}$ apresenta valores de EQM superiores aos restantes parâmetros, qualquer que seja a dimensão da amostra considerada.

Na tabela 5.18 apresentam-se os valores obtidos de *coverage*, que se encontram acima do aceitável para todos os casos considerados. Todavia, para os parâmetros μ_{β} , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ os valores de *coverage* são menores quando a amostra é de pequena dimensão (50 observações), se a amostra é de dimensão elevada (1 000 observações) $\psi_{\alpha\alpha}$ é o parâmetro que apresenta menores valores de *coverage*.

Modelos A com $T = 4$ e correlação(α, β) = 0.9

Na análise da tabela 5.19 verifica-se que os valores obtidos de enviesamento relativo na estimativa média dos parâmetros $\psi_{\alpha\beta}$ e $\psi_{\beta\beta}$, diminui com o aumento do valor populacional de $\psi_{\beta\beta}$, de 0.2 para 4, enquanto que o enviesamento relativo na estimativa média de $\psi_{\alpha\alpha}$ aumenta, em particular sendo a dimensão da amostra de 250 ou mais observações. Para μ_{α} o enviesamento bruto na sua estimativa média mantém-se constante com o aumento do valor populacional de $\psi_{\beta\beta}$, enquanto que μ_{β} vê o enviesamento na sua estimativa média aumentar. Em amostras de pequena dimensão (50 observações) o parâmetro $\psi_{\beta\beta}$ apresenta um enviesamento no limite do negligenciável, -0.0490. Para amostras de dimensão elevada (1 000 observações) o parâmetro que apresenta maior enviesamento é $\psi_{\alpha\alpha}$, -0.0167, quando $\psi_{\beta\beta}=4$.

CAPÍTULO 5. ESTUDO DE SIMULAÇÃO

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	-0.0013 (-0.0059)	0.0031 (0.0141)	0.0056 (0.0255)	0.0036 (0.0164)
	1	-0.0064 (-0.0128)	0.0046 (0.0092)	0.0086 (0.0172)	0.0058 (0.0116)
	2	-0.0105 (-0.0148)	0.0057 (0.0080)	0.0112 (0.0158)	0.0075 (0.0106)
	4	-0.0162 (-0.0162)	0.0074 (0.0074)	0.0150 (0.0150)	0.0100 (0.0100)
μ_{α}	0.2	-0.0010	0.0041	0.0016	-0.0001
	1	-0.0015	0.0041	0.0017	0.0001
	2	-0.0017	0.0042	0.0017	0.0001
	4	-0.0016	0.0042	0.0017	0.0002
μ_{β}	0.2	0.0078	0	0.0002	-0.0007
	1	0.0128	-0.0009	0.0007	-0.0015
	2	0.0162	0.0016	0.0010	-0.0022
	4	0.0209	0.0025	0.0014	-0.0031
$\psi_{\alpha\alpha}$	0.2	-0.0104 (-0.0104)	-0.0105 (-0.0105)	-0.0091 (-0.0091)	-0.0080 (-0.0080)
	1	-0.0032 (-0.0032)	-0.0135 (-0.0135)	-0.0133 (-0.0133)	-0.0117 (-0.0117)
	2	0.0025 (0.0025)	-0.0157 (-0.0157)	-0.0167 (-0.0167)	-0.0147 (-0.0147)
	4	0.0108 (0.0108)	-0.0192 (-0.0192)	-0.0217 (-0.0217)	-0.0190 (-0.0190)
$\psi_{\beta\beta}$	0.2	-0.0086 (-0.0430)	-0.0029 (-0.0145)	-0.0018 (-0.0090)	-0.0018 (-0.0090)
	1	-0.0296 (-0.0296)	-0.0088 (-0.0088)	-0.0028 (-0.0028)	-0.0041 (-0.0041)
	2	-0.0556 (-0.0278)	-0.0157 (-0.0079)	-0.0035 (-0.0018)	-0.0064 (-0.0032)
	4	-0.1074 (-0.0269)	-0.0289 (-0.0072)	-0.0043 (-0.0011)	-0.0104 (-0.0026)

Tabela 5.16: Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A, com correlação(α, β) = 0.5

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.0486	0.0095	0.0050	0.0025
	1	0.1545	0.0307	0.0158	0.0078
	2	0.2825	0.0562	0.0287	0.0143
	4	0.5340	0.1063	0.0542	0.0269
μ_{α}	0.2	0.0335	0.0067	0.0032	0.0017
	1	0.0353	0.0070	0.0033	0.0018
	2	0.0360	0.0071	0.0034	0.0018
	4	0.0365	0.0072	0.0034	0.0019
μ_{β}	0.2	0.0133	0.0025	0.0013	0.0006
	1	0.0400	0.0076	0.0040	0.0019
	2	0.0706	0.0135	0.0070	0.0034
	4	0.1295	0.0247	0.0129	0.0062
$\psi_{\alpha\alpha}$	0.2	0.2407	0.0449	0.0239	0.0119
	1	0.4364	0.0805	0.0429	0.0216
	2	0.6491	0.1199	0.0637	0.0321
	4	1.0466	0.1941	0.1027	0.0519
$\psi_{\beta\beta}$	0.2	0.0289	0.0054	0.0028	0.0014
	1	0.1905	0.0363	0.0186	0.0094
	2	0.5364	0.1040	0.0529	0.0270
	4	1.6807	0.3320	0.1675	0.0844

Tabela 5.17: Erro quadrático médio para os parâmetros dos modelos A, com correlação(α, β) = 0.5

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.950	0.944	0.943	0.941
	1	0.953	0.945	0.947	0.944
	2	0.951	0.944	0.947	0.945
	4	0.950	0.944	0.950	0.948
μ_α	0.2	0.964	0.957	0.963	0.953
	1	0.957	0.957	0.966	0.952
	2	0.952	0.954	0.965	0.949
	4	0.951	0.935	0.967	0.951
μ_β	0.2	0.925	0.945	0.946	0.940
	1	0.929	0.945	0.941	0.951
	2	0.936	0.948	0.934	0.947
	4	0.933	0.945	0.934	0.947
$\psi_{\alpha\alpha}$	0.2	0.924	0.951	0.938	0.937
	1	0.931	0.947	0.937	0.936
	2	0.930	0.944	0.938	0.936
	4	0.931	0.944	0.942	0.935
$\psi_{\beta\beta}$	0.2	0.935	0.946	0.944	0.945
	1	0.938	0.948	0.948	0.955
	2	0.933	0.951	0.943	0.960
	4	0.929	0.948	0.941	0.954

Tabela 5.18: *Coverage* para os parâmetros dos modelos A, com correlação(α, β) = 0.5

O EQM para o parâmetro μ_α mantém-se constante quando o valor populacional de $\psi_{\beta\beta}$ aumenta, exceto quando a dimensão da amostra é pequena (50 observações). Neste caso, aumenta ligeiramente, de 0.0340 para 0.0366, quando $\psi_{\beta\beta}$ aumenta de 0.2 para 4. Para os restantes parâmetros do modelo com trajetória latente, o EQM diminui quando aumenta a dimensão da amostra e aumenta com o aumento do valor considerado no modelo para o parâmetro populacional $\psi_{\beta\beta}$. O parâmetro $\psi_{\alpha\alpha}$ apresenta um EQM superior ao dos restantes parâmetros do modelo, quando a variância populacional do declive $\psi_{\beta\beta}$ assume valores pequenos ($\psi_{\beta\beta}=0.2, 1$ e 2), qualquer que seja a dimensão da amostra, tabela 5.20.

Os valores de *coverage* apresentados na tabela 5.21, mostram que estes são sempre superiores a 0.90, para todos os parâmetros em consideração, sendo de salientar que para $N=50$ os valores de *coverage* para $\mu_\beta, \psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ são inferiores aos obtidos para outras dimensões amostrais. O parâmetro que apresenta menores valores de *coverage*, sendo a dimensão da amostra elevada (1 000 observações), é $\psi_{\alpha\alpha}$.

Modelos A com $T = 4$ e correlação(α, β) = 0.9 vs Modelos A com $T = 4$ e correlação(α, β) = -0.9

A comparação dos valores apresentados nas tabelas 5.7 e 5.19 para amostras grandes (1 000 observações), permite verificar que o enviesamento relativo na estimativa média dos parâmetros $\psi_{\alpha\beta}, \psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ é menor quando se consideram modelos com correlação entre o intercepto e declive igual a -0.9 face a modelos com correlação igual a 0.9, qualquer que

CAPÍTULO 5. ESTUDO DE SIMULAÇÃO

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	-0.0026 (-0.0065)	0.0025 (0.0063)	0.0056 (0.0140)	0.0036 (0.0090)
	1	-0.0090 (-0.0100)	0.0028 (0.0031)	0.0086 (0.0096)	0.0053 (0.0059)
	2	-0.0140 (-0.0110)	0.0031 (0.0024)	0.0110 (0.0087)	0.0067 (0.0053)
	4	-0.0212 (-0.0118)	0.0035 (0.0019)	0.0144 (0.0080)	0.0087 (0.0048)
μ_{α}	0.2	-0.0009	0.0041	0.0016	-0.0001
	1	-0.0014	0.0042	0.0017	0.0001
	2	-0.0014	0.0042	0.0017	0.0002
	4	-0.0014	0.0042	0.0017	0.0002
μ_{β}	0.2	0.0083	0.0004	0.0004	-0.0007
	1	0.0130	0.0017	0.0010	-0.0014
	2	0.0163	0.0027	0.0015	-0.0019
	4	0.0209	0.0041	0.0021	-0.0026
$\psi_{\alpha\alpha}$	0.2	-0.0101 (-0.0101)	-0.0104 (-0.0104)	-0.0091 (-0.0091)	-0.0079 (-0.0079)
	1	-0.0038 (-0.0038)	-0.0126 (-0.0126)	-0.0127 (-0.0127)	-0.0109 (-0.0109)
	2	0.0013 (0.0013)	-0.0144 (-0.0144)	-0.0154 (-0.0154)	-0.0133 (-0.0133)
	4	0.0087 (0.0087)	-0.0169 (-0.0169)	-0.0195 (-0.0195)	-0.0167 (-0.0167)
$\psi_{\beta\beta}$	0.2	-0.0098 (-0.0490)	-0.0029 (-0.0145)	-0.0016 (-0.0080)	-0.0017 (-0.0085)
	1	-0.0298 (-0.0298)	-0.0080 (-0.0080)	-0.0016 (-0.0016)	-0.0033 (-0.0033)
	2	-0.0523 (-0.0262)	-0.0137 (-0.0069)	-0.0012 (-0.0006)	-0.0050 (-0.0025)
	4	-0.0955 (-0.0239)	-0.0244 (-0.0061)	0.0001 (0)	-0.0078 (-0.0020)

Tabela 5.19: Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A, com correlação(α, β) = 0.9

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.0573	0.0113	0.0059	0.0029
	1	0.1761	0.0351	0.0182	0.0088
	2	0.3146	0.0628	0.0325	0.0157
	4	0.5839	0.1167	0.0602	0.0291
μ_{α}	0.2	0.0340	0.0068	0.0032	0.0017
	1	0.0355	0.0071	0.0033	0.0018
	2	0.0361	0.0072	0.0034	0.0018
	4	0.0366	0.0073	0.0034	0.0019
μ_{β}	0.2	0.0148	0.0028	0.0015	0.0007
	1	0.0420	0.0081	0.0042	0.0020
	2	0.0726	0.0140	0.0072	0.0035
	4	0.1311	0.0254	0.0131	0.0063
$\psi_{\alpha\alpha}$	0.2	0.2694	0.0498	0.0268	0.0132
	1	0.4841	0.0881	0.0479	0.0237
	2	0.7017	0.1275	0.0693	0.0343
	4	1.0942	0.1991	0.1081	0.0534
$\psi_{\beta\beta}$	0.2	0.0370	0.0069	0.0036	0.0017
	1	0.2258	0.0424	0.0217	0.0107
	2	0.6179	0.1174	0.0597	0.0297
	4	1.8996	0.3654	0.1850	0.0924

Tabela 5.20: Erro quadrático médio para os parâmetros dos modelos A com correlação(α, β) = 0.9

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.955	0.946	0.946	0.947
	1	0.946	0.948	0.945	0.947
	2	0.947	0.948	0.946	0.949
	4	0.944	0.947	0.946	0.953
μ_{α}	0.2	0.963	0.958	0.963	0.955
	1	0.956	0.956	0.965	0.950
	2	0.957	0.953	0.965	0.950
	4	0.951	0.952	0.968	0.951
μ_{β}	0.2	0.931	0.948	0.943	0.954
	1	0.940	0.950	0.938	0.950
	2	0.941	0.949	0.944	0.949
	4	0.943	0.948	0.944	0.949
$\psi_{\alpha\alpha}$	0.2	0.929	0.952	0.937	0.937
	1	0.933	0.952	0.934	0.942
	2	0.931	0.950	0.935	0.943
	4	0.936	0.949	0.935	0.937
$\psi_{\beta\beta}$	0.2	0.936	0.945	0.945	0.944
	1	0.942	0.947	0.951	0.950
	2	0.934	0.952	0.947	0.953
	4	0.926	0.954	0.945	0.954

Tabela 5.21: *Coverage* para os parâmetros dos modelos A com correlação $(\alpha, \beta) = 0.9$

seja o valor da variância populacional $\psi_{\beta\beta}$. Para os parâmetros μ_{α} e μ_{β} verifica-se que o enviesamento bruto não parece ser influenciado pelo valor da correlação considerado entre o intercepto e o declive.

Em amostras de pequena dimensão (50 observações) o enviesamento é maior para os parâmetros μ_{β} , $\psi_{\beta\beta}$ e $\psi_{\alpha\beta}$ (se o valor populacional $\psi_{\beta\beta}=4$), quando se comparam os resultados obtidos considerando uma correlação de 0.9 face aos valores obtidos considerando uma correlação de -0.9, entre o intercepto e o declive. Por outro lado, é menor para os parâmetros $\psi_{\alpha\beta}$ (se $\psi_{\beta\beta}=0.2$), μ_{α} e $\psi_{\alpha\alpha}$, quando a correlação é considerada igual a 0.9.

Os valores obtidos para o EQM, tabelas 5.8 e 5.20, para todos os parâmetros do modelo, são inferiores quando se consideram modelos com correlação entre intercepto e declive igual a -0.9, qualquer que sejam a dimensão da amostra e o valor populacional para $\psi_{\beta\beta}$ considerados.

Os valores de *coverage* são em ambos os casos aceitáveis, mas se $N=50$, sendo a correlação entre o intercepto e o declive igual a -0.9, os parâmetros mais afetados são $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, enquanto que se a correlação entre o intercepto e o declive é igual a 0.9, os parâmetros mais afetados são μ_{β} , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$. Para $N=1000$, o parâmetro que apresenta valores de *coverage* mais baixos é $\psi_{\alpha\alpha}$ em ambos os casos.

Ao contrário do esperado não existe um “efeito simetria” nos resultados obtidos, isto é, os valores de enviesamento na estimação dos parâmetros e de EQM não são semelhantes

seja a correlação entre intercepto e declive positiva ou negativa. Uma das razões para estes resultados pode ser a existência de maior número de réplicas com problemas na estimação do modelo, quando se assume uma correlação positiva entre intercepto e declive, face ao mesmo modelo com correlação negativa (neste caso existe 1 amostra em 1 000 com problemas), quando as amostras são de pequena dimensão. Todavia, quando as amostras são de dimensão elevada não existem amostras com problemas de estimação.

Modelos A com $T = 4$ e correlação $(\alpha, \beta) = 0.5$ vs Modelos A com $T = 4$ e correlação $(\alpha, \beta) = -0.5$

A comparação dos valores apresentados nas tabelas 5.10 e 5.16 para amostras grandes, mostram que o enviesamento nas estimativas médias obtidas para os parâmetros μ_α , μ_β e $\psi_{\beta\beta}$ nos modelos em que se considerou correlação 0.5, são semelhantes aos obtidos para os modelos nos quais foi considerada uma correlação entre o intercepto e o declive igual a -0.5. Todavia, para $\psi_{\alpha\alpha}$ o enviesamento é superior se no modelo considerado a correlação é 0.5, o mesmo acontece para $\psi_{\alpha\beta}$ (quando se considera $\psi_{\beta\beta}=0.2$).

Para amostras de dimensão pequena (50 observações), o enviesamento na estimativa média dos parâmetros $\psi_{\alpha\beta}$ (quando $\psi_{\beta\beta}=4$), μ_β , $\psi_{\alpha\alpha}$ (quando $\psi_{\beta\beta}=4$) e $\psi_{\beta\beta}$ é maior se a correlação considerada entre o intercepto e o declive é 0.5. Para o parâmetro μ_α o enviesamentos obtido na sua estimativa média é inferior, quando se considera uma correlação entre o intercepto e o declive igual a 0.5.

Os valores do EQM apresentados nas tabelas 5.11 e 5.17, são sempre inferiores nos modelos com correlação -0.5, exceto para o parâmetro μ_α em que os valores em ambas as tabelas são muito próximos.

Os valores de coverage são em ambos os casos aceitáveis, mas se $N=50$, sendo a correlação entre o intercepto e o declive -0.5, os parâmetros mais afetados são $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, enquanto que se a correlação entre o intercepto e o declive é 0.5, os parâmetros mais afetados são μ_β , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$. Para $N=1\,000$, o parâmetro que apresenta valores de *coverage* mais baixos é $\psi_{\alpha\alpha}$ em ambos os casos.

Tal como, no caso anterior, também com uma correlação entre o intercepto e o declive de 0.5 (positiva e negativa) não é possível verificar um efeito “simetria” nos resultados obtidos.

Modelos A com $T = 4$ e correlação $(\alpha, \beta) = -0.5$ vs Modelos A com $T = 4$ e correlação $(\alpha, \beta) = 0$

A comparação dos valores apresentados nas tabelas 5.10 e 5.13 permite afirmar que o enviesamento nas estimativas médias dos parâmetros é semelhante em modelos que consideram uma correlação entre o intercepto e o declive igual a -0.5 em relação a modelos com correlação 0, para amostras de dimensão grande (1 000 observações), exceto para o

parâmetro $\psi_{\alpha\alpha}$ que apresenta valores de enviesamento menores quando a correlação entre o intercepto e o declive é -0.5.

Para amostras de pequena dimensão (50 observações), os parâmetros μ_β e $\psi_{\beta\beta}$ apresentam maiores valores de enviesamento quando a correlação é 0, enquanto que os parâmetros $\psi_{\alpha\beta}$, μ_α e $\psi_{\alpha\alpha}$ apresentam valores de enviesamento menores quando a correlação é 0.

Os valores dispostos nas tabelas 5.11 e 5.14 mostram que para modelos em que se considerou um correlação entre o intercepto e o declive igual a -0.5, os valores de EQM são menores do que os obtidos quando é considerado um modelo com correlação 0 entre intercepto e declive. Se a dimensão da amostra é maior do que 50 observações, o EQM nos dois modelos é semelhante.

Os valores de *coverage* são em ambos os casos aceitáveis, mas se $N=50$, sendo a correlação entre o intercepto e o declive -0.5, os parâmetros mais afetados são $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, enquanto que se a correlação entre o intercepto e o declive é 0, os parâmetros mais afetados são μ_β , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$. Para $N=1000$, o parâmetro que apresenta valores de *coverage* mais baixos é $\psi_{\alpha\alpha}$ em ambos os casos.

Modelos A com $T = 4$ e correlação(α, β) = 0.5 vs Modelos A com $T = 4$ e correlação(α, β) = 0

A comparação dos valores apresentados nas tabelas 5.16 e 5.13 permite afirmar que, sendo a dimensão da amostra elevada (1000 observações), o enviesamento na estimativa média dos parâmetros é semelhante em ambos os modelos, com correlação entre o intercepto e o declive igual a 0 e igual a 0.5. Se a dimensão da amostra é pequena (50 observações), o enviesamento na estimativa média dos parâmetros $\psi_{\alpha\beta}$, μ_β e $\psi_{\beta\beta}$ (sendo o valor populacional $\psi_{\beta\beta}=0.2$) é superior no modelo que considera a correlação igual a 0.5. O parâmetro μ_α apresenta valores de enviesamento semelhante, em ambos os modelos, enquanto que $\psi_{\alpha\alpha}$ apresenta valores inferiores, quando a correlação é considerada igual a 0.5 e sendo o parâmetro populacional $\psi_{\beta\beta}<4$.

Quanto aos valores do EQM, estes são inferiores no modelo com correlação igual a 0 entre o intercepto e o declive, exceto para o parâmetro μ_α , caso em que os valores obtidos são semelhantes, tabelas 5.17 e 5.14.

Os valores de *coverage* são em ambos os casos aceitáveis, mas se $N=50$, os parâmetros mais afetados são μ_β , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, se $N=1000$, o parâmetro que apresenta valores de *coverage* mais baixos é $\psi_{\alpha\alpha}$ em ambos os casos.

Modelos A com $T = 4$ e correlação(α, β) = 0.5 vs Modelos A com $T = 4$ e correlação(α, β) = 0.9

A análise das tabelas 5.16 e 5.19 permite verificar, que se a dimensão da amostra é grande (1 000 observações), os valores obtidos de enviesamento na estimativa média dos parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ são inferiores quando se considera uma correlação de 0.9 entre o intercepto e o declive, e semelhantes, em ambos os modelos, para μ_α e μ_β .

Para $N=50$, o enviesamento nas estimativas médias dos parâmetros $\psi_{\alpha\beta}$ e $\psi_{\beta\beta}$ é maior quando se considera uma correlação de 0.9 e sendo o valor populacional de $\psi_{\beta\beta}=0.2$, mas se o valor populacional de $\psi_{\beta\beta}=4$, o enviesamento nas estimativas médias de $\psi_{\alpha\beta}$ e $\psi_{\beta\beta}$ é menor. Para os parâmetros μ_α e μ_β os valores de enviesamento obtidos são semelhantes, enquanto que para $\psi_{\alpha\alpha}$ são semelhantes quando o valor populacional de $\psi_{\beta\beta}=0.2$, e menor quando este valor populacional é 4.

Quanto aos valores do EQM para todos os parâmetros, tabelas 5.17 e 5.20, são menores no modelo em que a correlação entre o intercepto e o declive é menor, qualquer que seja a dimensão da amostra. De salientar que, os valores do EQM para o parâmetro μ_α nos modelos em comparação, são bastante próximos.

Os valores de coverage são em ambos os casos aceitáveis, mas se $N=50$, os parâmetros mais afetados são μ_β , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, se $N=1\ 000$, o parâmetro que apresenta valores de *coverage* mais baixos é $\psi_{\alpha\alpha}$ em ambos os casos.

Modelos A com $T = 4$ e correlação(α, β) = -0.5 vs Modelos A com $T = 4$ e correlação(α, β) = -0.9

A análise das tabelas 5.7 e 5.10 permite concluir que, para amostras grandes (1 000 observações) o enviesamento na estimativa média dos parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ são menores quando se considera um modelo com correlação entre o intercepto e o declive de -0.9. Para os parâmetros μ_α e μ_β os valores de enviesamento bruto obtido na sua estimativa média são análogos em ambos os modelos, com correlação 0.9 e com correlação 0.5.

Para amostras de pequena dimensão (50 observações) verifica-se que, quando a correlação considerada entre o intercepto e o declive é -0.9, os parâmetros μ_α , $\psi_{\alpha\beta}$ e $\psi_{\alpha\alpha}$ apresentam um enviesamento superior, e os parâmetros μ_β e $\psi_{\beta\beta}$ enviesamento menor, face aos resultados obtidos no mesmo modelo em que se considera uma correlação -0.5.

Os valores do EQM apresentados nas tabelas 5.8 e 5.11 permitem concluir que estes são menores nos modelos em que foi considerada uma correlação -0.9 entre intercepto e declive. Do conjunto dos parâmetros estimados μ_α é aquele que apresenta valores mais semelhantes nos modelos em comparação.

Os valores de coverage são em ambos os casos aceitáveis, mas se $N=50$, os parâmetros mais afetados são $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, se $N=1\ 000$, o parâmetro que apresenta valores de *coverage* mais baixos é $\psi_{\alpha\alpha}$ em ambos os casos.

Modelos A com $T = 4$ e fiabilidade dos indicadores de 0.4

No sentido de avaliar o efeito da fiabilidade na estimação dos parâmetros de um modelo com trajetória latente foram geradas amostras de dimensão 500 com distribuição normal, a partir dos modelos A1 e A4, para os quais se considerou valores de $\psi_{\alpha\beta}$ por forma a obter correlações entre o intercepto e o declive de -0.9, -0.5, 0, 0.5, 0.9. As variâncias dos termos residuais foram fixadas de modo a obter uma fiabilidade de 0.4 para cada uma das variáveis Y_1, Y_2, Y_3, Y_4 , que definem as observações em cada um dos momentos temporais do modelo.

Comparando os valores apresentados nas tabelas 5.7, 5.10, 5.13, 5.16 e 5.19 com os apresentados em 5.22 pode concluir-se que os enviesamentos nas estimativas médias dos parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ são menores nos modelos com maior fiabilidade dos indicadores. Os parâmetros μ_α e μ_β apresentam enviesamento nas estimativas médias semelhantes quando se varia o valor da fiabilidade dos indicadores, o que parece evidenciar que a estimativa média destes parâmetros não é influenciada por este fator.

Quanto ao EQM a comparação dos valores das tabelas 5.8, 5.11, 5.14, 5.17 e 5.20 com os valores de 5.23 permite verificar que o EQM é sempre superior, para os vários parâmetros do modelo, se a fiabilidade é menor. De referir que, por exemplo, para o modelo A1 ($\psi_{\beta\beta} = 0.2$) e correlação entre o intercepto e o declive de -0.5, o EQM para o parâmetro $\psi_{\alpha\beta}$, passa de 0.0027 para 0.0107, quase cinco vezes maior.

A análise da tabela 5.24 mostra que os valores de *coverage* para todos os parâmetros do modelo estão acima de 0.90, mas o parâmetro que apresenta menor valor de *coverage* é $\psi_{\alpha\alpha}$, em particular quando o parâmetro populacional $\psi_{\beta\beta}=0.2$, qualquer que seja a dimensão da amostra.

	$\psi_{\beta\beta}$	correlação(α, β)				
		-0.9	-0.5	0	0.5	0.9
$\psi_{\alpha\beta}$	0.2	0.0042 (-0.0105)	0.0094 (-0.0427)	0.0073	0.0082 (0.0373)	0.0087 (0.0218)
	4	0.0167 (-0.0093)	0.0197 (-0.0197)	0.0218	0.0225 (0.0225)	0.0224 (0.0124)
μ_α	0.2	0.0015	0.0019	0.0017	0.0018	0.0018
	4	0.0017	0.0018	0.0019	0.0019	0.0019
μ_β	0.2	-0.0004	-0.0003	-0.0001	0.0001	0.0003
	4	-0.0016	-0.0007	0.0003	0.0013	0.0020
$\psi_{\alpha\alpha}$	0.2	-0.0079 (-0.0079)	-0.0150 (-0.0150)	-0.0125 (-0.0125)	-0.0135 (-0.0135)	-0.0140 (-0.0140)
	4	-0.0228 (-0.0228)	-0.0288 (-0.0288)	-0.0321 (-0.0321)	-0.0323 (-0.0323)	-0.0310 (-0.0310)
$\psi_{\beta\beta}$	0.2	-0.0021 (-0.0105)	-0.0050 (-0.0250)	-0.0033 (-0.0165)	-0.0033 (-0.0165)	-0.0032 (-0.0160)
	4	-0.0136 (-0.0034)	-0.0135 (-0.0034)	-0.0117 (-0.0029)	-0.0078 (-0.0020)	-0.0035 (-0.0009)

Tabela 5.22: Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A, com fiabilidade dos indicadores de 0.4

	$\psi_{\beta\beta}$	correlação(α, β)				
		-0.9	-0.5	0	0.5	0.9
$\psi_{\alpha\beta}$	0.2	0.0027	0.0107	0.0071	0.0095	0.0115
	4	0.0569	0.0659	0.0791	0.0930	0.1037
μ_{α}	0.2	0.0031	0.0039	0.0038	0.0039	0.0040
	4	0.0041	0.0042	0.0042	0.0042	0.0043
μ_{β}	0.2	0.0007	0.0016	0.0014	0.0017	0.0020
	4	0.0111	0.0127	0.0141	0.0152	0.0158
$\psi_{\alpha\alpha}$	0.2	0.0140	0.0435	0.0331	0.0420	0.0486
	4	0.0623	0.1056	0.1565	0.1965	0.2174
$\psi_{\beta\beta}$	0.2	0.0007	0.0052	0.0036	0.0055	0.0071
	4	0.1378	0.1699	0.2082	0.2461	0.2783

Tabela 5.23: Erro quadrático médio para os parâmetros dos modelos A com fiabilidade dos indicadores de 0.4

	$\psi_{\beta\beta}$	correlação(α, β)				
		-0.9	-0.5	0	0.5	0.9
$\psi_{\alpha\beta}$	0.2	0.945	0.940	0.942	0.944	0.944
	4	0.962	0.957	0.949	0.944	0.939
μ_{α}	0.2	0.955	0.965	0.961	0.963	0.963
	4	0.966	0.966	0.967	0.968	0.967
μ_{β}	0.2	0.956	0.955	0.947	0.945	0.949
	4	0.957	0.948	0.945	0.936	0.939
$\psi_{\alpha\alpha}$	0.2	0.936	0.933	0.934	0.930	0.929
	4	0.947	0.944	0.947	0.943	0.934
$\psi_{\beta\beta}$	0.2	0.941	0.938	0.937	0.938	0.942
	4	0.952	0.948	0.953	0.949	0.948

Tabela 5.24: *Coverage* para os parâmetros dos modelos A com fiabilidade dos indicadores de 0.4

Modelos A1 e A4 com $T = 4$ vs Modelos A1 e A4 com $T = 3$

Com o objetivo de estudar o efeito de diminuir o número de momentos temporais na estimação de um modelo com trajetória latente foram gerados dados com distribuição normal a partir dos modelos A1 e A4, com 3 momentos temporais, em que a variância populacional do declive toma valores extremos 0.2 e 4, e a covariância entre o intercepto e o declive foi fixa de maneira a obter os níveis de correlação -0.9, -0.5, 0, 0.5 e 0.9. As variâncias dos termos residuais foram fixas a valores capazes de garantir uma fiabilidade dos indicadores de 0.5 em cada momento temporal. Posteriormente, foi estimado um modelo com trajetória latente com três momentos temporais tendo por base os dados assim gerados.

Os resultados obtidos para o enviesamento nas estimativas médias dos parâmetros do modelo, $\psi_{\alpha\beta}$, μ_{α} , μ_{β} , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ encontram-se nas tabelas F.1, F.2, F.3, F.4 e F.5, em anexo. Os valores destas tabelas devem ser comparados com os valores das tabelas 5.7, 5.10, 5.13, 5.16 e 5.19, respetivamente, tendo em consideração a mesma correlação entre o intercepto e o declive.

Assim, quando a correlação entre o intercepto e o declive considerada nos modelos A1 e A4 com 4 e 3 momentos temporais, é igual a 0.9 ou 0.5, o enviesamento nas estimativas médias de $\psi_{\alpha\beta}$, μ_{α} , μ_{β} e $\psi_{\alpha\alpha}$ aumenta, e diminui para $\psi_{\beta\beta}$, em amostras pequenas (50 observações), quando se comparam os resultados obtidos para o modelo com 3 momentos temporais face aos mesmos resultados para o modelo com 4 momentos temporais. Todavia, se a amostra é grande (1000 observações), o enviesamento obtido nas estimativas médias de $\psi_{\alpha\beta}$, μ_{β} (apenas se o valor populacional $\psi_{\beta\beta}=0.2$) e $\psi_{\alpha\alpha}$ diminui, aumentando para μ_{α} e $\psi_{\beta\beta}$, no modelo com 3 momentos temporais, tabelas 5.7 e 5.10 relativamente a F.1 e F.2, respetivamente.

De referir que, quando a correlação entre o intercepto e o declive é 0.5, o valor populacional de $\psi_{\beta\beta}=4$ e $N=50$, o enviesamento relativo na estimativa média de $\psi_{\alpha\beta}$ é -0.0415, valor no limite do negligenciável.

Se a correlação entre o intercepto e declive é igual a 0 verifica-se que o enviesamento nas estimativas médias dos parâmetros $\psi_{\alpha\beta}$, μ_{α} , μ_{β} e $\psi_{\alpha\alpha}$ aumenta, e diminui para $\psi_{\beta\beta}$, no modelo com 3 momentos temporais e para amostras de pequena dimensão (50 observações). Se a amostra é grande (1000 observações) então o enviesamento das estimativas médias de $\psi_{\alpha\beta}$, μ_{β} (apenas se o valor populacional $\psi_{\beta\beta}=4$) e $\psi_{\alpha\alpha}$ diminui, enquanto que aumenta para μ_{α} e mantém-se para $\psi_{\beta\beta}$, quando se comparam os resultados obtidos no modelo com 3 momentos temporais face aos resultados obtidos para o modelo com 4 momentos temporais.

Se a correlação entre o intercepto e o declive é igual a -0.5 e a dimensão da amostra pequena (50 observações), o enviesamento nas estimativas médias de $\psi_{\alpha\beta}$, μ_{β} , $\psi_{\alpha\alpha}$ aumenta, de $\psi_{\alpha\alpha}$ apenas no modelo A1 (valor populacional de $\psi_{\beta\beta} = 0.2$), no modelo com menor número de momentos temporais. Para $\psi_{\beta\beta}$ o enviesamento na estimativa média diminui.

Se a dimensão da amostra é grande (1000 observações) o enviesamento na estimativa média de $\psi_{\alpha\beta}$, μ_β (apenas no modelo A4), $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ diminui e aumenta o valor desta medida para μ_α .

Para uma correlação igual a -0.9 e amostra de dimensão pequena (50 observações) o enviesamento nas estimativas médias de $\psi_{\alpha\beta}$, μ_β e $\psi_{\alpha\alpha}$ (apenas no modelo A1) e $\psi_{\beta\beta}$ aumentam. O enviesamento na estimativa média de μ_α diminui, se o valor populacional de $\psi_{\beta\beta}=0.2$ e mantêm-se constante se $\psi_{\beta\beta}=4$. Todavia, se a dimensão da amostra é grande (1000 observações) os valores obtidos para esta mesma medida aumentam para μ_α e μ_β , para os restantes parâmetros do modelo diminui.

A comparação entre os valores obtidos de EQM para os mesmos modelos quando se consideram 3 e 4 momentos temporais, é feita analisando as tabelas 5.8 com F.10, 5.11 com F.9, 5.14 com F.8, 5.17 com F.7 e 5.20 com F.6, encontrando-se as tabelas F.10 a F.6 em anexo.

A análise destas tabelas permite verificar que o parâmetro para o qual aconteceu um maior aumento no EQM, com a diminuição do número de momentos temporais do modelo, é $\psi_{\beta\beta}$, seguido de $\psi_{\alpha\beta}$, no modelo A1. Por exemplo, para o modelo A1 com 3 momentos temporais e correlação 0.9, o EQM de estimação do parâmetro $\psi_{\beta\beta}$ aumenta mais de 6 vezes o valor obtido para o mesmo modelo com 4 momentos temporais. Os valores obtidos do EQM para o modelo A1 com 3 momentos temporais passam de 0.0370, 0.0069, 0.0036 e 0.0017 para 0.2237, 0.0461, 0.0222 e 0.0110, para amostras de dimensão 50, 250, 500 e 1000 observações, respetivamente. Do mesmo modo, o EQM de estimação do parâmetro $\psi_{\alpha\beta}$ no modelo A1 com 3 momentos temporais aumentou mais de 3 vezes, passam de 0.0573, 0.0113, 0.0059 e 0.0029 para 0.2196, 0.0426, 0.0277 e 0.0112, para as dimensões amostrais de 50, 250, 500 e 1000 observações, respetivamente.

Para os outros graus de correlação, 0.5, 0, -0.5 e -0.9, continua a verificar-se que os parâmetros para os quais os valores de EQM são mais afetados pela redução do número de momentos temporais nos modelos considerados, são o $\psi_{\beta\beta}$ e $\psi_{\alpha\beta}$ para o modelo A1. Quando a correlação é -0.9, também o parâmetro $\psi_{\alpha\alpha}$ mostra um aumento grande, quase 2 vezes mais para o modelo A1. Para $\psi_{\alpha\beta}$ o aumento é de quase 2 vezes, para o modelo A1, enquanto que $\psi_{\beta\beta}$ apresenta um aumento de quase 4 vezes mais, quando se considera o modelo A1. Por outro lado, o parâmetro μ_α não parece ser muito afetado pela diminuição do número de momentos temporais do modelo, uma vez que os respetivos valores do EQM são análogos, qualquer que seja o modelo, A1 ou A4, o grau de correlação e a dimensão da amostra considerada.

Quando se considera o modelo com três momentos temporais, o parâmetro que apresenta menor valor de *coverage*, em amostras pequenas (N=50), é μ_β , quando se considera uma correlação entre o intercepto e o declive de 0.9, 0.5, 0 e -0.5. Se a correlação considerada no modelo é -0.9, verifica-se que os parâmetros que apresentam menores valores de *coverage* são $\psi_{\alpha\alpha}$, quando o valor populacional $\psi_{\beta\beta}=0.2$ e para amostras de pequena di-

mensão, μ_β e $\psi_{\beta\beta}$, quando valor populacional $\psi_{\beta\beta}=4$ e em amostras de pequena dimensão.

Modelos A1 e A4 com $T = 4$ vs Modelos A1 e A4 com $T = 5$

Com o objetivo de estudar o efeito de aumentar o número de momentos temporais na estimação de um modelo com trajetória latente foram gerados dados com distribuição normal a partir dos modelos A1 e A4, em que a variância populacional do declive toma valores extremos 0.2 e 4, com cinco momentos temporais, e a covariância entre o intercepto e o declive a tomar valores por forma a obter os graus de correlação -0.9, -0.5, 0, 0.5 e 0.9. As variâncias dos termos residuais foram fixadas a valores capazes de assegurar uma fiabilidade de 0.5. Posteriormente, foi estimado um modelo com trajetória latente com cinco momentos temporais aos dados assim gerados.

Os resultados obtidos para o enviesamento nas estimativas médias dos parâmetros do modelo, $\psi_{\alpha\beta}$, μ_α , μ_β , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ e o erro quadrático médio foram comparados, com os obtidos para os mesmos modelos com 4 momentos temporais.

Para valores de correlação 0.5 e 0.9 e amostras de pequena dimensão (50 observações), tabelas H.4 e H.5, em anexo, verifica-se que o enviesamento nas estimativas médias de μ_α e $\psi_{\alpha\alpha}$ (o enviesamento para este parâmetro está no limite do negligenciável, se $\psi_{\beta\beta}=4$) aumentou no modelo com 5 momentos temporais face ao modelo com 4 momentos temporais, tabelas 5.16 e 5.19, respetivamente. Para os parâmetros μ_β e $\psi_{\beta\beta}$ este valor diminui. Para $\psi_{\alpha\beta}$ aumenta para o modelo A1 e diminui para o modelo A4. Em amostras de dimensão elevada (1 000 observações) os valores de $\psi_{\alpha\beta}$, μ_β e $\psi_{\alpha\alpha}$ diminuem e os valores de μ_α e $\psi_{\beta\beta}$ (apenas se $\psi_{\beta\beta}=0.2$) mantêm-se no modelo com 5 momentos temporais.

Para os modelos A1 e A4 com valores de correlação entre o intercepto e o declive igual a 0, tabela H.3 (T=5, em anexo) e tabela 5.13 (T=4), e amostra de pequena dimensão (50 observações) o enviesamento nas estimativas médias aumenta para os parâmetros $\psi_{\alpha\beta}$, μ_α e $\psi_{\alpha\alpha}$ (o enviesamento para este parâmetro está próximo de 0.05, se $\psi_{\beta\beta}=4$) e diminui para os restantes. Todavia, se a amostra é grande (1 000 observações) verifica-se que o enviesamento nas estimativas médias dos parâmetros $\psi_{\alpha\beta}$, μ_β e $\psi_{\alpha\alpha}$ diminuem, enquanto que para o parâmetro μ_α o valor mantém (no modelo A1) e para $\psi_{\beta\beta}$ diminui se o valor populacional $\psi_{\beta\beta}=0.2$, e aumenta, caso contrário.

Sendo a correlação entre o intercepto e o declive -0.5 e -0.9 e amostras de dimensão pequena (50 observações), o enviesamento nas estimativas médias de $\psi_{\alpha\beta}$, μ_α e $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ (apenas no modelo A1) aumenta, para este último parâmetro o valor de enviesamento diminui no modelo A4. Para μ_β o enviesamento na estimativa média diminui. Quando a correlação é 0.5 e N=250, os valores de enviesamento para os parâmetros $\psi_{\alpha\beta}$ e $\psi_{\alpha\alpha}$ estão próximos de 0.05, quando $\psi_{\beta\beta}=0.2$ e 4, respetivamente. Para amostras de dimensão elevada (1 000 observações) o enviesamento nas estimativas médias de $\psi_{\alpha\beta}$, μ_β (apenas no modelo A4, mantendo-se constante caso contrário), $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ (apenas no modelo A1) diminui. Para o parâmetro μ_α o enviesamento na sua estimativa média mantêm-se e para

μ_β também se considerado o modelo A1, diminuindo caso contrário, tabelas H.1 e H.2 relativamente a 5.10 e 5.7.

Os valores do EQM para os vários parâmetros dos modelos com 5 momentos temporais são apresentados nas tabelas H.6, H.7, H.8, H.9 e H.10, em anexo.

Para o modelo A1 com correlação igual a 0.9 (tabela H.10 em anexo) verifica-se que a maior redução no EQM, face aos mesmos valores para o modelo com 4 momentos temporais, acontece para os parâmetros $\psi_{\beta\beta}$ e $\psi_{\alpha\beta}$. O EQM para o parâmetro $\psi_{\beta\beta}$ reduz para 1/3 do valor inicial, para amostras pequenas (50 observações) e para amostras grandes (1 000 observações) reduz para 1/5 do valor encontrado para o modelo com 4 momentos temporais. Para o parâmetro $\psi_{\alpha\beta}$ o seu EQM de estimação diminui mais ou menos para metade, qualquer que seja a dimensão da amostra. Para o modelo A4 a maior redução ocorre nos valores do EQM de $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, de quase metade do valor obtido para o modelo com 4 momentos temporais. Os valores do EQM para μ_α mantiveram-se idênticos nos modelos A1 e A4 com 4 e 5 momentos temporais.

Conclusões semelhantes podem ser estabelecidas para os mesmos modelos com outros níveis de correlação entre o intercepto e o declive, embora que quando a correlação é igual a -0.5 e o modelo A4, também $\psi_{\alpha\alpha}$ é um dos parâmetros que apresenta maior redução no valor do EQM.

Assim, comparando as tabelas H.6, H.7, H.8, H.9 e H.10, em anexo, pode concluir-se que nos modelos com correlação positiva é maior a diminuição nos valores do EQM dos parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, quando o modelo tem maior número de momentos temporais. O EQM de estimação para o parâmetro μ_α parece não ser afetado pelo número de momentos temporais do modelo utilizado qualquer que seja o modelo, A1 ou A4, o grau de correlação ou a dimensão da amostra considerada.

A análise das tabelas H.11 a H.15, em anexo, permite verificar que os valores de *coverage* obtidos para os parâmetros $\psi_{\alpha\beta}$ e $\psi_{\alpha\alpha}$ encontram-se no limite do aceitável, 0.908 e 0.903, respetivamente, sendo a variância populacional $\psi_{\beta\beta}=0.2$, a correlação entre o intercepto e o declive -0.9 e a dimensão da amostra pequena (50 observações). A comparação entre as diferentes tabelas permite verificar que para amostras de pequena dimensão, os valores de *coverage* são inferiores e próximos do limite do aceitável, em modelos nos quais foi considerada uma correlação negativa ou inexistente entre o intercepto e o declive, -0.9, -0.5 e 0.

5.3.2 Dados gerados com distribuição normal - com omissões

Com o objetivo de avaliar o efeito da existência de omissões resultantes de um *planned missing design* em dados gerados com distribuição normal, foram considerados os modelos A1 e A4 com 3, 4 e 5 momentos temporais, nos quais se permitiram vários graus de

correlação entre o intercepto e o declive, -0.5, 0 e 0.5, dimensões amostrais de 250 e 500 observações e a existência de omissões em diferentes momentos temporais e em diferentes percentagens (padrões de omissão presentes no ICOR e cujo desenho é apresentado nas tabelas 5.1, 5.2 e 5.3). Assim, para o modelo com 4 momentos temporais foi considerada a existência de 50% de omissões no 1º momento temporal e 25% no 2º e 4º momentos temporais. Para o modelo com 3 momentos temporais foi considerada a existência de 25% de omissões no 1º e 3º momento temporal, enquanto que para o modelo com 5 momentos temporais foram consideradas 75% de omissões no 1º momento temporal, 50% no 2º e 25% no 3º e 5º. Amostras de dimensão 50 não foram consideradas, pois a existência de omissões em amostras desta dimensão, resultam num número muito diminuto de observações para posterior estimação de um modelo. Por outro lado, também não foram consideradas amostras de dimensão 1 000, porque se afigura como mais interessante estudar o efeito das omissões em amostras pequenas. Apresenta-se de seguida uma síntese dos principais resultados obtidos.

Modelos A1 e A4 com $T = 4$, correlação(α, β)= 0.5, 0 e -0.5 e omissões

A síntese dos resultados obtidos na estimação dos modelos que integram as 1 000 réplicas geradas a partir de um modelo com 4 momentos temporais são apresentados nas tabelas 5.25 a 5.27.

A análise destas tabelas permite verificar que o enviesamento na estimativa média dos parâmetros $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ é inferior no modelo A1 ($\psi_{\beta\beta} = 0-2$), quando se considera uma correlação entre o intercepto e o declive negativa, -0.5, qualquer que seja a dimensão da amostra. Para os parâmetros μ_{α} e μ_{β} o enviesamento na estimativa média não varia com a dimensão da amostra, nem com a correlação entre o intercepto e o declive considerada no modelo, pois os valores nos três modelos são semelhantes. Para o parâmetro $\psi_{\alpha\beta}$ o enviesamento é menor, no modelo A1, quando se considera a correlação igual a -0.5 face ao mesmo modelo com correlação 0.5.

Para o modelo A4, o parâmetro $\psi_{\alpha\alpha}$ mostra um menor enviesamento na estimativa média, quando o modelo considerado assume uma correlação positiva e sendo a amostra de dimensão pequena (250 observações). O parâmetro $\psi_{\alpha\beta}$ mostra um enviesamento menor, quando considerado um modelo A4 com correlação positiva, qualquer que seja a dimensão da amostra face ao mesmo modelo com correlação negativa. Para o modelo A4 com correlação 0, o enviesamento absoluto de $\psi_{\alpha\beta}$ é de 0.0077. Para $\psi_{\beta\beta}$ o enviesamento na estimativa média é menor quando se considera correlação 0 entre o intercepto e o declive. Se é considerada uma correlação de -0.5, o valor do enviesamento é de -0.1520, isto é, a estimativa média da variância do declive apresenta 15.20% de enviesamento face ao valor populacional, em amostras de dimensão 250 e um enviesamento de 6% se a dimensão da amostra é de 500 observações. Para μ_{α} os valores de enviesamento são semelhantes se $N=250$, mas se $N=500$ o modelo que apresenta um menor enviesamento na estimação deste parâmetro é o que considera correlação igual a -0.5.

Para o EQM a análise das tabelas 5.25, 5.26 e 5.27 permite verificar que os valores obtidos para esta medida são menores, quando a correlação considerada é negativa (-0.5), para os parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, para ambos os modelos, A1 e A4, qualquer que seja a dimensão da amostra. Os parâmetros μ_{α} e μ_{β} apresentam valores de EQM de estimação análogos, em ambos os modelos, A1 e A4, qualquer que seja a dimensão da amostra. Os valores obtidos de *coverage* para todos os parâmetros do modelo mostram-se dentro do limite do aceitável, isto é valores superiores a 0.90.

	$\psi_{\beta\beta}$	Enviesamento		EQM		Coverage	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
$\psi_{\alpha\beta}$	0.2	0.0023(0.0105)	0.0079 (0.0359)	0.0232	0.0115	0.949	0.951
	4	0.0077 (0.0077)	0.0258(0.0258)	0.2626	0.1298	0.942	0.954
μ_{α}	0.2	0.0049	0.0028	0.0111	0.0056	0.955	0.950
	4	0.0055	0.0028	0.0137	0.0068	0.954	0.954
μ_{β}	0.2	0.0003	0	0.0039	0.0020	0.935	0.952
	4	0.0038	0.0018	0.0287	0.0148	0.946	0.932
$\psi_{\alpha\alpha}$	0.2	-0.0075 (-0.0075)	-0.0108 (-0.0108)	0.0903	0.0464	0.950	0.948
	4	-0.0125 (-0.0125)	-0.0289 (-0.0289)	0.4373	0.2258	0.956	0.961
$\psi_{\beta\beta}$	0.2	-0.0042 (-0.0210)	-0.0051 (-0.0255)	0.0119	0.0063	0.939	0.947
	4	-0.0412 (-0.0103)	-0.0231 (-0.0058)	0.4650	0.2490	0.949	0.946

Tabela 5.25: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos A1 e A4, com correlação(α, β) = 0.5, $T = 4$ e omissões

	$\psi_{\beta\beta}$	Enviesamento		EQM		Coverage	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
$\psi_{\alpha\beta}$	0.2	0.0021	0.0066	0.0179	0.0090	0.949	0.949
	4	0.0105	0.0215	0.2421	0.1205	0.944	0.954
μ_{α}	0.2	0.0046	0.0024	0.0107	0.0054	0.952	0.949
	4	0.0054	0.0022	0.0142	0.0073	0.952	0.951
μ_{β}	0.2	-0.0001	-0.0001	0.0034	0.0017	0.940	0.950
	4	0.0015	0.0011	0.0274	0.0142	0.945	0.930
$\psi_{\alpha\alpha}$	0.2	-0.0062 (-0.0062)	-0.0098 (-0.0098)	0.0739	0.0377	0.953	0.953
	4	-0.0139 (-0.0139)	-0.0278 (-0.0278)	0.3739	0.1921	0.960	0.953
$\psi_{\beta\beta}$	0.2	-0.0030 (-0.0150)	-0.0041 (-0.0205)	0.0082	0.0044	0.953	0.951
	4	-0.0382 (-0.0096)	-0.0198 (-0.0050)	0.4198	0.2157	0.951	0.947

Tabela 5.26: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos A1 e A4, com correlação(α, β) = 0, $T = 4$ e omissões

Modelos A1 e A4 com $T = 4$, correlação(α, β)= 0.5, 0 e -0.5 com omissões vs sem omissões

Quando a correlação considerada entre o intercepto e o declive é 0.5 é possível verificar, analisando as tabelas 5.25 e 5.16, que o enviesamento nas estimativas médias dos

	$\psi_{\beta\beta}$	Enviesamento		EQM		Coverage	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
$\psi_{\alpha\beta}$	0.2	0.0018 (-0.0082)	0.0047 (-0.0214)	0.0119	0.0061	0.951	0.950
	4	0.0111 (-0.0111)	0.0133 (-0.0133)	0.2022	0.1004	0.946	0.957
μ_α	0.2	0.0042	0.0019	0.0096	0.0049	0.952	0.949
	4	0.0050	0.0014	0.0136	0.0071	0.943	0.944
μ_β	0.2	-0.0005	-0.0002	0.0026	0.0013	0.940	0.950
	4	-0.0010	0.0003	0.0255	0.0131	0.943	0.945
$\psi_{\alpha\alpha}$	0.2	-0.0047 (-0.0047)	-0.0079 (-0.0079)	0.0538	0.0271	0.957	0.952
	4	-0.0150 (-0.0150)	-0.0217 (-0.0217)	0.2687	0.1356	0.963	0.953
$\psi_{\beta\beta}$	0.2	-0.0018 (-0.0090)	-0.0027 (-0.0135)	0.0045	0.0024	0.952	0.956
	4	-0.0304 (-0.1520)	-0.0120 (-0.0600)	0.3672	0.1803	0.946	0.944

Tabela 5.27: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.5$, $T = 4$ e omissões

parâmetros $\psi_{\alpha\beta}$, μ_α , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ aumentou com a presença de omissões nos dados, para os modelos A1 e A4, qualquer que seja a dimensão da amostra, exceto para $\psi_{\alpha\alpha}$ em amostras de pequena dimensão (250 observações), o enviesamento reduz de -0.0105 para -0.0075, no modelo A1, e de -0.0192 para -0.0125, no modelo A4, e para $\psi_{\alpha\beta}$ em amostras de pequena dimensão (50 observações) e modelo A1, em que reduz de 0.0031 para 0.0023.

O enviesamento na estimativa média do parâmetro μ_β não parece ser influenciada pela existência de omissões nos dados. Para dados completos, o enviesamento é 0.0002, e é 0 para dados com omissões, modelo A1 e amostra de dimensão 500. Para o modelo A4 e amostra de dimensão 500, o enviesamento na estimativa média de μ_β é 0.0014 para dados completos, e 0.0018 para dados com omissões.

A comparação das tabelas 5.25 e 5.17 permite verificar que o erro quadrático médio de estimação aumentou com a existência de omissões, para o conjunto de todos os parâmetros do modelo.

Se para os modelos A1 e A4 é considerada uma correlação 0, então o enviesamento nas estimativas médias dos parâmetros aumenta se os dados considerados apresentam omissões. A exceção verifica-se para os parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e μ_β . Para os parâmetros $\psi_{\alpha\alpha}$ e $\psi_{\alpha\beta}$ o enviesamento apresenta uma diminuição, quando a dimensão da amostra é pequena (250 observações). Por exemplo, no modelo A1 o enviesamento na estimativa média de $\psi_{\alpha\beta}$ diminui de 0.0038, tabela 5.13, para 0.0021, tabela 5.26, quando se compara o modelo com dados completos em relação ao modelo com dados omissos. Para o parâmetro μ_β e em particular para o modelo A1, o enviesamento na estimativa média não se altera com a presença de omissões nos dados, sendo no caso completo de -0.0004 e no caso com omissões de -0.0001, para amostra de dimensão 250, e para amostras de dimensão 500 passa de 0 para -0.0001 no caso com omissões.

Os valores do EQM de estimação para o conjunto de todos os parâmetros do modelo,

aumentam com a existência de omissões nos dados, qualquer que seja o modelo e a dimensão amostral considerada, se a correlação considerada é 0, tabelas 5.26 e 5.14.

Quando se considera uma correlação entre o intercepto e o declive igual a -0.5, a comparação entre as tabelas 5.27 e 5.10 permite observar que não existirá qualquer efeito das omissões presentes nos dados, no enviesamento das estimativas médias dos parâmetros $\psi_{\alpha\beta}$, μ_α , μ_β e $\psi_{\alpha\alpha}$, sendo a dimensão da amostra de 500 observações. Apenas o enviesamento da estimativa média do parâmetro $\psi_{\beta\beta}$ aumenta de -0.0019 para -0.0027, no modelo A1 e de -0.0094 para -0.0120, no modelo A4. Todavia, se a amostra é de dimensão pequena (250 observações), o enviesamento nas estimativas médias dos parâmetros $\psi_{\alpha\beta}$ e $\psi_{\alpha\alpha}$ diminui em relação ao caso completo. Para $\psi_{\alpha\beta}$ passa de um valor de 0.0041 para 0.0018, no modelo A1 e de 0.0155 para 0.0111, no modelo A4. Para $\psi_{\alpha\alpha}$ reduz de -0.0097 para -0.0047, no modelo A1, e de -0.0211 para -0.0150, no modelo A4. Para os restantes parâmetros os valores obtidos mantêm-se idênticos ao obtido com dados completos.

A comparação das tabelas 5.27 e 5.11 permite concluir que se a correlação considerada é -0.5, então existe um aumento no erro quadrático médio de estimação dos parâmetros do modelo sempre que existem omissões nos dados.

Modelos A1 e A4 com $T = 3$, correlação(α, β)= 0.5, 0 e -0.5 e omissões

Nas tabelas F.16, F.17 e F.18, em anexo, são apresentados os resultados obtidos na estimação de um modelo com 3 momentos temporais com dados gerados a partir de um modelo com trajetória latente com igual número de momentos temporais, tendo considerado para este modelo três diferentes graus de correlação, 0.5, 0 e -0.5, e diferentes dimensões amostrais ($N=250$ e $N=500$).

A comparação dos resultados apresentados nestas tabelas permite concluir que o enviesamento nas estimativas médias dos parâmetros $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ é menor nos modelos em que se considera uma correlação -0.5, qualquer que seja o modelo, A1 ou A4 (isto é, qualquer que seja o valor populacional de $\psi_{\beta\beta}$, 0.2 ou 4), e qualquer que seja a dimensão da amostra. Para os parâmetros μ_α e μ_β o enviesamento absoluto é semelhante nos dois modelos e qualquer que seja a dimensão da amostra, exceto para amostras pequenas (250 observações) o enviesamento na estimativa média de μ_β é menor no modelo A4 ($\psi_{\beta\beta}=4$) com correlação entre o intercepto e o declive igual a 0.5.

Para o parâmetro $\psi_{\alpha\beta}$, o enviesamento é menor no modelo A4 com correlação 0.5, qualquer que seja a dimensão da amostra. Todavia, para o modelo A1 o enviesamento na estimativa média deste parâmetro, em amostras pequenas, é menor no modelo no qual se considera correlação -0.5, enquanto que para amostras com 500 observações, o enviesamento face ao seu valor populacional é maior no modelo que considera uma correlação entre o intercepto e o declive negativa, -0.5.

De referir ainda que os valores de enviesamento para o parâmetro $\psi_{\alpha\beta}$, quando o valor

populacional $\psi_{\beta\beta}=0.2$ e $N=250$, são superiores ao negligenciável, 6.86% e 5.64%, nos modelos com correlação entre o intercepto e o declive, positiva ou negativa, respetivamente.

Por outro lado, o parâmetro $\psi_{\alpha\alpha}$ apresenta valores de enviesamento no limite do negligenciável, -0.0492 e -0.0452, quando $N=250$ e o parâmetro populacional $\psi_{\beta\beta}=4$, qualquer que seja o grau de correlação entre o intercepto e o declive considerado. O parâmetro $\psi_{\beta\beta}$, no modelo A1, também apresenta valores de enviesamento acima do desejável, -0.0840 (8.4%), -0.0730 (7.3%) e -0.0560 (5.6%), qualquer que seja a correlação considerada, positiva, negativa ou inexistente.

Quanto aos valores do EQM de estimação dos parâmetros $\psi_{\alpha\beta}$, μ_β , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, estes são menores quando se consideram modelos com correlação -0.5, qualquer que seja a dimensão amostral. A exceção é o parâmetro μ_β que para o modelo A1 e sendo a amostra de dimensão elevada ($N=500$), apresenta um EQM de estimação semelhante, independentemente do grau de correlação considerado no modelo. Para o parâmetro μ_α estes valores são semelhantes nos vários modelos, independentemente do valor que a correlação entre o intercepto e o declive assume.

Os valores de *coverage* obtidos para o conjunto dos parâmetros do modelo apresentaram-se superiores a 0.90, o limite do aceitável, em todas as situações em análise.

	$\psi_{\beta\beta}$	Enviesamento		EQM		Coverage	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
$\psi_{\alpha\beta}$	0.2	0.0151 (0.0686)	0.0037 (0.0168)	0.0615	0.0309	0.959	0.950
	4	0.0364 (0.0364)	0.0074 (0.0074)	0.4936	0.2462	0.950	0.941
μ_α	0.2	-0.0042	-0.0002	0.0093	0.0048	0.946	0.947
	4	-0.0043	-0.0001	0.0101	0.0052	0.944	0.942
μ_β	0.2	0.0034	0.0008	0.0060	0.0029	0.950	0.956
	4	0.0051	0.0014	0.0342	0.0172	0.952	0.952
$\psi_{\alpha\alpha}$	0.2	-0.0244 (-0.0244)	-0.0072 (-0.0072)	0.1186	0.0599	0.952	0.950
	4	-0.0500 (-0.0500)	-0.0122 (-0.0122)	0.6539	0.3277	0.959	0.952
$\psi_{\beta\beta}$	0.2	-0.0168 (-0.0840)	-0.0054 (-0.0270)	0.0593	0.0280	0.941	0.952
	4	-0.0671 (-0.0168)	-0.0250 (-0.0063)	1.0053	0.4707	0.942	0.947

Tabela 5.28: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos A1 e A4, com correlação(α, β) = 0.5, $T = 3$ e omissões

Modelos A1 e A4 com $T = 3$, correlação(α, β) = 0.5, 0 e -0.5 com omissões vs sem omissões

Os valores do enviesamento na estimativa média e do EQM obtidos para os dados gerados com omissões que constituem um *planned missing design*, tabelas F.16, F.17 e F.18, em anexo, devem ser comparados com os valores obtidos para os dados completos, tabelas F.4, F.3 e F.2, em anexo, no sentido de avaliar o efeito de dados omissos em

	$\psi_{\beta\beta}$	Enviesamento		EQM		Coverage	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
$\psi_{\alpha\beta}$	0.2	0.0144	0.0037	0.0444	0.0224	0.959	0.946
	4	0.0418	0.0098	0.4361	0.2156	0.941	0.937
μ_{α}	0.2	-0.0042	-0.0003	0.0091	0.0047	0.945	0.945
	4	-0.0045	-0.0003	0.0104	0.0053	0.943	0.941
μ_{β}	0.2	0.0035	0.0008	0.0051	0.0025	0.942	0.959
	4	0.0062	0.0013	0.0315	0.0156	0.948	0.956
$\psi_{\alpha\alpha}$	0.2	0.0222 (0.0222)	0.0067 (0.0067)	0.0895	0.0454	0.947	0.949
	4	-0.0492 (-0.0492)	-0.0121 (-0.0121)	0.5382	0.2696	0.953	0.952
$\psi_{\beta\beta}$	0.2	0.0146 (0.0730)	-0.0045 (-0.0225)	0.0401	0.0191	0.944	0.957
	4	-0.0629 (-0.0157)	-0.0219 (-0.0055)	0.8187	0.3783	0.938	0.950

Tabela 5.29: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos A1 e A4, com correlação(α, β) = 0, $T = 3$ e omissões

	$\psi_{\beta\beta}$	Enviesamento		EQM		Coverage	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
$\psi_{\alpha\beta}$	0.2	0.0124 (-0.0564)	0.0044 (-0.0200)	0.0260	0.0132	0.955	0.946
	4	0.0427 (-0.0427)	0.0105 (-0.0105)	0.3454	0.1691	0.936	0.944
μ_{α}	0.2	-0.0041	-0.0005	0.0086	0.0045	0.945	0.946
	4	-0.0045	-0.0005	0.0103	0.0052	0.945	0.941
μ_{β}	0.2	0.0033	0.0008	0.0039	0.0019	0.945	0.952
	4	0.0070	0.0011	0.0284	0.0137	0.947	0.948
$\psi_{\alpha\alpha}$	0.2	-0.0188 (-0.0188)	-0.0059 (-0.0059)	0.0587	0.0300	0.949	0.952
	4	-0.0452 (-0.0452)	-0.0108 (-0.0108)	0.3865	0.1937	0.959	0.952
$\psi_{\beta\beta}$	0.2	-0.0112 (-0.0560)	-0.0034 (-0.0170)	0.0210	0.0101	0.940	0.950
	4	-0.0554 (-0.0139)	-0.0168 (-0.0042)	0.6210	0.2827	0.940	0.953

Tabela 5.30: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos A1 e A4, com correlação(α, β) = -0.5, $T = 3$ e omissões

modelos com 3 momentos temporais.

Esta comparação permite concluir que qualquer que seja a correlação considerada no modelo, o enviesamento na estimativa média de todos os parâmetros aumenta quando se consideram dados com omissões e sendo a amostra de pequena dimensão ($N=250$). Para amostras de maior dimensão ($N=500$) o enviesamento na estimativa média dos parâmetros μ_α e μ_β mantem-se constante.

De salientar que, a existência de omissões nos dados gerados provocou enviesamentos na estimativa média dos parâmetros $\psi_{\alpha\beta}$ e $\psi_{\beta\beta}$, sendo o parâmetro populacional $\psi_{\beta\beta}=0.2$, acima do considerado negligenciável, enquanto que para $\psi_{\alpha\alpha}$ (se $\psi_{\beta\beta}=4$) se encontra no limite do considerado negligenciável.

Os valores do EQM de estimação dos parâmetros do modelo aumentam, quando se consideram dados com omissões, qualquer que seja o modelo considerado, o grau de correlação e a dimensão da amostra, sendo os parâmetros mais afetados neste aumento $\psi_{\alpha\beta}$ e $\psi_{\alpha\alpha}$ no modelo A4 e $\psi_{\beta\beta}$ no modelo A1, qualquer que seja o grau de correlação entre o intercepto e o declive. O aumento no EQM de estimação de $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ é de quase duas vezes mais do que o valor obtido para dados sem omissões.

Modelos A1 e A4 com $T = 5$, correlação(α, β)= $0.5, 0$ e -0.5 e omissões

Nas tabelas H.16, H.17 e H.18, em anexo, são apresentados os resultados obtidos na estimação de um modelo com 5 momentos temporais com dados gerados a partir de um modelo com trajetória latente com igual número de momentos temporais, considerando para este modelo três diferentes graus de correlação, 0.5, 0 e -0.5, e diferentes dimensões amostrais ($N=250$ e $N=500$).

Quando se comparam os resultados obtidos para os modelos A1 e A4 com dados omissos e com diferentes graus de correlação, tabelas H.16, H.17 e H.18, em anexo, é possível concluir que para amostras pequenas (250 observações) o enviesamento nas estimativas médias dos parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$, $\psi_{\beta\beta}$ e μ_β é inferior no modelo com correlação -0.5, mas para μ_β apenas no modelo A4, uma vez que para o modelo A1, o enviesamento na estimativa média de μ_β apresenta valores idênticos independentemente do valor considerado para correlação entre o intercepto e o declive.

Para o parâmetro μ_α o enviesamento é maior no modelo em que se considera correlação igual a -0.5. Para amostras de dimensão mais elevada (500 observações), verifica-se que o enviesamento nas estimativas médias dos parâmetros $\psi_{\alpha\beta}$, μ_α , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ (estes dois últimos apenas para o modelo A1) é mais pequeno nos modelos em que se considera uma correlação entre o intercepto e o declive é de -0.5. O parâmetro $\psi_{\alpha\alpha}$ apresenta enviesamento, face ao seu valor populacional, menor quando a correlação é 0, e $\psi_{\beta\beta}$ quando a correlação é 0.5, sendo o valor populacional de $\psi_{\beta\beta}=4$ (modelo A4). O parâmetro μ_β apresenta valores de enviesamento na sua estimativa média que são independentes do grau

de correlação entre o intercepto e o declive do modelo considerado.

A análise dos valores do EQM apresentados nas tabelas H.16, H.17 e H.18, em anexo, permite verificar que nos modelos em que se considera a correlação entre o intercepto e o declive é igual a -0.5, estes valores são menores para a totalidade dos parâmetros do modelo, independentemente da dimensão amostral considerada.

Os valores de *coverage* para todos os parâmetros do modelo mantêm-se aceitáveis, apesar das omissões existentes nos dados.

Modelos A1 e A4 com $T = 5$, correlação $(\alpha, \beta) = 0.5, 0$ e -0.5 com omissões *vs* sem omissões

Da comparação dos resultados apresentados nas tabelas H.16, H.17 e H.18, em anexo, com os resultados apresentados nas tabelas H.9, H.8 e H.7, em anexo, isto é, comparando o enviesamento nas estimativas médias dos parâmetros obtidas para dados com omissões com a mesma medida para dados completos, conclui-se que para os parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ o enviesamento nas estimativas médias obtidas é superior quando os dados apresentam omissões, qualquer que seja a dimensão da amostra e a correlação considerada. Para o parâmetro μ_{β} o enviesamento na estimativa média é semelhante em ambas as situações, quando a dimensão da amostra é grande ($N=500$). Se a dimensão da amostra é de 250 observações, o enviesamento na estimativa média de μ_{β} é inferior na situação em que os dados apresentam omissões. Qualquer que seja a dimensão da amostra ($N=250$ ou $N=500$), aumenta o enviesamento na estimativa média de μ_{α} quando existem omissões nos dados, exceto quando o valor populacional $\psi_{\beta\beta}=4$, correlação -0.5 e $N=500$.

De referir ainda que, os valores de enviesamento na estimativa média dos vários parâmetros do modelo, obtidos na estimação dos vários modelos com dados gerados com omissões, apresenta valores inferiores a 0.05.

Os valores do EQM de estimação dos parâmetros para dados com omissões são superiores aos correspondentes valores para dados completos, tabelas H.16, H.17 e H.18, em anexo, *versus* H.9, H.8 e H.7, respetivamente.

Os parâmetros que apresentam um maior aumento no EQM de estimação são $\psi_{\alpha\beta}$ e $\psi_{\alpha\alpha}$, apenas para o modelo A4. Por exemplo, o valor do EQM de estimação de $\psi_{\alpha\alpha}$ para o modelo A4 aumenta mais de 5 vezes, se a correlação entre o intercepto e o declive é 0.5 ou 0 e quase 5 vezes mais, se a correlação é -0.5, quando se consideram dados omissos. Para $\psi_{\alpha\beta}$ o aumento é de quase 5 vezes mais, se a correlação entre o intercepto e o declive é 0.5 ou 0, e de 4 vezes mais se a correlação é -0.5. O μ_{β} é o parâmetro cujo EQM menos aumenta, com a existência de omissões nos dados, em particular quando o modelo considerado é o A4, e existindo correlação entre o intercepto e o declive.

5.4 Resultados do estudo de simulação para dados gerados a partir de uma mistura de distribuições normais

Neste estudo o modelo com trajetória latente é examinado sob diferentes condições: 1) dados resultantes de uma mistura de distribuições normais, sem omissões, diferentes dimensões amostrais e diferente número de momentos temporais do modelo; 2) dados resultantes de uma mistura de distribuições normais, sem omissões, diferentes dimensões amostrais e diferentes métodos de estimação; 3) dados resultantes de uma mistura de distribuições normais, com omissões que constituem um *planned missing design*, diferentes dimensões amostrais e modelo com diferente número de momentos temporais. Na presença de dados resultantes de uma mistura de distribuições normais, o método de estimação utilizado foi o método da máxima verossimilhança robusto (MLR), mas sempre que nos dados gerados se permite a existência de omissões, a estimação de um modelo com trajetória latente é realizada recorrendo ao método do *full information maximum likelihood*. Considerou-se ainda, o método de estimação da máxima verossimilhança (ML) com o intuito de avaliar o efeito, na estimação de um modelo com trajetória latente a dados que não apresentam distribuição normal, da utilização de um estimador que não é robusto.

Tal como quando se geraram amostras de dados com distribuição normal, também ao gerar como uma mistura de distribuições normais com e sem omissões, ocorreram problemas de estimação devido à existência de variâncias negativas, em amostras com dimensão de 50 observações, aumentando a percentagem de amostras nesta situação se o número de momentos temporais diminui.

5.4.1 Dados gerados a partir de uma mistura de distribuições normais - completos

Com o objetivo de estudar o efeito da não normalidade da distribuição dos dados na estimação de um modelo com trajetória latente foram considerados 5 diferentes modelos de mistura, H1 a H5 (ver especificações na tabela 5.5), a partir dos quais são gerados dados com distintos graus de desvio face à normalidade.

O pacote estatístico Mplus gera dados resultantes de uma mistura de distribuições normais recorrendo a uma junção de dados gerados a partir de duas classes, sendo que em cada uma destas classes os dados apresentam uma distribuição normal. Deste modo, foram considerados os modelos de mistura H1 a H5 (conforme especificado na tabela 5.5), com duas classes, uma maioritária, correspondendo a 88% das observações da amostra, que apresenta distribuição normal e uma classe minoritária, correspondendo a 12% das observações da amostra, também com distribuição normal, mas diferentes parâmetros caracterizadores. Os dados foram gerados de acordo com a mistura definida em cada modelo induzindo diferentes graus de achatamento e de assimetria positiva. Após a geração de

1 000 réplicas os dados são considerados como pertencendo a apenas uma classe e para cada amostra é estimado um modelo com trajetória latente, cujos valores estabelecidos para os parâmetros são previamente obtidos na estimação de um modelo com apenas uma réplica de dimensão elevada, isto é, 100 000 observações. Dado a dimensão elevada desta amostra é possível assumir os valores estimados para o conjunto dos parâmetros do modelo como valores populacionais.

O painel esquerdo das figuras 5.1 a 5.10 mostra as distribuições univariadas para cada uma das quatro variáveis, Y_1 , Y_2 , Y_3 e Y_4 , geradas por cada um dos modelos H1 a H5, para amostras de dimensão 1 000, respetivamente. Nestes gráficos é possível observar diferentes graus de assimetria positiva, bem como diferentes graus de achatamento. Para os dados gerados a partir do modelo H1, gráfico 5.1, pode observar-se que as variáveis Y_1 , Y_2 , Y_3 e Y_4 , apresentam uma assimetria e achatamento fracos. Os dados gerados a partir do modelo H2 mostram uma assimetria fraca a suave e um achatamento suave a moderado, figura 5.3. Por outro lado, os dados gerados a partir do modelo H3 têm uma assimetria suave a moderada e um achatamento moderado, figura 5.5, enquanto que os dados obtidos a partir de H4 apresentam assimetria e achatamento moderados, figura 5.7. Os dados gerados a partir do modelo H5 apresentam assimetria e achatamento severos, figura 5.9.

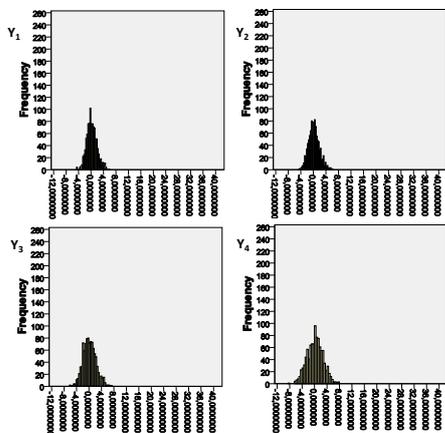


Figura 5.1: Distribuição univariada das variáveis Y_1 , Y_2 , Y_3 e Y_4 , na primeira amostra gerada a partir do modelo H1

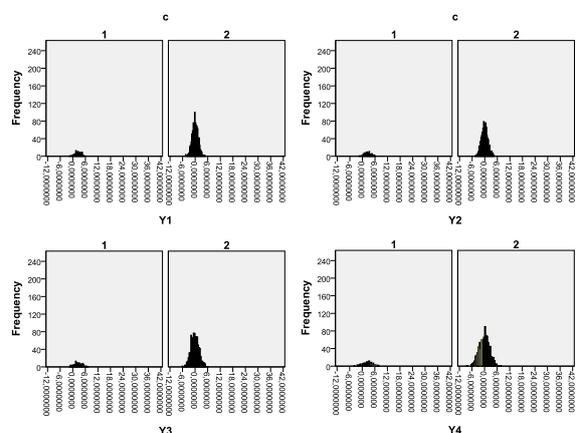


Figura 5.2: Distribuição univariada das variáveis Y_1 , Y_2 , Y_3 e Y_4 , em cada uma das classes definidas pelo modelo H1

O painel direito das figuras 5.1 a 5.10 mostra para cada um dos cinco modelos, H1 a H5, a distribuição dos dados em cada uma das classes considerada na geração de dados não normais, minoritária e maioritária. Em todos estes gráficos é possível observar que os dados gerados em cada uma das classes, apresentam uma distribuição normal, embora com diferentes parâmetros caracterizadores.

A tabela 5.31 apresenta para estes cinco diferentes modelos a assimetria e achatamento resultantes para as quatro variáveis univariadas, Y_1 , Y_2 , Y_3 e Y_4 , sendo estas medidas con-

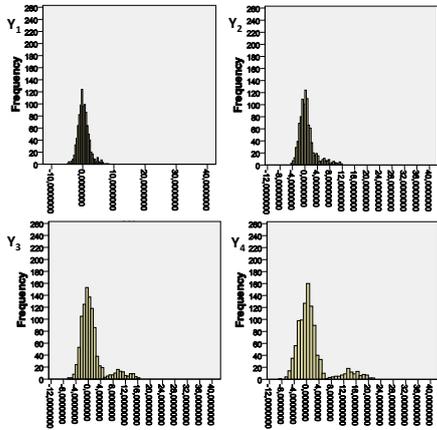


Figura 5.3: Distribuição univariada das variáveis Y_1 , Y_2 , Y_3 e Y_4 , na primeira amostra gerada a partir do modelo H2

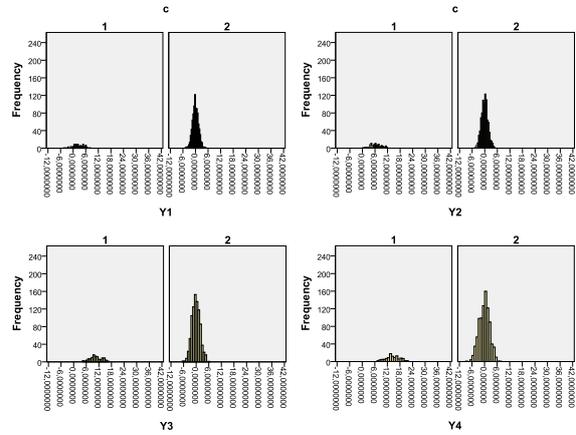


Figura 5.4: Distribuição univariada das variáveis Y_1 , Y_2 , Y_3 e Y_4 , em cada uma das classes definidas pelo modelo H2

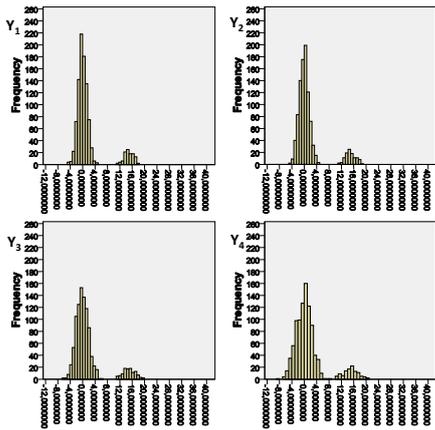


Figura 5.5: Distribuição univariada das variáveis Y_1 , Y_2 , Y_3 e Y_4 , na primeira amostra gerada a partir do modelo H3

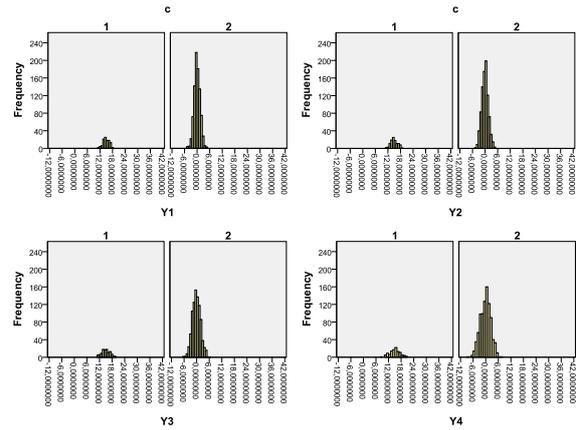


Figura 5.6: Distribuição univariada das variáveis Y_1 , Y_2 , Y_3 e Y_4 , em cada uma das classes definidas pelo modelo H3

sequência das combinações dos valores assumidos pelos parâmetros do modelo a partir do qual os dados são gerados. Os valores de assimetria e achatamento são calculados para uma amostra de dimensão muito elevada, $N = 100\,000$, sendo possível considerarem-se como os verdadeiros valores da distribuição destes dados. Para os dados gerados a partir do modelo H1 as variáveis univariadas apresentam valores de assimetria entre 0.10 e 0.29, enquanto que o achatamento toma valores entre 0.05 e 0.25. Para os dados gerados a partir de H2, a assimetria varia entre 0.82 e 1.75, e o achatamento entre 1.99 e 2.95. Para o modelo H3 os valores de assimetria e achatamento variam entre 1.70 e 2.06, e 2.23 e 2.89, respetivamente, enquanto que, para o modelo H4 estas medidas tomam valores entre 2.12 e 2.18, e 3.13 e 3.44, respetivamente. Para o modelo H5 obtêm-se valores de assimetria entre 2.31 e 2.65 e de achatamento no intervalo 4.35 a 6.78.

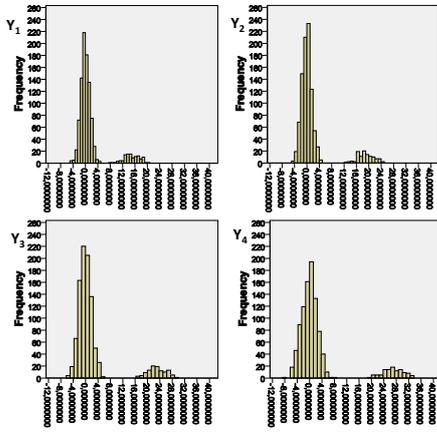


Figura 5.7: Distribuição univariada das variáveis Y_1 , Y_2 , Y_3 e Y_4 , na primeira amostra gerada a partir do modelo H4

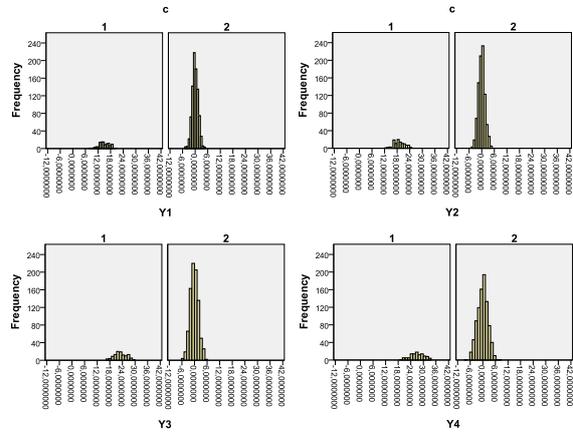


Figura 5.8: Distribuição univariada das variáveis Y_1 , Y_2 , Y_3 e Y_4 , em cada uma das classes definidas pelo modelo H4

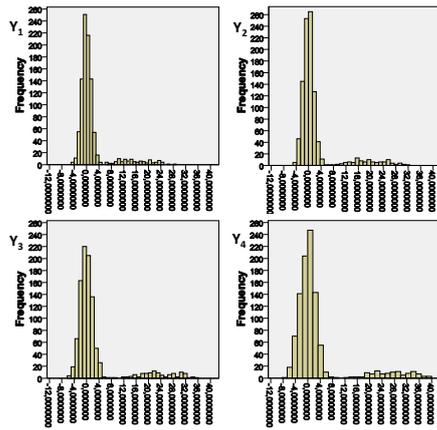


Figura 5.9: Distribuição univariada das variáveis Y_1 , Y_2 , Y_3 e Y_4 , na primeira amostra gerada a partir do modelo H5

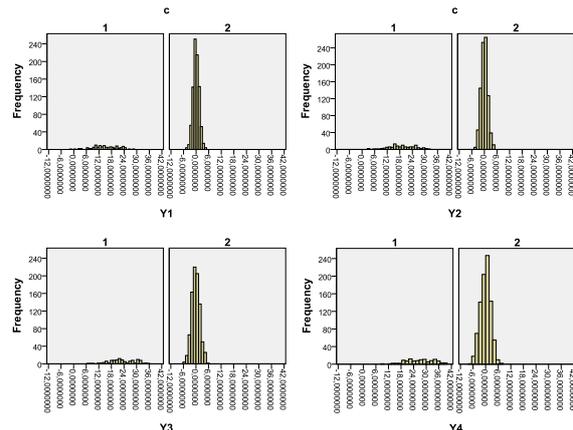


Figura 5.10: Distribuição univariada das variáveis Y_1 , Y_2 , Y_3 e Y_4 , em cada uma das classes definidas pelo modelo H5

Modelos H1 a H5 com $T = 4$ e estimação MLR

Na tabela 5.32 são apresentados os valores de enviesamento absoluto e relativo que se verificam na estimativa média dos parâmetros do modelo, quando para os dados gerados a partir dos modelos H1 a H5 é estimado um modelo com trajetória latente, cujos parâmetros populacionais são obtidos gerando uma amostra de 100 000 observações.

Os dados apresentados nesta tabela mostram que o parâmetro $\psi_{\alpha\beta}$ é aquele que apresenta maior valor de enviesamento relativo na sua estimativa média, -9, -7.6, -7.2 e 4.6, qualquer que seja a dimensão da amostra considerada (50 a 1 000 observações), quando os dados são gerados a partir do modelo H1, isto é, dados que apresentam assimetria e achatamento fracos. Também, o parâmetro μ_β apresenta um enviesamento relativo eleva-

Modelo	Medidas	Y_1	Y_2	Y_3	Y_4
H1	assimetria	0.290754	0.246025	0.147609	0.095011
	achatamento	0.250411	0.171933	0.96977	0.048937
H2	assimetria	0.815137	1.539574	1.703300	1.746838
	achatamento	1.994921	2.949045	2.776742	2.631650
H3	assimetria	2.058081	2.015224	1.882022	1.696652
	achatamento	2.892773	2.808353	2.564597	2.234632
H4	assimetria	2.145613	2.177906	2.154056	2.119419
	achatamento	3.436419	3.361090	3.238353	3.125429
H5	assimetria	2.654646	2.535176	2.411996	2.312204
	achatamento	6.778189	5.705369	4.903543	4.353472

Tabela 5.31: Valores da assimetria e achatamento para cada uma das variáveis univariadas geradas pelos modelos H1 a H5 (com 100 000 observações)

do, 2.6316, -1.2105, -1 e -1.2632, para N a variar entre 50 a 1 000 observações. Os restantes parâmetros apresentam valores de enviesamento relativo abaixo de 0.05, o recomendado na literatura. Todavia, se $N=50$ o enviesamento relativo na estimativa média de $\psi_{\beta\beta}$ encontra-se no limite do negligenciável, -0.0455.

Quando os dados são gerados a partir do modelo H2, dados que obtidos como uma mistura de distribuições não normais com assimetria fraca a suave e achatamento suave a moderado, observa-se que o enviesamento relativo nas estimativas médias dos diferentes parâmetros do modelo apresenta valores abaixo de 0.05.

Quando os dados são gerados a partir do modelo H3, isto é, dados resultantes de uma mistura de distribuições normais com assimetria suave a moderada e achatamento moderado, verifica-se que os valores de enviesamento relativo para a estimativa média de $\psi_{\alpha\beta}$ e μ_β são superiores a 0.05. O parâmetro $\psi_{\alpha\beta}$ apresenta valores de 2.3125, -0.5625, -1.7708 e -2.4792, enquanto que μ_β apresenta valores de 2.6313, -1.2105, -1 e -1.2632, sendo $N=50$, 250, 500 e 1 000, respetivamente. O enviesamento relativo na estimativa média de $\psi_{\beta\beta}$, sendo $N=50$, é de -0.0489, valor considerado no limite do negligenciável.

Quando os dados são gerados a partir dos modelos H4 ou H5, isto é, dados provenientes de uma mistura de distribuições normais, com assimetria e achatamento moderados ou severos, respetivamente, os resultados apresentados na tabela 5.32 permitem verificar que nenhum dos parâmetros do modelo apresenta enviesamento acima de 0.05, limite do negligenciável. Todavia, para $N=50$, os parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ são os que apresentam valores de enviesamento mais elevados, em ambos os modelos.

De salientar ainda que, se a dimensão da amostra é pequena ($N=50$), o parâmetro μ_α apresenta um enviesamento relativo na sua estimativa média semelhante, quando se comparam os valores obtidos para os dados gerados a partir dos modelos H1 e H2, e dos modelos H3, H4 e H5. O parâmetro μ_β apresenta valores idênticos para H1 e H3, e para H2, H4 e H5. O parâmetro $\psi_{\alpha\alpha}$ apresenta valores idênticos para H3, H4 e H5. O

parâmetro $\psi_{\beta\beta}$ apresenta valores idênticos para H1 e H3, e para H2, H4 e H5.

Modelo	Parâmetro	N=50	N=250	N=500	N = 1 000
H1	$\psi_{\alpha\beta}$	0.0045 (-9)	0.0038 (-7.6)	0.0036 (-7.2)	0.0023 (-4.6)
	μ_{α}	-0.0054 (-0.0177)	-0.0011 (0.0036)	-0.0035 (-0.0115)	-0.0052 (-0.0171)
	μ_{β}	0.0050 (2.6316)	-0.0023 (-1.2105)	-0.0019 (-1)	-0.0024(-1.2632)
	$\psi_{\alpha\alpha}$	-0.0357 (-0.0213)	-0.0271 (-0.0162)	-0.0225 (-0.0135)	-0.0208 (-0.0120)
	$\psi_{\beta\beta}$	-0.0092 (-0.0455)	-0.0048 (-0.0237)	-0.0038 (-0.0188)	-0.0040 (-0.0198)
H2	$\psi_{\alpha\beta}$	-0.0226 (-0.0212)	-0.0133 (-0.0125)	-0.0126 (-0.0118)	-0.0116 (-0.0109)
	μ_{α}	-0.0056 (-0.0184)	-0.0002 (-0.0007)	-0.0044 (-0.0144)	-0.0058 (-0.0190)
	μ_{β}	0 (0)	-0.0050 (-0.0103)	-0.0080 (-0.0164)	-0.0084 (-0.0173)
	$\psi_{\alpha\alpha}$	-0.0381 (-0.0175)	-0.04655 (-0.0214)	-0.0430 (-0.0198)	-0.0408 (-0.0188)
	$\psi_{\beta\beta}$	-0.0509 (-0.0266)	-0.0270 (-0.0141)	-0.0360 (-0.0188)	-0.0353 (-0.0184)
H3	$\psi_{\alpha\beta}$	0.0111 (2.3125)	-0.0027 (-0.5625)	-0.0085 (-1.7708)	-0.0119 (-2.4792)
	μ_{α}	-0.0203 (-0.0112)	-0.0072 (-0.0040)	-0.0226 (-0.0124)	-0.0234 (-0.0129)
	μ_{β}	0.0050 (2.6313)	-0.0023 (-1.2105)	-0.0019 (-1)	-0.0024 (-1.2632)
	$\psi_{\alpha\alpha}$	-0.6697 (-0.0268)	-0.2606 (-0.0104)	-0.3250 (-0.0130)	-0.2874 (-0.0115)
	$\psi_{\beta\beta}$	-0.0100 (-0.0489)	-0.0060 (-0.0293)	-0.0054 (-0.0264)	-0.0055 (-0.0269)
H4	$\psi_{\alpha\beta}$	-0.1616 (-0.0252)	-0.0686 (-0.0107)	-0.0945 (-0.0148)	-0.0883 (-0.0138)
	μ_{α}	-0.0201 (-0.0110)	-0.0084 (-0.0046)	-0.0234 (-0.0129)	-0.0240 (-0.0132)
	μ_{β}	0.0001 (0.0002)	-0.0050 (-0.0103)	-0.0080 (-0.0164)	-0.0083 (-0.0171)
	$\psi_{\alpha\alpha}$	-0.6623 (-0.0260)	-0.3015 (-0.0002)	-0.3609 (-0.0142)	-0.3184 (-0.0125)
	$\psi_{\beta\beta}$	-0.0522 (-0.0273)	-0.0277 (-0.0145)	-0.0373 (-0.0195)	-0.0365 (-0.0191)
H5	$\psi_{\alpha\beta}$	-0.1580 (-0.0246)	-0.0862 (-0.0134)	-0.1077 (-0.0168)	-0.0989 (-0.0154)
	μ_{α}	-0.0197 (-0.0108)	-0.0121 (-0.0066)	-0.0261 (-0.0143)	-0.0259 (-0.0142)
	μ_{β}	0.0001 (0.0002)	-0.0051 (-0.0105)	-0.0081 (-0.0166)	-0.0085 (-0.0175)
	$\psi_{\alpha\alpha}$	-0.7025 (-0.0240)	-0.4984 (-0.0170)	-0.5521 (-0.0189)	-0.4985 (-0.0170)
	$\psi_{\beta\beta}$	-0.0512 (-0.0267)	-0.0283 (-0.0148)	-0.0376 (-0.0196)	-0.0367 (-0.0192)

Tabela 5.32: Enviesamento bruto (enviesamento relativo) da estimativa média para os parâmetros dos modelos H1 a H5, e $T = 4$

Os valores do EQM, na tabela 5.33, permitem concluir que o parâmetro que apresenta piores valores é $\psi_{\alpha\alpha}$, qualquer que seja o modelo, H1 a H5, e qualquer que seja a dimensão da amostra (50 a 1000 observações). Estes valores são especialmente elevados, quando a amostra é de pequena dimensão, 61.2873, 70.4316 e 149.6839, e o modelo considerado é H3, H4 e H5, respetivamente. Estes três modelos geram dados com um desvio face à normalidade mais acentuado (assimetria e achatamento mais elevados).

O parâmetro $\psi_{\alpha\beta}$ apresenta um valor de EQM bastante superior, 4.4717 e 5.3674, quando a amostra é de dimensão pequena (50 observações) e os dados apresentam assimetria e achatamento moderados e severos, respetivamente, face aos valores obtidos para dados gerados a partir de H1, H2 e H3, respetivamente, 0.0487, 0.3248 e 0.3225. O parâmetro μ_{α} apresenta valores de EQM análogos, quando comparados os valores obtidos para os modelos H3 e H4, e quando comparados os valores obtidos para os modelos H1 e H2. Todavia, o valor de EQM mais elevado verifica-se para o modelo H5 e N=50, 0.5632. Os parâmetros μ_{β} e $\psi_{\beta\beta}$ apresentam valores de EQM semelhantes, quando comparados os valores obtidos para os modelos H1 e H3 e os modelos H2, H4 e H5, qualquer que seja a

dimensão da amostra. Para $\psi_{\beta\beta}$, os resultados obtidos a partir do modelo H5 são exatamente iguais aos obtidos a partir do modelo H4, qualquer que seja a dimensão da amostra.

Modelo	Parâmetro	N=50	N=250	N=500	N = 1 000
H1	$\psi_{\alpha\beta}$	0.0487	0.0091	0.0046	0.0023
	μ_{α}	0.0455	0.0094	0.0043	0.0024
	μ_{β}	0.0114	0.0022	0.0011	0.0006
	$\psi_{\alpha\alpha}$	0.3530	0.0694	0.0348	0.169
	$\psi_{\beta\beta}$	0.0201	0.0038	0.0020	0.0010
H2	$\psi_{\alpha\beta}$	0.3248	0.0622	0.0315	0.0155
	μ_{α}	0.0542	0.0112	0.0053	0.0028
	μ_{β}	0.0456	0.0090	0.0044	0.0023
	$\psi_{\alpha\alpha}$	0.8036	0.1496	0.0783	0.0399
	$\psi_{\beta\beta}$	0.3999	0.0810	0.0426	0.0220
H3	$\psi_{\alpha\beta}$	0.3225	0.0622	0.0304	0.0147
	μ_{α}	0.4945	0.1029	0.0484	0.0260
	μ_{β}	0.0113	0.0022	0.0011	0.0006
	$\psi_{\alpha\alpha}$	61.2873	12.4475	6.2797	3.2637
	$\psi_{\beta\beta}$	0.0233	0.0045	0.0023	0.0011
H4	$\psi_{\alpha\beta}$	4.4717	0.9207	0.4597	0.2448
	μ_{α}	0.4995	0.1042	0.0493	0.0264
	μ_{β}	0.0452	0.0090	0.0044	0.0023
	$\psi_{\alpha\alpha}$	70.4316	14.2374	7.2121	3.7344
	$\psi_{\beta\beta}$	0.4007	0.0811	0.0429	0.0220
H5	$\psi_{\alpha\beta}$	5.3674	1.0884	0.5510	0.2877
	μ_{α}	0.5632	0.1170	0.0565	0.0297
	μ_{β}	0.0452	0.0090	0.0044	0.0023
	$\psi_{\alpha\alpha}$	149.6839	29.8818	15.0955	7.7352
	$\psi_{\beta\beta}$	0.4000	0.0809	0.0429	0.0221

Tabela 5.33: Erro quadrático médio de estimação para os parâmetros dos modelos H1 a H5, e $T = 4$

Os valores de *coverage* apresentados na tabela 5.34 permitem observar que todos os parâmetros apresentam valores de *coverage* baixa, em amostras de dimensão pequena (50 observações) e quando os dados apresentam assimetria e achatamento moderados e severos, variando entre 0.890 e 0.928, no modelo H4 e entre 0.877 e 0.926, no modelo H5, sendo $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ os parâmetros que apresentam piores resultados, no modelo H4, e $\psi_{\alpha\beta}$, μ_{α} , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ no modelo H5. Nos restantes modelos também se verifica que estes 3 parâmetros, $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, apresentam os piores valores de *coverage*, exceto no modelo H3 e para o valor do parâmetro $\psi_{\alpha\beta}$. No modelo H2 e H5 o valor de *coverage* para $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ são inferiores a 0.90 e para o modelo H1 o valor obtido para $\psi_{\alpha\alpha}$ encontra-se no limite do recomendado na literatura, 0.90.

Modelos H1 a H5 com $T = 4$ e estimação ML

Quando se considera o método da máxima verosimilhança na estimação de um modelo com trajetória latente a dados gerados a partir dos modelos H1 a H5, e se comparam os

Modelo	Parâmetro	N=50	N=250	N=500	$N = 1\ 000$
H1	$\psi_{\alpha\beta}$	0.918	0.943	0.943	0.949
	μ_{α}	0.947	0.960	0.960	0.948
	μ_{β}	0.929	0.941	0.948	0.939
	$\psi_{\alpha\alpha}$	0.908	0.939	0.935	0.942
	$\psi_{\beta\beta}$	0.912	0.940	0.936	0.939
H2	$\psi_{\alpha\beta}$	0.910	0.939	0.949	0.957
	μ_{α}	0.952	0.957	0.961	0.948
	μ_{β}	0.922	0.949	0.946	0.941
	$\psi_{\alpha\alpha}$	0.872	0.925	0.935	0.937
	$\psi_{\beta\beta}$	0.849	0.933	0.945	0.938
H3	$\psi_{\alpha\beta}$	0.951	0.951	0.953	0.946
	μ_{α}	0.928	0.944	0.948	0.949
	μ_{β}	0.933	0.941	0.949	0.940
	$\psi_{\alpha\alpha}$	0.918	0.940	0.950	0.949
	$\psi_{\beta\beta}$	0.911	0.944	0.933	0.944
H4	$\psi_{\alpha\beta}$	0.913	0.949	0.955	0.947
	μ_{α}	0.925	0.944	0.949	0.948
	μ_{β}	0.928	0.950	0.946	0.942
	$\psi_{\alpha\alpha}$	0.918	0.941	0.947	0.947
	$\psi_{\beta\beta}$	0.890	0.935	0.944	0.933
H5	$\psi_{\alpha\beta}$	0.916	0.937	0.951	0.944
	μ_{α}	0.912	0.937	0.949	0.954
	μ_{β}	0.926	0.948	0.946	0.944
	$\psi_{\alpha\alpha}$	0.877	0.924	0.944	0.945
	$\psi_{\beta\beta}$	0.897	0.935	0.943	0.932

Tabela 5.34: *Coverage* para os parâmetros dos modelos H1 a H5, e $T = 4$

resultados com os obtidos com um estimador robusto, a diferença de resultados não reside nas estimativas médias dos parâmetros, mas no erro padrão médio e consequentemente nos valores de *coverage* obtidos, sendo os novos valores de *coverage* apresentados na tabela 5.35.

A comparação destes valores obtidos a partir do estimador ML, tabela 5.35, com os obtidos quando considerada a versão robusta do mesmo estimador (MLR), tabela 5.34, permite verificar que para os parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ os valores de *coverage* se encontram abaixo do aceitável, qualquer que seja a dimensão da amostra e para dados cuja assimetria e achatamento são classificados como suave a severo (modelos H2 a H5). Em particular, quando o grau de assimetria e achatamento dos dados são severos, modelo H5, os valores de *coverage* para o parâmetro $\psi_{\alpha\alpha}$ variam entre 0.651 e 0.661. De salientar que, para os dados gerados a partir de H3, os valores de *coverage* para os parâmetros $\psi_{\alpha\beta}$ e $\psi_{\beta\beta}$ são aceitáveis, acima de 0.90, enquanto os valores para o parâmetro $\psi_{\alpha\alpha}$ são baixos, entre 0.777 e 0.798. Relembre-se que no modelo H3, o grau de assimetria e achatamento foi obtido fazendo variar apenas a média de intercepto, tendo sido considerado um valor elevado na classe minoritária face ao assumido na classe majoritária, os restantes parâmetros mantiveram-se iguais em ambas as classes.

Modelo	Parâmetro	N=50	N=250	N=500	$N = 1\ 000$
H1	$\psi_{\alpha\beta}$	0.938	0.946	0.945	0.950
	μ_{α}	0.950	0.960	0.960	0.948
	μ_{β}	0.931	0.941	0.948	0.939
	$\psi_{\alpha\alpha}$	0.916	0.931	0.922	0.923
	$\psi_{\beta\beta}$	0.928	0.944	0.934	0.942
H2	$\psi_{\alpha\beta}$	0.816	0.822	0.844	0.828
	μ_{α}	0.953	0.957	0.961	0.948
	μ_{β}	0.926	0.949	0.946	0.941
	$\psi_{\alpha\alpha}$	0.829	0.835	0.828	0.831
	$\psi_{\beta\beta}$	0.810	0.834	0.801	0.818
H3	$\psi_{\alpha\beta}$	0.943	0.954	0.951	0.944
	μ_{α}	0.929	0.944	0.948	0.949
	μ_{β}	0.933	0.941	0.949	0.940
	$\psi_{\alpha\alpha}$	0.782	0.798	0.777	0.780
	$\psi_{\beta\beta}$	0.930	0.942	0.938	0.946
H4	$\psi_{\alpha\beta}$	0.784	0.803	0.773	0.764
	μ_{α}	0.925	0.944	0.948	0.947
	μ_{β}	0.930	0.950	0.947	0.942
	$\psi_{\alpha\alpha}$	0.768	0.775	0.762	0.747
	$\psi_{\beta\beta}$	0.806	0.842	0.803	0.823
H5	$\psi_{\alpha\beta}$	0.768	0.781	0.757	0.755
	μ_{α}	0.912	0.937	0.949	0.954
	μ_{β}	0.929	0.947	0.946	0.943
	$\psi_{\alpha\alpha}$	0.652	0.661	0.651	0.651
	$\psi_{\beta\beta}$	0.807	0.839	0.803	0.819

Tabela 5.35: *Coverage* para os parâmetros dos modelos H1 a H5, com $T = 4$ e estimador não robusto

Modelos H1 a H5 vs Modelo A1

Quando se comparam os resultados obtidos na estimação do modelo aos dados gerados com distribuição normal a partir do modelo A1, tabela 5.13, com os resultados obtidos na estimação do modelo aos dados gerados a partir de uma mistura de distribuições normais, modelos H1 a H5 (e estimador robusto), tabela 5.32, é possível verificar que o enviesamento relativo na estimativa média dos parâmetros $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ aumenta, em amostras de dimensão elevada (1000 observações). Para amostras de pequena dimensão, $N=50$, verifica-se que o enviesamento relativo na estimativa média de $\psi_{\alpha\alpha}$ aumenta qualquer que seja o modelo que gera os dados, enquanto que para o parâmetro $\psi_{\beta\beta}$ aumenta quando se consideram os dados gerados a partir dos modelos H1 e H3, e diminui se os dados são gerados a partir dos modelos H2, H4 e H5.

A comparação para os parâmetros μ_α , μ_β e $\psi_{\alpha\beta}$ faz-se recorrendo ao valor de enviesamento bruto. Assim, verifica-se que sendo a amostra grande ($N=1000$) o enviesamento da estimativa média, face ao valor populacional, dos parâmetros μ_α , μ_β e $\psi_{\alpha\beta}$ aumenta, quando os dados são provenientes de uma mistura de distribuições de normais, qualquer que seja o grau de assimetria e achatamento nos dados, exceto $\psi_{\alpha\beta}$ em dados com assimetria e achatamento fracos, modelo H1. Se a amostra é de pequena dimensão ($N=50$) observa-se que o enviesamento na estimativa média dos parâmetros $\psi_{\alpha\beta}$ e μ_α aumenta e para μ_β diminui, nos modelos que geram dados não normais, modelo H1 a H5.

O EQM de estimação aumenta quando os dados não apresentam distribuição normal, sendo de salientar que para μ_β e $\psi_{\beta\beta}$ os valores apresentados na tabela 5.14, para dados normais são análogos aos obtidos para dados não normais, gerados a partir dos modelos H1 e H3, tabela 5.33.

De salientar ainda que o EQM de estimação para o parâmetro $\psi_{\alpha\alpha}$, com dados não normais, com assimetria e achatamento moderados, modelo H4, aumenta 355 vezes mais face a dados com distribuição normal em amostras de pequena dimensão (50 observação), o valor passou de 0.1981 para 70.4316, enquanto que para amostras de dimensão elevada (1000 observações), aumenta mais de 377 vezes o valor obtido para dados normais, passa de 0.0099 para 3.7344. Quando os dados provêm de uma mistura de distribuições normais com assimetria e achatamento severos, modelo H5, aumenta 755 vezes mais, passando de 0.1981 para 149.6839, se $N=50$, e de 0.0099 para 7.7352 se $N=1000$.

Para o parâmetro $\psi_{\alpha\beta}$ o aumento no EQM é quase 200 vezes, quando se comparam os resultados para dados não normais obtidos do modelo H4, com dados normais, em amostras pequenas (50 observações), sendo o aumento de 140 vezes mais se os dados têm distribuição normal com assimetria e achatamento severos, modelo H5.

Os valores de *coverage*, em especial para os parâmetros $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ tomam sempre valores inferiores, quando os dados são não normais face a dados normais, tabelas 5.34 e 5.15, respetivamente. Para dados normais os valores obtidos são 0.926 e 0.929, para $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$, respetivamente, enquanto que para dados não normais, $\psi_{\alpha\alpha}$ varia entre 0.872

e 0.918, e $\psi_{\beta\beta}$ entre 0.890 e 0.912, nos diferentes modelos H1 a H5.

Modelos H1 a H5 com $T = 3$ vs $T = 4$

Com o objetivo de avaliar o efeito da diminuição do número de momentos temporais do modelo na qualidade de estimação de dados obtidos como uma mistura de distribuições normais, com diferentes graus de assimetria e achatamento, são comparados os resultados da tabela G.1, em anexo, valores obtidos na estimação de um modelo com 3 momentos temporais, com os resultados da tabela 5.32, valores obtidos quando o modelo considerado apresenta 4 momentos temporais.

A comparação destas tabelas permite observar que o enviesamento relativo na estimativa média do parâmetro $\psi_{\alpha\alpha}$ aumenta, enquanto que para o parâmetro μ_{α} diminui, quando se considera um modelo com 3 momentos temporais em amostras de pequena dimensão (50 observações) e qualquer que seja o grau de não normalidade existente nos dados. Todavia, se a amostra é de dimensão elevada ($N=1\ 000$), o enviesamento relativo na estimativa média de $\psi_{\alpha\alpha}$ aumenta, se os dados são gerados a partir de H3 e H4, e diminui caso contrário. Para o parâmetro μ_{α} o enviesamento apresenta o mesmo comportamento que em amostras pequenas. De salientar que, o enviesamento na estimativa média dos parâmetros $\psi_{\alpha\beta}$ e μ_{β} apresenta valores superiores a 0.05 (valor considerado como o limite do negligenciável na literatura), este último apenas para amostras com dimensão de 250 ou mais observações, quando os dados são gerados a partir de H1 ou H3.

Por outro lado, os valores do EQM (tabela G.2 em anexo) aumentam, quando o número de momentos temporais do modelo é menor, sendo os parâmetros mais afetados $\psi_{\alpha\beta}$ e $\psi_{\beta\beta}$, quando se consideram os dados gerados a partir do modelo H1. O EQM para estes parâmetros aumenta quase 4 vezes e mais de 6 vezes, face ao modelo com 4 momentos temporais, respetivamente. Por exemplo, $\psi_{\beta\beta}$ aumenta de 0.0201 para 0.1327, com a diminuição do número de momentos temporais, tendo a amostra dimensão 50.

Quando se consideram os dados gerados a partir do modelo H2, o aumento no EQM é maior nos parâmetros $\psi_{\alpha\beta}$ e $\psi_{\beta\beta}$, sendo este aumento de quase 2 vezes o valor obtido para o modelo com 4 momentos temporais, qualquer que seja a dimensão da amostra. Por exemplo, para $\psi_{\beta\beta}$ o valor aumenta de 0.3999 para 0.6918, quando se compara o modelo com 3 momentos temporais face ao mesmo modelo com 4 momentos temporais, para amostras de dimensão 50.

Quando se consideram os resultados da estimação de um modelo com trajetória latente a dados gerados a partir do modelo H3, com 3 momentos temporais em relação ao mesmo modelo com 4 momentos temporais, verifica-se que os parâmetros mais afetados são $\psi_{\alpha\beta}$ e $\psi_{\beta\beta}$, sendo o aumento para $\psi_{\alpha\beta}$ de 5 vezes mais, enquanto que para $\psi_{\beta\beta}$ este aumento é de 44 vezes mais, qualquer que seja a dimensão da amostra. O valor de $\psi_{\beta\beta}$ para o modelo com 3 momentos temporais e $N=50$, é 1.0556, enquanto que para o modelo com 4 momentos temporais é 0.0233. Para amostras de dimensão elevada, $N=1\ 000$, o EQM de

estimação do parâmetro $\psi_{\beta\beta}$ é de 0.0484 no modelo com $T=3$, sendo de 0.0011 no modelo com $T=4$.

Quando se consideram os dados gerados a partir de H4, o parâmetro que apresenta um maior aumento no EQM de estimação é $\psi_{\beta\beta}$, de quase 5 vezes mais, se $N=50$, e de quase 4 vezes mais, se $N=1\ 000$.

Quando se consideram os dados gerados a partir do modelo H5, o parâmetro que apresenta um maior aumento no EQM de estimação é $\psi_{\beta\beta}$, aumentando 5 vezes mais, se $N=50$, e aumentando 4 vezes mais, se $N=1\ 000$, se o modelo considerado tem apenas 3 momentos temporais.

Os valores de *coverage* para os dados gerados a partir dos modelos H1 a H5 com 3 momentos temporais são apresentados na tabela G.3, em anexo. Para os dados obtidos a partir do modelo H1 o valor de *coverage* para o parâmetro $\psi_{\alpha\alpha}$ é 0.914, valor próximo do limite do aceitável 0.90, sendo a dimensão da amostra de 50 observações.

Para os dados gerados a partir do modelo H2, os parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ apresentam valores baixos de *coverage* de 0.902, 0.889 e 0.902, respectivamente, quando a dimensão da amostra é 50 observações. Todavia, quando a dimensão da amostra é 250 observações, o valor de *coverage* para $\psi_{\alpha\alpha}$ encontra-se no limite do aceitável, 0.919.

Para os dados obtidos gerados a partir do modelo H3, os parâmetros que apresentam valores mais baixos de *coverage* são μ_{α} e $\psi_{\alpha\alpha}$, 0.914 e 0.919, respectivamente, quando a dimensão da amostra é pequena (50 observações).

Para os dados obtidos a partir dos modelos H4 e H5, todos os parâmetros do modelo apresentam valores baixos de *coverage*, quando a dimensão da amostra é pequena. Para os dados provenientes do modelo H4, estes valores são 0.902, 0.911, 0.928, 0.918 e 0.894, respectivamente, estão no limite do aceitável, sendo que para $\psi_{\beta\beta}$ se encontra abaixo do recomendado como aceitável nas literatura da área. Para os dados provenientes do modelo H5, os valores são 0.885, 0.906, 0.927, 0.874 e 0.900, respectivamente, sendo que para $\psi_{\alpha\beta}$ e $\psi_{\alpha\alpha}$ os valores de *coverage* estão abaixo do aceitável e para $\psi_{\beta\beta}$ e μ_{α} estão no limite.

Modelos H1 e H5 com $T = 5$ vs $T = 4$

Com o objetivo de avaliar o efeito de aumentar o número de momentos temporais na geração de dados a partir de uma mistura de distribuições normais e posterior estimação de um modelo com trajetória latente, foram considerados os modelos H1 a H4 com 5 momentos temporais e os resultados obtidos da estimação destes dados apresentados nas tabelas I.1 a I.3, em anexo.

Os valores da tabela I.1, em anexo, quando comparados com os valores da tabela 5.32, que apresenta os valores de enviesamento na estimativa média dos parâmetros do

modelo com 4 momentos temporais, permitem verificar que qualquer que seja o grau de não normalidade dos dados, se a amostra é de dimensão elevada (1000 observações) o enviesamento relativo na estimativa média de todos os parâmetros do modelo diminui, exceto para $\psi_{\alpha\alpha}$, quando os dados apresentam assimetria e achatamento fracos (modelo H1). Para amostras de pequena dimensão (50 observações), o enviesamento na estimativa média do parâmetro $\psi_{\alpha\alpha}$ aumenta se os dados apresentam assimetria, fraca ou fraca a suave e fraco a moderado, e achatamento fraco ou moderado, respetivamente.

Para o parâmetro μ_α o enviesamento diminui se os dados são gerados a partir dos modelos H1, H3, H4 e H5. Para os parâmetros μ_β e $\psi_{\beta\beta}$ diminui se os dados são obtidos a partir de H1 e de H3, aumentando para dados com assimetria e achatamento, moderado e severo (modelo H4 e H5). De referir ainda, que os parâmetros $\psi_{\alpha\beta}$ e μ_β apresentam valores de enviesamento relativo superiores ao considerado negligenciável, quando os dados são gerados a partir dos modelos H1 e H3. Apenas quando a dimensão da amostra é pequena ($N=50$) e os dados apresentam um grau de não normalidade fraca (modelo H1), o enviesamento na estimativa média de μ_β é 0.

A análise das tabelas I.2 e 5.33, em anexo, permite avaliar o efeito do aumento do número de momentos temporais no modelo, no EQM de estimação dos parâmetros do modelo. Para os dados gerados a partir dos modelos H1 e H2 verifica-se que existe uma diminuição do EQM de estimação para todos os parâmetros do modelo, exceto para μ_α , qualquer que seja a dimensão da amostra. O parâmetro μ_α apresenta valores de erro quadrático médio semelhantes, não parecendo ser afetado pelo aumento do número de momentos temporais, de 4 para 5.

Para os dados gerados a partir dos modelos H3 a H4, dados com níveis de assimetria e achatamento mais acentuados, verifica-se que existe uma diminuição no EQM de estimação dos parâmetros $\psi_{\alpha\beta}$, μ_β e $\psi_{\beta\beta}$. Os parâmetros μ_α e $\psi_{\alpha\alpha}$ apresentam valores análogos, quando se comparam os resultados dos modelos com 4 e 5 momentos temporais. Para os dados gerados a partir do modelo H5, dados que apresentam valores de assimetria e achatamento severos, o EQM de estimação dos parâmetros $\psi_{\alpha\beta}$ e $\psi_{\beta\beta}$ diminui, para $\psi_{\alpha\alpha}$ aumenta, enquanto que para μ_α e μ_β , se a amostra é grande (1000 observações) os valores mantêm-se constantes.

A comparação dos valores de *coverage* obtidos para o mesmo modelo com diferente número de momentos temporais, permite verificar que se as amostras são pequenas (50 observações) estes valores são próximos de 0.90, sendo que alguns deles se encontram abaixo. Por exemplo, para os dados obtidos a partir do modelo H1, os valores de *coverage* para os parâmetros $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ são 0.889 e 0.909, respetivamente. Para os dados gerados a partir do modelo H2, os parâmetros $\psi_{\alpha\beta}$, $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ apresentam valores de *coverage* de 0.883, 0.844 e 0.914, respetivamente. Enquanto que, se os dados são gerados a partir do modelo H4, para estes parâmetros foram obtidos os seguintes valores, 0.911, 0.910 e 0.917, respetivamente. Para os dados obtidos a partir do modelo H3 são os parâmetros $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ que apresentam valores de *coverage* mais baixos, 0.916 e 0.907, respetivamente.

Para os dados gerados a partir do modelo H5, os parâmetros que apresentam valores de *coverage* mais baixos são $\psi_{\alpha\beta}$, μ_α e $\psi_{\alpha\alpha}$, 0.893, 0.908 e 0.865, respectivamente.

5.4.2 Dados gerados a partir de uma mistura de distribuições normais - com omissões

Com o objetivo de avaliar o efeito da existência de omissões resultantes de um *planned missing design* em dados gerados a partir de uma mistura de distribuições normais, com diferentes graus de assimetria e achatamento, foram considerados na geração dos dados os modelos H1 e H5 com 3, 4 e 5 momentos temporais e com omissões que representam os padrões de omissão presentes no ICOR (tabelas 5.1, 5.2 e 5.3). Assim, para o modelo com 4 momentos temporais foi considerada a existência de 50% de omissões no 1º momento temporal e 25% no 2º e 4º momentos temporais. Para o modelo com 3 momentos temporais foi considerada a existência de 25% de omissões no 1º e 3º momento temporal, enquanto que para o modelo com 5 momentos temporais foram consideradas 75% de omissões no 1º momento temporal, 50% no 2º e 25% no 3º e 5º. As dimensão de amostra consideradas foram $N=250$ e $N=500$.

Modelos H1 a H5 com $T = 4$ e omissões

Na tabela 5.36 são apresentados os resultados obtidos na estimação de um modelo com trajetória latente com 4 momentos temporais a dados gerados como uma mistura de distribuições normais com diferentes níveis de assimetria e achatamento. Os dados considerados foram gerados a partir da estrutura dos modelos H1 a H5 com 4 momentos temporais. A existência de omissões com um padrão igual ao apresentado na tabela 5.2, foi considerada nos dados gerados. As amostras geradas foram de dimensão 250 e 500 observações.

A análise da tabela 5.36, permite verificar que o enviesamento na estimativa média dos parâmetros $\psi_{\alpha\beta}$ e μ_β é superior a 0.05, o limite do negligenciável, quando os dados são gerados a partir dos modelos H1 e H3. Para os dados gerados a partir de H1 os valores para $\psi_{\alpha\beta}$ são bastante elevados, -6.6 e -17.2, sendo $N=250$ e $N=500$, respectivamente. Para μ_β os valores de enviesamento obtidos são -0.7895 e -0.8421, com $N=250$ e $N=500$. Para os dados gerados a partir do modelo H3, o enviesamento observado para $\psi_{\alpha\beta}$ é de 2.2917 e 3.1042, se $N=250$ e $N=500$. Para o parâmetro μ_β , os valores obtidos são -0.5263 e -1.0526, para $N=250$ e $N=500$. De salientar que, estes valores de enviesamento são superiores em amostras de dimensão mais elevada ($N=500$).

Os valores do EQM de estimação apresentados na tabela 5.36 mostram que para os parâmetros $\psi_{\alpha\beta}$, μ_α e $\psi_{\alpha\alpha}$ estes valores aumentam quando se comparam os valores obtidos na estimação de dados gerados a partir dos modelos H1, H2, H3, H4 e H5, sendo bastante mais elevados se os dados são gerados a partir dos modelos H3, H4 e H5, dados que apre-

sentam maior assimetria e achatamento. O EQM para os parâmetros μ_β e $\psi_{\beta\beta}$ é menor se os dados são gerados a partir dos modelos H1 e H3, apresentando valores semelhantes para dados gerados a partir dos modelos H2, H4 e H5. Para todos os parâmetros do modelo o EQM de estimação é maior se os dados são gerados a partir do modelo H5, dados com assimetria e achatamentos severos.

Os valores de *coverage* apresentados na tabela 5.36, mostram que para todos os modelos o parâmetro que apresenta menor valor de *coverage* é $\psi_{\alpha\alpha}$, sendo estes valores menores se os dados são gerados a partir dos modelos H2 e H5, em que se obtém 0.902 e 0.901 para amostras de dimensão 250, e 0.918 e 0.910 para amostras de dimensão 500. De salientar que, estes valores se encontram no limite do considerado aceitável na literatura: 0.90.

Modelo		Enviesamento		EQM		Coverage	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
H1	$\psi_{\alpha\beta}$	0.0033 (- 6.6)	0.0086 (- 17.2)	0.0211	0.0105	0.950	0.939
	μ_α	-0.0051 (-0.0168)	-0.0073 (-0.0240)	0.0140	0.0069	0.951	0.954
	μ_β	-0.0015 (- 0.7895)	-0.0016 (- 0.8421)	0.0035	0.0018	0.941	0.948
	$\psi_{\alpha\alpha}$	-0.0384 (-0.0230)	-0.0395 (-0.0236)	0.1260	0.0624	0.940	0.934
	$\psi_{\beta\beta}$	-0.0043 (-0.0213)	-0.0061 (-0.0302)	0.0086	0.0046	0.944	0.944
H2	$\psi_{\alpha\beta}$	-0.0361 (-0.0338)	-0.0255 (-0.0239)	0.0908	0.0459	0.936	0.935
	μ_α	-0.0065 (-0.0213)	-0.0080 (-0.0262)	0.0161	0.0080	0.948	0.958
	μ_β	-0.0136 (-0.0279)	-0.0140 (-0.0288)	0.0106	0.0056	0.936	0.935
	$\psi_{\alpha\alpha}$	-0.0719 (-0.0331)	-0.0671 (-0.0308)	0.2431	0.1247	0.902	0.918
	$\psi_{\beta\beta}$	-0.0494 (-0.0258)	-0.0502 (-0.0262)	0.1007	0.0548	0.931	0.928
H3	$\psi_{\alpha\beta}$	0.0110 (2.2917)	0.0149 (3.1042)	0.1159	0.0543	0.941	0.947
	μ_α	-0.0432 (-0.0238)	-0.0436 (-0.0240)	0.1147	0.0573	0.933	0.944
	μ_β	-0.0010 (- 0.5263)	-0.0020 (- 1.0526)	0.0039	0.0020	0.939	0.957
	$\psi_{\alpha\alpha}$	-0.6959 (-0.0279)	-0.6375 (-0.0255)	15.1062	7.6018	0.926	0.932
	$\psi_{\beta\beta}$	-0.0058 (-0.0284)	-0.0084 (-0.0411)	0.0099	0.0053	0.946	0.946
H4	$\psi_{\alpha\beta}$	-0.1683 (-0.0263)	-0.1506 (-0.0235)	1.0335	0.5407	0.931	0.928
	μ_α	-0.0442 (-0.0243)	-0.0450 (-0.0247)	0.1147	0.0577	0.932	0.944
	μ_β	-0.0132 (-0.0271)	-0.0141 (-0.0290)	0.0106	0.0056	0.939	0.936
	$\psi_{\alpha\alpha}$	-0.7587 (-0.0298)	-0.6860 (-0.0269)	16.9398	8.6007	0.921	0.925
	$\psi_{\beta\beta}$	-0.0499 (-0.0261)	-0.0504 (-0.0263)	0.0964	0.0528	0.940	0.928
H5	$\psi_{\alpha\beta}$	-0.1937 (-0.0302)	-0.1707 (-0.0266)	1.2489	0.6519	0.927	0.925
	μ_α	-0.0510 (-0.0280)	-0.0504 (-0.0277)	0.1318	0.0659	0.931	0.927
	μ_β	-0.0134 (-0.0275)	-0.0143 (-0.0294)	0.0107	0.0057	0.934	0.934
	$\psi_{\alpha\alpha}$	-1.1545 (-0.0394)	-1.0247 (-0.0350)	34.7984	17.7495	0.901	0.910
	$\psi_{\beta\beta}$	-0.0480 (-0.0251)	-0.0505 (-0.0264)	0.1038	0.0570	0.935	0.936

Tabela 5.36: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos H1 e H5, com $T = 4$ e omissões

Modelos H1 a H5 com $T = 4$ e com omissões *vs* sem omissões

Os valores apresentados nas tabelas 5.36 e 5.32 permitem verificar que o enviesamento na estimativa média de todos os parâmetros do modelo aumenta, para amostras de dimensão 500, quando os dados apresentam omissões, qualquer que sejam os graus de assimetria

e achatamento. A exceção é o parâmetro μ_β , se os dados são obtidos a partir do modelo H1. Se a dimensão da amostra é 250 observações, todos os parâmetros do modelo apresentam um maior enviesamento relativo na sua estimativa média, sendo os dados obtidos a partir de H2, H4 e H5. Para os dados gerados a partir de H1 e H3, os parâmetros μ_α e $\psi_{\alpha\alpha}$ apresentam um enviesamento maior na presença de omissões, enquanto que μ_β e $\psi_{\beta\beta}$ apresentam um enviesamento menor. O parâmetro $\psi_{\alpha\beta}$ apresenta um enviesamento maior apenas se os dados são obtidos a partir de H3. A comparação dos valores apresentados nas tabelas 5.36 e 5.33 permitem verificar que os valores de EQM de estimação aumentam quando os dados apresentam omissões, independentemente do desvio face à normalidade dos dados.

Os valores de *coverage* apresentados nas tabelas 5.36 e 5.34 permitem concluir que para o parâmetro $\psi_{\alpha\alpha}$ os valores de *coverage* são menores, em particular quando se consideram os dados obtidos a partir dos modelos H2 e H5 e uma amostra de dimensão mais pequena (250 observações), os valores obtidos 0.902 e 0.901 encontram-se no limite do aceitável, 0.90.

Modelos H1 a H5 com $T = 4$ e com omissões vs Modelo A1 com omissões

A análise das tabelas 5.36 e 5.25 a 5.27 permite comparar os resultados obtidos na estimação do modelo com trajetória latente, quando se consideram dados normais com omissões e dados não normais (com diferentes graus de não normalidade) com omissões.

Quando se consideram os resultados obtidos na estimação do modelo aos dados gerados a partir dos modelos H1, H2, H3 e H4, modelos que geram dados não normais (com diferentes níveis de assimetria e achatamento) verifica-se que, face aos resultados obtidos na estimação do modelo aos dados gerados a partir do modelo A1, modelo que gera dados com distribuição normal, ocorreu um aumento no enviesamento relativo da estimativa média para os parâmetros $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ e qualquer que seja a dimensão da amostra. Para os parâmetros $\psi_{\alpha\beta}$, μ_α e μ_β são comparados os valores do enviesamento bruto, uma vez que para os dados com distribuição normal não foi possível calcular o enviesamento relativo, para estes parâmetros do modelo. De relembrar que, o enviesamento bruto apresenta o desvio da estimativa média obtida das 1 000 réplicas que foram geradas face ao valor populacional do parâmetro. Assim, verifica-se que para os parâmetros $\psi_{\alpha\beta}$, μ_α e μ_β , os valores das suas estimativas médias apresentavam-se mais distantes dos seus valores populacionais, quando os dados provêm de uma mistura de distribuições normais.

Os valores do EQM também verificam, de um modo geral, um aumento, se os dados subjacentes à estimação provêm de uma mistura de distribuições normais. Para os dados obtidos a partir do modelo H1, existe um aumento para todos os parâmetros do modelo, exceto para μ_β e $\psi_{\beta\beta}$ que apresentam valores análogos. O parâmetro que verifica maior aumento é $\psi_{\alpha\alpha}$, passando quase para o dobro, isto é, passa de 0.0739 para 0.1260 se $N = 250$ e de 0.037 para 0.0624 se $N = 500$.

Para os dados gerados a partir do modelo H2 todos os parâmetros verificam um aumento no valor do EQM de estimação, sendo este aumento de 12 vezes mais para o parâmetro $\psi_{\beta\beta}$, 3 vezes mais para os parâmetros μ_β e $\psi_{\alpha\alpha}$ e 5 vezes mais para $\psi_{\alpha\beta}$.

Para os dados gerados a partir do modelo H3, apenas para o parâmetro μ_β os valores do EQM são semelhantes. Para o parâmetro $\psi_{\alpha\alpha}$ o aumento é de 205 vezes mais, enquanto que para μ_α é de quase 11 vezes mais e para $\psi_{\alpha\beta}$ quase sete vezes mais.

Para os dados obtidos a partir do modelo H4, todos os parâmetros apresentam aumento no EQM de estimação, sendo este aumento de 230 vezes mais para $\psi_{\alpha\alpha}$, 57 vezes mais para $\psi_{\alpha\beta}$, 12 vezes mais para $\psi_{\beta\beta}$, 11 vezes mais para μ_α e 3 vezes mais para μ_β .

Para os dados obtidos a partir de H5, os parâmetros que apresentam um maior aumento no EQM é $\psi_{\alpha\alpha}$, cujo valor passa de 0.0739 e 0.0377 para 34.7984 e 17.7495, se $N=250$ e $N=500$, respectivamente. O aumento do valor de EQM para este parâmetro é de 470 vezes mais, sendo de 70 vezes mais o aumento para $\psi_{\alpha\beta}$, de cerca de 12 vezes mais para μ_α e de 3 vezes mais para μ_β . Para todos os parâmetros do modelo é possível concluir que o EQM se acentua com o aumento do desvio face à normalidade dos dados.

A comparação dos valores de *coverage* obtidos para dados não normais com omissões, tabela 5.36, e para os dados normais com omissões, tabela 5.27, permitem concluir que não existem grandes diferenças entre resultados apresentados para os modelos H1 e A1. Para os modelos H2, H3, H4 e H5 face ao modelo A1 verifica-se que os valores de *coverage* obtidos diminuem nos modelos com dados obtidos a partir de uma mistura de distribuições normais.

Modelos H1 a H5 com $T = 3$ e omissões

Na tabela G.4, em anexo, são apresentados os valores obtidos na estimação de um modelo com trajetória latente com 3 momentos temporais, aos dados gerados com omissões a partir dos modelos H1 a H5, considerando apenas 3 momentos temporais.

A análise desta tabela permite verificar que para amostras de dimensão $N=250$, os valores de enviesamento na estimativa média dos parâmetros $\psi_{\alpha\beta}$, μ_β e $\psi_{\beta\beta}$ obtidos na estimação do modelo aos dados gerados a partir dos modelos H1 e H3, são superiores a 0.05, limite do negligenciável. Para amostras de dimensão $N=500$, os valores de enviesamento na estimativa média dos parâmetros $\psi_{\alpha\beta}$ e μ_β obtidos na estimação do modelo aos dados gerados a partir do modelo H1, são superiores a 0.05, enquanto que se os dados são obtidos a partir do modelo H3, apenas μ_β apresenta um enviesamento superior a 0.05.

Quando se comparam estes resultados com os obtidos para dados sem omissões, tabela G.1 verifica-se que a presença de omissões nos dados obtidos como uma mistura de distribuições normais gerados a partir dos modelos H1 e H3, provocou um enviesamento nas estimativas médias dos parâmetro μ_β e $\psi_{\beta\beta}$ que não pode ser considerado negligenciável,

quando $N=250$. Os valores do EQM de estimação apresentados na tabela G.2, mostram que o parâmetro que apresenta valores mais elevados é $\psi_{\alpha\alpha}$, seguido de $\psi_{\alpha\beta}$, qualquer que seja o modelo H1 a H5 considerado na geração dos dados, e que aumenta com o aumento do grau de não normalidade que os dados apresentam.

Comparando os valores de EQM de obtidos na estimação de dados com omissões face aos mesmos dados sem omissões, tabelas G.4 e G.1, verifica-se que existe um aumento nestes valores para todos os parâmetros do modelo e qualquer que seja o grau de assimetria achatamento dos dados (modelos H1 a H5). Todavia, o maior aumento do EQM verifica-se na estimação do modelo aos dados obtidos a partir do modelo H1.

Modelos H1 a H5 com $T = 5$ e omissões

Na tabela I.4, em anexo, são apresentados os resultados da estimação de um modelo com trajetória latente aos dados gerados a partir dos modelos H1 a H4 com 5 momentos temporais e omissões que configuram um *planned missing design*. A análise desta tabela permite verificar que para os parâmetros $\psi_{\alpha\beta}$ e μ_{β} o enviesamento nas suas estimativas médias é superior a 0.05, valor considerado o limite do negligenciável, quando se consideram os dados gerados a partir dos modelos H1 e H3, e sendo a dimensão da amostra $N=250$. Para amostras de dimensão $N=500$, verifica-se que para dados gerados a partir de H3, o enviesamento nas estimativas médias dos parâmetros $\psi_{\alpha\beta}$ e μ_{β} apresenta valores superiores a 0.05, enquanto que para os dados gerados a partir de H1, apenas o parâmetro μ_{β} apresenta um enviesamento superior a 0.05.

A comparação dos resultados apresentados nas tabelas I.4 e I.3 permite verificar que a existência de omissões nos dados fez com que o enviesamento na estimativa média de μ_{β} deixasse de ser negligenciável, e que na estimativa média de $\psi_{\alpha\beta}$, passasse a ser, quando são considerados os dados obtidos a partir do modelo H1 e para amostras de dimensão $N=250$ e $N=500$, respetivamente.

Para o EQM de estimação verifica-se que para o parâmetro $\psi_{\alpha\beta}$, μ_{α} , $\psi_{\alpha\alpha}$ e $\psi_{\beta\beta}$ esta medida é maior, quando se consideram os dados obtidos a partir do modelo H5, para amostras de dimensão 500. Para amostras de dimensão $N=250$, apenas os parâmetros $\psi_{\alpha\beta}$, μ_{α} e $\psi_{\alpha\alpha}$ apresentam valores superiores de EQM. Se os dados são obtidos do modelo H5, o EQM de estimação de μ_{β} é semelhante, quando se consideram os dados gerados a partir dos modelo H2, H4 e H5.

Os valores de *coverage* para o conjunto dos parâmetros do modelo é bom, exceto para $\psi_{\alpha\alpha}$ quando considerados os dados gerados a partir do modelo H2.

5.5 Principais conclusões do estudo de simulação

5.5.1 Conclusões do estudo para dados gerados com distribuição normal

Neste estudo de simulação começou-se por gerar dados com distribuição normal e sem omissões, a partir de um modelo com trajetória latente, com quatro momentos temporais, considerando diferentes valores para a média e variância do declive. A covariância entre o intercepto e o declive assumiu valores de modo a assegurar diferentes graus de correlação entre os fatores latentes do modelo, sendo que as variâncias dos termos residuais tomaram valores por forma a verificar uma fiabilidade dos indicadores de 0.5, em todos os momentos temporais, valor usualmente encontrado na prática. Os valores populacionais da média e da variância do intercepto foram fixados a zero e um, respetivamente.

Variação da média do declive

Quando se faz variar o parâmetro populacional média do declive, mantendo constantes os restantes parâmetros populacionais no modelo a partir do qual os dados são gerados, é possível verificar que esta variação não influencia as estimativas médias obtidas para cada um dos parâmetros do modelo em consideração, qualquer que seja a dimensão da amostra, a variância do declive e o grau de correlação entre intercepto e declive, exceto para o próprio parâmetro. Note-se que o enviesamento bruto se mantém, enquanto que o enviesamento relativo diminui à medida que o valor populacional do próprio parâmetro aumenta (em valor absoluto).

Variação da variância do declive

Quando se aumenta a variância populacional do declive, para amostras grandes (1 000 observações) aumenta o enviesamento relativo na estimativa média da variância do intercepto, diminuindo para a variância do declive. O enviesamento bruto da média do intercepto mantém-se (tomando valores próximos de zero) e aumenta para a média do declive, qualquer que seja a correlação considerada entre intercepto e declive. Para a covariância entre intercepto e declive o enviesamento relativo diminui quando aumenta o valor populacional da variância do declive, mas apenas quando existe correlação entre o intercepto e o declive. Considerando a inexistência de correlação entre o intercepto e o declive só é possível o cálculo do enviesamento absoluto, sendo que este aumenta com o aumento da variância populacional do declive.

Se a dimensão da amostra é pequena (50 observações) verifica-se que o aumento da variância populacional do declive provoca um aumento no enviesamento relativo da covariância entre intercepto e declive, uma diminuição no enviesamento relativo da variância do declive, bem como uma diminuição, até um determinado valor $\psi_{\beta\beta}=2$, do enviesamento relativo da variância do intercepto, sendo a correlação considerada no modelo que gera os

dados positiva.

Os valores do erro quadrático médio de estimação para o conjunto de todos os parâmetros do modelo aumentam com o aumento do valor populacional da variância do declive. Todavia, o parâmetro para o qual esse aumento é menos expressivo é a média do intercepto, sendo os valores semelhantes em amostras de dimensão elevada (1 000 observações), qualquer que seja o grau de correlação considerado no modelo.

Variação da correlação entre os fatores latentes

Se a correlação é negativa e elevada, o enviesamento relativo obtido na estimativa média da covariância e das variâncias do intercepto e declive aumenta, se $\psi_{\beta\beta}=0.2$ ou 1, e diminui para valores mais elevados de $\psi_{\beta\beta}$. Se a correlação é negativa mas baixa o enviesamento relativo obtido na estimativa média da covariância e das variâncias do intercepto e do declive aumenta, se $\psi_{\beta\beta}=0.2$ ou 1, e diminui para valores mais elevados deste parâmetro. O enviesamento relativo no valor estimado da variância do intercepto diminui, enquanto que para a estimativa da variância do intercepto aumenta, com o aumento da variância populacional do declive. Se não é considerada correlação entre intercepto e declive, verifica-se que o enviesamento relativo na estimativa média da variância do intercepto diminui se $\psi_{\beta\beta}=0.2, 1$ ou 2, e aumenta caso contrário; para a variância do declive aumenta para valores populacionais de $\psi_{\beta\beta}=0.2$ ou 1 e aumenta caso contrário. O enviesamento bruto do valor estimado da covariância entre intercepto e declive aumenta com o aumento de $\psi_{\beta\beta}$.

O enviesamento absoluto na estimativa média do parâmetro média do intercepto mantém-se constante, aumentando para a média do declive, com o aumento do valor populacional da variância do declive, qualquer que seja o grau de correlação considerado.

O enviesamento relativo obtido na estimação dos diferentes parâmetros do modelo, qualquer que seja o modelo considerado, a dimensão da amostra e o grau de correlação entre o intercepto e o declive, de acordo com a literatura da área pode ser considerado negligenciável, isto é, os valores estimados podem ser considerados não enviesados. Por outro lado, para os parâmetros para os quais só é possível o cálculo do enviesamento bruto, os valores obtidos também são inferiores a 0.05 (limite acima do qual se considera existir enviesamento).

Quando se aumenta a variância populacional do declive, independentemente da correlação considerada entre intercepto e declive e a dimensão da amostra, o valor do erro quadrático médio de estimação para todos os parâmetros do modelo aumenta.

A comparação do enviesamento obtido na estimativa média dos parâmetros do modelo em amostras de dimensão elevada (1 000 observações), quando se considera o mesmo grau de correlação positivo e negativo, permite dizer que para as variâncias do intercepto e do declive, e covariância entre eles, o enviesamento relativo é menor, se a correlação é

elevada e negativa (-0.9). Neste caso, o enviesamento absoluto das médias do intercepto e do declive não é influenciado pelo valor de correlação entre o intercepto e o declive considerado no modelo que gera os dados.

Em amostras de pequena dimensão (50 observações), a média do declive, a variância do declive e a covariância entre intercepto e declive (se a variância populacional do declive é elevada) são os parâmetros cujas estimativas médias apresentam maior enviesamento, sendo o grau de correlação positivo e elevado, enquanto que a média do intercepto, a variância do intercepto e a covariância entre intercepto e declive (se a variância populacional do declive é baixo) são os parâmetros cujas estimativas médias apresentam menor enviesamento.

Para amostras de grande dimensão (1 000 observações), quando se comparam os enviesamentos obtidos em modelos com correlação baixa (0.5) face ao mesmo grau de correlação negativo, verifica-se que estes são semelhantes para as médias do intercepto e do declive e para a variância do declive, aumentando para a variância do intercepto, bem como para a covariância entre intercepto e declive, sendo a variância populacional do declive baixa. Para amostras de pequena dimensão (50 observações), o enviesamento nas estimativas médias da covariância entre intercepto e declive (se o valor populacional de $\psi_{\beta\beta}$ é elevado), da média do declive, da variância do intercepto (se o valor populacional de $\psi_{\beta\beta}$ é elevado) e da variância do declive, é maior quando a correlação é positiva. Para a média do intercepto o enviesamento é inferior, quando a correlação é positiva.

A comparação entre os resultados obtidos para modelos com grau de correlação baixo, positivo ou negativo, ou zero, em amostras de dimensão elevada (1 000 observações), permite verificar que o enviesamento nas estimativas médias dos parâmetros é semelhante, exceto para a variância do intercepto, o qual apresenta valores de enviesamento menores se a correlação considerada é negativa.

Quando se comparam os valores do erro quadrático médio de estimação obtidos para todos os parâmetros do modelo considerando uma correlação positiva, com o mesmo modelo, mas com um grau de correlação negativo, verifica-se que os valores são menores se a correlação é negativa. Por outro lado, quando se comparam os valores do erro quadrático médio com correlação igual a 0 com os obtidos para uma correlação igual a -0.5 verifica-se que são menores no modelo com correlação negativa. Todavia, se a amostra tem dimensão superior a 50 observações os valores obtidos são semelhantes em ambas as situações.

Quando se comparam os valores do erro quadrático médio com correlação igual a 0 com os obtidos para uma correlação igual a 0.5, verifica-se que são menores no modelo com correlação nula, exceto para a média do intercepto que apresenta valores semelhantes em ambas as situações.

Os valores de *coverage* obtidos, qualquer que seja o grau de correlação considerado entre o intercepto e o declive, estão acima do considerado aceitável, constatando-se que o

parâmetro que apresenta menores valores é a variância do intercepto, qualquer que seja a dimensão da amostra.

Variação da fiabilidade dos indicadores

Com o objetivo de avaliar o efeito de um valor de fiabilidade dos indicadores inferior a 0.5, as variâncias dos termos residuais foram fixadas por forma a garantir uma fiabilidade dos indicadores de 0.4 em todos os momentos temporais. Os resultados obtidos permitem concluir que considerar um valor de fiabilidade dos indicadores mais baixo provoca um aumento no enviesamento das estimativas médias das variâncias do intercepto e do declive, bem como na estimativa média da covariância entre o intercepto e o declive. O enviesamento bruto nas estimativas médias das médias do intercepto e do declive não é afetado pela variação da fiabilidade do modelo. Os valores do erro quadrático médio de estimação aumentam quando o valor da fiabilidade dos indicadores diminui para todos os parâmetros do modelo. No entanto, os valores de *coverage* obtidos mantêm-se aceitáveis.

Variação do número de momentos temporais

A diminuição do número de momentos temporais no modelo, de 4 para 3, embora implique variações (aumento ou diminuição, de acordo com a situação) no enviesamento da estimativa média dos parâmetros do modelo, não apresenta valores de enviesamento superiores a 5%. A mesma situação se verifica quando se aumenta o número de momentos temporais no modelo de 4 para 5.

Por outro lado, verifica-se um aumento nos valores do erro quadrático médio de estimação para todos os parâmetros do modelo. Em sentido contrário, o aumento do número de momentos temporais no modelo com trajetória latente provocou uma diminuição dos valores do erro quadrático médio de estimação do conjunto dos parâmetros do modelo, exceto para a média do intercepto. Não se registam alterações do valor do erro quadrático médio da média do intercepto, aumentando o número de momentos temporais do modelo, qualquer que seja o valor populacional da variância do declive, o grau de correlação ou a dimensão da amostra considerada.

Existência de omissões nos dados

Foram geradas várias amostras a partir de um modelo com trajetória latente com quatro momentos temporais com distribuição normal e com omissões, configurando um *planned missing design*. Os valores obtidos de erro quadrático médio permitem verificar que quando os dados têm omissões, para as variâncias do intercepto e do declive e para a covariância entre intercepto e declive, estes valores são menores quando a correlação considerada no modelo é negativa. Para as médias do intercepto e do declive os valores

são análogos.

Se a correlação considerada é positiva, o enviesamento nas estimativas médias da média do intercepto, das variâncias do intercepto e do declive e da covariância entre intercepto e declive aumenta com a existência de omissões nos dados, qualquer que seja o valor populacional da variância do declive e a dimensão da amostra, exceto para a variância do intercepto quando a dimensão da amostra é de 250 observações. A média do declive não parece ser influenciada pela existência de omissões nos dados.

Quando se assume que não existe correlação entre o intercepto e o declive, então o enviesamento nas estimativas médias do conjunto dos parâmetros do modelo aumenta com a existência de omissões nos dados, exceto para a covariância entre intercepto e declive e variância do intercepto, quando a amostra tem dimensão de 250 observações. A média do declive apresenta valores de enviesamento semelhantes, independentemente da existência de omissões ou não.

Se a correlação entre intercepto e declive é negativa, não existe alteração no enviesamento das estimativas médias do conjunto dos parâmetros do modelo, exceto para a variância do declive, quando a amostra é de dimensão 500, e para a covariância entre o intercepto e o declive e variância do intercepto, sendo a amostra de pequena dimensão, isto é, de 250 observações.

De salientar que apesar de na maioria das situações o enviesamento nas estimativas médias dos parâmetros aumentar na presença de dados omissos, estes valores podem ser negligenciados (inferiores a 0.05), exceto o enviesamento na estimativa média da variância do declive, quando a correlação é negativa e a variância populacional do declive é elevada.

Os valores do erro quadrático médio de estimação para o conjunto de todos os parâmetros do modelo aumentam com a existência de omissões nos dados, qualquer que seja o modelo, a dimensão amostral considerada, e o grau de correlação entre fatores latente.

Os valores de *coverage* para todos os parâmetros do modelo são sempre superiores a 0.90, se os dados gerados apresentam distribuição normal, qualquer que seja a dimensão da amostra, o valor populacional dos parâmetros do modelo, o número de momentos temporais, ou a existência de omissões.

Quando se considera um modelo com apenas 3 momentos temporais, a existência de omissões nos dados provoca um aumento do enviesamento na estimativa média do conjunto dos parâmetros do modelo. Todavia, sendo a amostra de dimensão igual a 500, o enviesamento na estimativa média das médias do intercepto e do declive mantêm-se idêntico ao que se verifica na presença de dados completos.

Sublinha-se que a diminuição do número de momentos temporais no modelo provocou que, em amostras de pequena dimensão (50 observações), o enviesamento nas estimativas

médias da covariância e da variância do declive deixasse de ser negligenciável, quando a variância populacional toma um valor baixo, e que o enviesamento na estimativa média da variância do intercepto apresente um valor no limite do negligenciável, qualquer que seja o grau de correlação considerado.

Considerando um modelo com 5 momentos temporais, verifica-se um aumento no enviesamento das estimativas médias da covariância e das variâncias do intercepto e do declive, qualquer que seja a correlação considerada no modelo e a dimensão da amostra. O enviesamento na estimativa média da média do declive é semelhante com dados completos ou com omissões, em amostras de dimensão 500 observações, tendo apresentado uma diminuição, em amostras de dimensão 250 observações. O enviesamento na estimativa média da média do intercepto aumenta se o valor populacional da variância do declive é elevado e a correlação entre o intercepto e o declive é positiva.

5.5.2 Conclusões do estudo para dados gerados a partir de uma mistura de distribuições normais

Foi também considerada a geração de dados resultantes de uma mistura de distribuições normais recorrendo aos procedimentos disponíveis no *software* Mplus no âmbito dos modelos de mistura. No modelo de mistura são consideradas duas classes com diferentes percentagens de observações. Em cada uma das classes os dados são gerados a partir de um modelo com trajetória latente, apresentando conseqüentemente distribuição normal em cada classe.

Constata-se que assumir valores elevados para as médias e variâncias dos fatores latentes do modelo com 4 momentos temporais, numa das classes, enquanto na outra considerar valores baixos, permite obter diferentes níveis de assimetria e achatamento. Assim, foram considerados dados com valores para as referidas medidas a variar de fraco a severo. Foi estimado um modelo com trajetória latente com 4 momentos temporais, cujos valores populacionais dos seus parâmetros são previamente obtidos gerando uma amostra de 100 000 observações (dimensão muito elevada). O método de estimação considerado na presença de dados obtidos como uma mistura de distribuições normais foi o da máxima verosimilhança robusto.

Variação dos níveis de assimetria e achatamento

Os valores de enviesamento relativo na estimativa média da covariância entre intercepto e declive, bem como da média do declive permitem concluir que estes não podem ser considerados negligenciáveis. Por outro lado, o enviesamento relativo da estimativa média da variância do declive encontra-se no limite do considerado negligenciável, quando os dados apresentam assimetria e achatamento fracos, bem como quando apresentam assimetria suave a moderada e achatamento moderado, qualquer que seja a dimensão da

amostra.

Se os dados gerados, a partir de uma mistura de distribuições normais, apresentam assimetria fraca a suave, moderada ou severa, e achatamento suave a moderado, moderado ou severo, os valores de enviesamento relativo obtidos na estimação de um modelo com trajetória latente para todos os parâmetros do modelo podem ser considerados negligenciáveis.

Os valores do erro quadrático médio de estimação para todos os parâmetros do modelo são de um modo geral, mais elevados quando os dados gerados apresentam níveis de assimetria e achatamento severos, qualquer que seja a dimensão da amostra, sendo a variância do intercepto o parâmetro mais afetado. A covariância entre o intercepto e o declive apresenta valores mais elevados, quando a dimensão da amostra é pequena (50 observações) e os dados apresentam assimetria e achatamento moderados ou severos face a dados com valores de assimetria e achatamento mais baixos. A média do intercepto apresenta valores de erro quadrático médio de estimação semelhantes quando os dados têm distribuição provêm de uma mistura de distribuições normais com assimetria suave a moderada ou moderada, e achatamento moderado, bem como quando os dados têm assimetria fraca ou fraca a suave, e achatamento fraco ou suave a moderado. Os valores do erro quadrático médio de estimação da média e variância do declive são análogos, quando os dados apresentam assimetria fraca ou suave a moderada, e achatamento fraco ou moderado. O mesmo acontece para dados com assimetria fraca a suave, moderada ou severa, e achatamento moderado ou severo. Quando os dados apresentam assimetria e achatamento moderados ou severos, a variância do declive apresenta um valor de erro quadrático médio de estimação semelhante. Por outro lado, conclui-se que para todos os parâmetros do modelo os valores de erro quadrático médio de estimação diminuem com o aumento da dimensão da amostra.

Os valores de *coverage* para todos os parâmetros do modelo são baixos, se os dados provêm de uma mistura de distribuições normais e a amostra é de pequena dimensão (50 observações). Para dados com assimetria e achatamento severos, os parâmetros que apresentam piores resultados são a variância do intercepto e do declive, os quais apresentam valores abaixo de recomendável, e os parâmetros covariância entre intercepto e declive e a média do intercepto apresentam valores muito próximos do limite aceitável. Para dados com assimetria e achatamento moderados, a variância do declive apresenta valor abaixo do recomendável, enquanto que para a covariância e variância do intercepto os valores são baixos. Para dados com assimetria suave a moderada e achatamento moderado, os valores de *coverage* para as variâncias do intercepto e do declive apresentam valores baixos, próximo do limite do aceitável. Para dados com assimetria fraca a suave e achatamento suave a moderado as variâncias do intercepto e do declive estão abaixo do aceitável, sendo que a covariância apresenta um valor baixo. Para dados com assimetria e achatamento fracos, as variâncias do intercepto e declive e a covariância entre eles apresentam valores baixos.

Utilização do estimador ML

A utilização do estimador da máxima verosimilhança na estimação de um modelo com trajetória latente com 4 momentos temporais, a dados gerados como uma mistura de distribuições normais, provocou na covariância e nas variâncias do intercepto e do declive uma diminuição dos valores de *coverage* para valores abaixo do aceitável. Estes resultados revelaram-se ainda muito abaixo do aceitável quando a assimetria e achatamento dos dados são severos. O efeito da utilização de um estimador que não é robusto só não se manifesta quando os dados apresentam assimetria e achatamento fracos. De salientar que este efeito só se faz sentir nos valores de *coverage*, uma vez que apenas os erros padrão são calculados de forma diferente quando se considera uma versão robusta do estimador de máxima verosimilhança.

Existência de omissões nos dados

Considerando dados gerados a partir de uma mistura de distribuições normais e omissões, que constituem um *planned missing design*, o enviesamento na estimativa média de todos os parâmetros do modelo aumenta, qualquer que seja o nível de assimetria e achatamento presente nos dados, exceto para a média do declive se os dados apresentam assimetria e achatamento fracos, em amostras de dimensão 500. Em amostras de 250 observações, apresentando os dados assimetria fraca a suave, moderada ou severa e achatamento suave a moderado, moderado ou severo, verifica-se que o enviesamento aumenta com a existência de omissões nos dados. Para dados gerados com assimetria fraca ou suave a moderada e achatamento fraco ou moderado, e com omissões, o enviesamento na estimativa média da média e variância do intercepto aumenta, diminuindo o enviesamento na estimativa média da média e variância do declive. O enviesamento na estimativa média da covariância aumenta quando os dados são gerados com omissões e a partir de uma mistura de normais, com assimetria suave a moderada e achatamento moderado.

Os valores do erro quadrático de estimação para todos os parâmetros do modelo aumentam com a presença de omissões nos dados, independentemente do grau de não normalidade presente nos dados. Por outro lado, a variância do intercepto apresenta valores de *coverage* no limite do aceitável.

Quando se consideram dados com distribuição com assimetria fraca ou suave a moderada e achatamento fraco ou moderado, gerados com omissões que constituem um *planned missing design*, a partir da utilização de modelos com trajetória latente com 3 ou com 5 momentos temporais, verifica-se que os valores de enviesamento obtidos para a covariância e para a média e variância do declive não podem ser negligenciados, para amostras de dimensão igual a 250 observações. Para amostras com 500 observações, se os dados apresentam valores de assimetria e achatamento fracos, então o enviesamento obtido para a covariância e para a média do declive não pode ser negligenciado. Se os dados provêm de uma mistura de distribuições normais com assimetria suave a moderada e achatamento

moderado, o enviesamento na estimativa média do parâmetro média do declive é superior ao considerado negligenciável.

Tal como acontece para o modelo com 4 momentos temporais, também para o modelo com 3 momentos temporais, os valores do erro quadrático médio de estimação para todos os parâmetros são superiores se os dados apresentam omissões, qualquer que seja o grau de não normalidade dos dados. Todavia, o maior aumento verifica-se na estimação do modelo aos dados que apresentam assimetria e achatamento fraco. Se o modelo tem 5 momentos temporais, então os valores do erro quadrático médio são superiores quando os dados apresentam assimetria e achatamento severos, exceto para a média do declive, sendo as amostras de dimensão 500.

5.5.3 Conclusões do estudo para dados gerados com distribuição normal *vs* dados gerados a partir de uma mistura de distribuições normais

A comparação dos valores de enviesamento obtidos na estimação de dados resultantes de uma mistura de distribuições normais com os obtidos na estimação de dados com distribuição normal, permite concluir que o enviesamento relativo na estimativa média das variâncias do intercepto e do declive aumenta, sendo a dimensão da amostra elevada (1 000 observações). Por outro lado, se a amostra é de dimensão pequena (50 observações), o enviesamento relativo na estimativa média da variância do intercepto aumenta, qualquer que seja o nível de assimetria e de achatamento que os dados apresentam. O enviesamento relativo na estimativa média da variância do declive aumenta se os dados têm assimetria fraca ou suave a moderada e o achatamento é fraco ou moderado, diminuindo em caso contrário. Para as médias do intercepto e do declive, o enviesamento face ao seu valor populacional aumenta, qualquer que seja a dimensão do desvio face à normalidade dos dados, sendo a amostra de dimensão elevada. No entanto, se a amostra é de pequena dimensão, o enviesamento bruto da média do intercepto aumenta e o da média do declive diminui. O enviesamento bruto no valor estimado da covariância entre o intercepto e o declive aumenta, qualquer que seja a dimensão da amostra e o grau de não normalidade dos dados, exceto quando a amostra é de dimensão elevada e os dados apresentam assimetria e achatamento fracos.

De salientar que, se os dados apresentam assimetria e achatamento fraco ou assimetria suave a moderada e achatamento moderado, o enviesamento na estimativa média da covariância e da média do declive deixa de ser negligenciável, qualquer que seja a dimensão da amostra.

Os valores do erro quadrático médio de estimação aumentam para todos os parâmetros do modelo quando os dados resultam de uma mistura de distribuições normais, relativamente aos mesmos valores para dados com distribuição normal. Os valores de *coverage*, em particular para as variâncias do intercepto e do declive, são menores, se os dados

provêm de uma mistura de distribuições normais.

Quando se diminui o número de momentos temporais no modelo que gera os dados da mistura de distribuições normais, com diferentes valores de assimetria e achatamento, verifica-se que o enviesamento obtido na estimação desses dados aumenta para a variância do declive, diminuindo para a média do intercepto, em amostras de pequena dimensão. Para amostras de dimensão elevada, o enviesamento na estimativa média da média do intercepto diminui e o da variância do intercepto aumenta, quando os dados apresentam assimetria suave a moderada ou moderada, e achatamento moderado, caso contrário, observa-se uma diminuição. De salientar que o enviesamento na estimativa média da covariância e média do declive apresentam valores que não podem ser considerados negligenciáveis, se os dados provêm de uma mistura de distribuições normais com assimetria fraca ou suave a moderada e o achatamento é fraco ou moderado, respetivamente. Os valores do erro quadrático médio de estimação aumentam quando o número de momentos temporais no modelo diminui.

Por outro lado, o aumento do número de momentos temporais no modelo que gera os dados provoca uma diminuição no enviesamento da estimativa média de todos os parâmetros do modelo, exceto a variância do intercepto, em amostras de dimensão elevada. Para amostras de pequena dimensão, o enviesamento na estimativa média da variância do intercepto aumenta se os dados apresentam assimetria fraca ou fraca a suave, e achatamento fraco ou moderado. Relativamente, à média do intercepto observa-se uma diminuição, independentemente do grau de não normalidade dos dados, exceto se estes apresentam assimetria fraca a suave e achatamento suave a moderado. Para a média e variância do declive o enviesamento diminui se os dados apresentam assimetria fraca ou suave a moderada e achatamento fraco ou moderado, aumentando para dados com assimetria e achatamento moderado ou severo. Todavia, a covariância e a média do declive apresentam um enviesamento na sua estimativa média que não pode ser considerado negligenciável.

Os valores do erro quadrático médio de estimação diminuem com o aumento do número de momentos temporais no modelo, exceto a média do intercepto que não revela ser afetada pelo aumento do número de momentos temporais, quando os dados apresentam assimetria fraca ou fraca a suave e achatamento fraco ou suave a moderado. Todavia, se os dados apresentam assimetria e achatamento severos, o enviesamento na estimativa média da variância do intercepto aumenta, mantendo-se constante os valores das médias do intercepto e do declive, em amostras de dimensão elevada.

Capítulo 6

Conclusões

Recentemente tem-se assistido a uma cada vez maior utilização de modelos com variáveis latentes para dados longitudinais com o objetivo de descrever, explicar e quantificar a mudança. Dos vários modelos com variáveis latentes contínuas são de destacar os modelos com trajetória latente devido às inúmeras vantagens que este tipo de modelação longitudinal apresenta, no âmbito do estudo da mudança. Por outro lado, a existência de um cada vez maior número de bases de dados longitudinais, bem como o desenvolvimento de pacotes estatísticos utilizando estes modelos não é alheio a este crescimento.

Neste momento, o EUROSTAT tem a decorrer vários estudos longitudinais, sendo de destacar o EU-SILC, pela sua grande utilização por parte da comunidade científica em grande parte devido à importância que atualmente as temáticas subjacentes apresentam, sendo de destacar as condições de vida das famílias. O estudo da pobreza e da qualidade de vida das famílias pode ser realizado tendo por base o conceito de privação material, ao contrário da perspetiva tradicional de se basear apenas no rendimento disponível dos agregados. O ICOR - Inquérito às Condições de Vida e ao Rendimento para o qual existem dados longitudinais relativos aos anos de 2004 a 2010 e dados transversais para os anos de 2004 a 2011, resulta da participação da população portuguesa na base de dados estatística europeia EU-SILC.

A existência de dados omissos é um problema bastante comum em inquéritos, sendo praticamente impossível que este problema não ocorra se o inquérito em estudo é um painel, qualquer que seja a sua periodicidade. O ICOR apresenta diferentes tipos de omissões, as que são devido ao esquema de rotatividade do próprio painel e ainda outro tipo de omissões. Deste modo, a utilização de abordagens estatísticas para lidar com dados omissos apresenta-se essencial.

Os estudos realizados no âmbito desta tese foram divididos em quatro grupos distintos, primeiro os que consideraram dados completos, segundo os que consideraram dados com omissões cujo mecanismo de omissão de dados é MAR, terceiro os que consideraram dados com omissões cujo mecanismo é MNAR, e ainda um estudo de simulação.

6.1 Principais conclusões dos estudos realizados

Assim, numa primeira etapa foram apenas considerados para modelação dados com observações válidas e três diferentes estudos foram realizados. No **primeiro estudo**, uma vez que não se encontravam ainda disponíveis ao público os dados longitudinais do inquérito ICOR, foram considerados os dados transversais para os anos 2007 e 2008 na análise do conceito de privação material em quatro dimensões, capacidade financeira do agregado para satisfazer necessidades básicas, posse de bens duradouros, condições da habitação, e vizinhança, nos diferentes anos. Um modelo de análise fatorial confirmatória foi considerado para cada um dos diferentes anos, 2007 e 2008, com o objetivo de validar a estrutura proposta, bem como avaliar o peso das diferentes dimensões da privação material na definição deste conceito, não se tendo mostrado significativa a dimensão vizinhança.

De seguida, com o intuito de explicar o conceito de privação material em três diferentes dimensões, capacidade financeira, posse de bens e condições da habitação, é estimado um modelo de análise fatorial confirmatória considerando as três seguintes variáveis caracterizadoras dos agregados: a área de urbanização do local onde o agregado reside, a dimensão do agregado e a existência de crianças no agregado, em cada momento temporal. Para os anos de 2007 e 2008, é possível concluir que a privação é menor quando existem crianças no agregado e em agregados de maior dimensão. Por outro lado, verifica-se que no ano de 2007, em áreas muito povoadas e intermédias existe menor privação relativamente a áreas pouco povoadas, enquanto que, no ano de 2008, em áreas muito povoadas e intermédias existe maior privação relativamente a áreas pouco povoadas. Uma análise longitudinal dos dados é realizada recorrendo a uma análise fatorial confirmatória multigrupos no sentido de aferir da existência de comportamentos diferentes por parte dos agregados nos dois momentos temporais considerados, 2007 e 2008, uma vez que os dados subjacentes ao estudo são transversais. Este procedimento não permitiu concluir quanto à existência de diferenças na privação material nos dois momentos temporais considerados.

No **segundo estudo** apresentado são utilizados os dados longitudinais do ICOR dos anos 2004 a 2007 para estudar o conceito de privação material em duas dimensões, capacidade financeira do agregado para satisfazer necessidades básicas e posse de bens duradouros. Os referidos dados são utilizados na modelação longitudinal do conceito de privação material através de um modelo com equações estruturais com efeitos cruzados e desfasados, assim como de um modelo com trajetória latente bivariada. Posteriormente os respetivos resultados são comparados, procedendo-se ainda à avaliação das diferenças entre metodologias utilizadas. Os valores obtidos na estimação de um modelo com equações estruturais com efeitos cruzados e desfasados permitiu concluir que a capacidade financeira e a posse de bens duradouros, em cada momento temporal, é influenciada por si própria e pela outra variável, no momento temporal anterior, exceto a capacidade financeira, em 2005, a qual não mostrou influenciar a posse de bens em 2006.

Por outro lado, os valores obtidos na estimação de um modelo com trajetória latente bivariada permitiram concluir que a privação material média mantém-se constante entre

2004 e 2007. Todavia, os agregados variam na sua trajetória de mudança ao longo do período em análise, bem como no momento inicial (ano de 2004). Foi ainda possível verificar face aos valores obtidos que o nível inicial de capacidade financeira está associado com o nível inicial de posse de bens, e que níveis mais elevados de capacidade financeira, em 2004, estão relacionados com taxas médias de mudança da capacidade financeira mais lentas, ao longo do período de 2004 a 2007.

No **terceiro estudo** utilizaram-se os dados longitudinais do ICOR correspondentes aos anos de 2007 a 2010, com o intuito de analisar longitudinalmente o conceito latente de privação material recorrendo a modelos com trajetória latente de segunda ordem. A privação material foi considerada medida em três dimensões: constrangimentos económicos, posse de bens duradouros e condições da habitação. A estimação de um modelo com trajetória latente de segunda ordem permitiu concluir que a privação material, para o conjunto dos agregados, diminui entre 2007 e 2010. No ano de 2007 parece existir evidência de que os agregados diferem no nível de privação material, não sendo no entanto, possível afirmar que existem diferenças na taxa média de mudança da privação material para o conjunto dos agregados. Também não foi possível concluir que existe relação entre a privação material no ano de 2007 e a taxa média de mudança de 2007 a 2010. Uma vez descrita a trajetória da privação material no período em análise, são consideradas como variáveis explicativas da privação material e da sua trajetória, o rendimento do agregado, a dimensão do agregado e a área de urbanização. Deste modo, foi possível concluir que agregados com rendimentos mais elevados têm níveis de privação menores, enquanto que agregados com três ou mais filhos e que residem em grandes ou médias cidades sentem maior privação comparativamente aos que têm menor número de filhos e que habitam em meio rural, no ano de 2007, respetivamente. A variável logaritmo do rendimento foi ainda introduzida como explicativa da trajetória da privação, mas como variante no tempo, tendo sido possível verificar que, para os anos de 2007, 2008 e 2010, valores mais elevados de rendimento implicam menor privação material.

Numa segunda etapa foram considerados dados com omissões que resultam do esquema de rotatividade, podendo considerar-se estas omissões como tendo um mecanismo MAR, conforme apresentado nos estudos 4 e 5.

No **quarto estudo** foi proposta a modelação através de um modelo com trajetória latente de um indicador de privação material para os anos de 2006 a 2009, para um conjunto de agregados que apresentam 50% de omissões no quarto momento temporal. Uma abordagem de *full information maximum likelihood* foi considerada no sentido de lidar com estas omissões na estimação da trajetória latente do *score* de privação material entre 2006 e 2009, considerando a totalidade dos agregados e os resultados comparados com os obtidos na estimação da trajetória latente do *score* de privação material entre 2006 e 2009 para os agregados que se mantiveram no inquérito durante os quatro momentos temporais (isto é, aqueles que apresentam dados completos). Todavia, os resultados mostraram-se contraditórios, pois para o conjunto dos agregados que constituíam a amostra com dados omissos foi possível concluir pela existência de uma diminuição na privação material entre

2006 e 2009, enquanto que para o conjunto dos agregados que se mantiveram no estudo entre 2006 e 2009 (amostra com dados completos) os resultados não permitiram retirar conclusão idêntica. Consequentemente, foi realizado um pequeno estudo de simulação recorrendo a procedimentos de Monte Carlo. Assumindo como verdadeiros os valores dos parâmetros estimados pelo modelo com trajetória latente para os dados completos, foram geradas a partir desta estrutura amostras de dados com um padrão de omissão de 50% no último momento temporal, ano de 2009. Para cada uma das diferentes amostras foram estimados utilizando o método do *full information maximum likelihood* os parâmetros do modelo com trajetória latente, verificando-se que o parâmetro cuja estimativa média apresenta um maior enviesamento é a média do declive e , para este mesmo parâmetro, existe uma percentagem de amostras geradas, nas quais este se mostrou significativo.

No **quinto estudo** as duas diferentes abordagens estatísticas, *full information maximum likelihood* e imputação múltipla são utilizadas para lidar com as omissões existentes nos dados. Neste trabalho o objetivo foi comparar os resultados obtidos na estimação dos parâmetros de um modelo com trajetória latente (com quatro momentos temporais) para um *score* de privação material, na presença de dados omissos que configuram um *planned missing design*, isto é, os dados do painel rotativo ICOR. Distintas subamostras com diferentes desenhos amostrais foram consideradas. As duas abordagens estatísticas, *full information maximum likelihood* e imputação múltipla são utilizadas para lidar com as omissões. As estimativas dos parâmetros obtidas por aplicação das duas diferentes abordagens estatísticas na presença de dados omissos, para as várias subamostras, foram comparadas entre si, bem como com aquelas que resultaram da estimação do modelo com dados completos. Apesar dos resultados obtidos pelo recurso às duas metodologias para lidar com omissões serem análogas, os resultados mostraram-se contraditórios quando se compararam os valores obtidos para a subamostra com dados completos e as outras subamostras. Tal como no estudo quatro recorreu-se a procedimentos de Monte Carlo para gerar dados com omissões semelhantes às das diferentes subamostras, tendo-se verificado existir uma percentagem de amostras geradas, nas quais a média do declive é significativo.

Numa terceira etapa foram considerados dados com omissões com um mecanismo MNAR, correspondendo aos estudos 6 e 7.

Os estudos 6 e 7, embora realizados em momentos diferentes e apresentados em diferentes congressos da área, utilizam os mesmos dados, os dados longitudinais 2006 a 2009, e complementam-se em termos metodológicos. Assim, nesta tese optou-se por apresentar os seus resultados na mesma secção, secção 4.4.3.

No **sexto estudo** são considerados os dados longitudinais 2006 a 2009. Todavia, no painel rotativo ICOR existem, além das omissões devido ao seu esquema de rotatividade, outros tipos de omissões, tais como saídas antes do tempo, entradas tardias ou agregados que não respondem em determinado momento temporal e voltam a responder no momento seguinte, resultando assim em diferentes padrões de omissão. Assim, as não respostas existentes neste painel apresentam diferentes mecanismos de omissão, tendo sido consideradas

MAR as omissões que resultam do esquema de rotatividade do painel e MNAR as outras omissões.

Neste estudo foi proposta a modelação através de um modelo com trajetória latente de um indicador de privação material para os anos de 2006 a 2009, primeiro considerando apenas os agregados que se mantiveram no painel durante os quatro anos, e de seguida considerando a totalidade dos agregados. A modelação da trajetória latente do indicador de privação material entre 2006 e 2009, considerando a totalidade dos agregados, foi realizada recorrendo a uma abordagem *full information maximum likelihood*. Posteriormente, uma abordagem *bootstrap* permitiu obter intervalos de confiança não paramétricos para os parâmetros estimados com diferentes níveis de confiança. Uma abordagem *pattern mixture model* foi considerada com o objetivo de incorporar na modelação informação sobre a distribuição das omissões. Para tal, os diferentes padrões de omissões observados na base de dados em estudo foram divididos em dois grupos, os que resultam da rotatividade do painel e os outros, uma vez que se acredita que estes dois grupos têm mecanismos de omissão distintos. Outras divisões foram ainda discutidas, tendo-se concluído que um modelo considerando que os dados apresentam dois padrões (padrão “caso completo+MAR” e padrão “MNAR”) apresentava melhor ajustamento. A utilização de um *pattern mixture model* permitiu confirmar que os agregados experimentam uma diminuição da privação material entre 2006 e 2009, conclusão semelhante à obtida com recurso a um *full information maximum likelihood*.

No **sétimo estudo** foi modelada a trajetória de um indicador de privação material, no período de 2006 a 2009, recorrendo a modelos com trajetória latente. Primeiro foram apenas considerados os agregados que se mantiveram no painel entre 2006 e 2009 e posteriormente foram considerados todos os agregados (incluindo os que apresentam omissões). Dado que na base de dados longitudinal 2006 a 2009 do ICOR existem omissões devido ao desenho do estudo, existindo igualmente outro tipo de não respostas. Diferentes abordagens estatísticas para lidar com a existência dos vários tipos de omissões presentes nos dados foram consideradas. Uma abordagem com *full information maximum likelihood* assumindo um mecanismo de omissão MAR é considerada, sendo discutida a utilização dos *pattern mixture model* e dos *selection model* se o mecanismo de omissão é MNAR. Padrões de omissão distintos foram levados em consideração na modelação de uma trajetória de mudança do indicador de privação material e a estabilidade dos resultados obtidos nas diferentes abordagens foi comparada. No *pattern mixture model* duas parametrizações distintas foram utilizadas, sendo que os resultados obtidos em ambas foram coincidentes e permitiram corroborar os resultados do estudo seis.

Assim, os estudos seis e sete permitem verificar que as conclusões são idênticas para o conjunto total dos agregados, independentemente do procedimento estatístico para lidar com as omissões presentes no painel ICOR se basear num mecanismo de omissão MAR ou MNAR.

Numa quarta etapa foi elaborado um **estudo de simulação** recorrendo a procedi-

mentos de Monte Carlo. Este estudo mostrou-se necessário uma vez que nos estudos 4 a 7 as abordagens estatísticas utilizadas para lidar com as omissões existentes nos dados, assentam no pressuposto de normalidade dos dados. Por outro lado, uma análise ao comportamento do indicador de privação material, no período de 2006 a 2009, mostrou que este não apresenta uma distribuição normal. Também a presença de omissões devido ao desenho do inquérito ICOR ou o insuficiente número de observações na amostra, ou mesmo o número de momentos temporais do modelo, podem motivar conclusões contraditórias entre “caso completo” e “caso com todas as observações”, nos estudos anteriores.

Assim, neste estudo de simulação foi analisado o efeito da não normalidade dos dados na estimação de um modelo com trajetória latente, o efeito da dimensão da amostra, do número de momentos temporais do modelo, do tipo de estimador utilizado e de diferentes padrões de omissão (resultantes de um *planned missing design*) com diferentes percentagens de não respostas em cada um dos momentos temporais.

Os resultados permitem concluir que o enviesamento obtido na estimativa média dos vários parâmetros do modelo pode ser negligenciável, se os dados são gerados com distribuição normal a partir de um modelo com trajetória latente com 4 momentos temporais, qualquer que seja a dimensão da amostra, o grau de correlação, a média e a variância do declive considerados no modelo, bem como o número de momentos temporais. Todavia, se são consideradas omissões nos dados com distribuição normal, verifica-se que o enviesamento na estimativa da variância do declive se encontra acima do recomendável, sendo a variância populacional do declive elevada, num modelo com 4 momentos temporais, em amostras de pequena dimensão.

Se os dados são gerados com distribuição normal a partir de um modelo com 3 momentos temporais (pequeno número de momentos temporais) verifica-se que o enviesamento na estimativa da covariância e variância do declive, se a variância populacional do declive é baixa, e na estimativa da variância do intercepto, se a variância populacional do declive é elevada, assume valores que não podem ser considerados negligenciáveis, em amostras de pequena dimensão.

Os valores do erro quadrático médio de estimação para os parâmetros do modelo são menores em amostras de dimensão mais elevada, em modelos que apresentam maior número de momentos temporais, e em modelos em que se considera correlação entre fatores latentes negativa.

Se os dados têm distribuição não normal com assimetria e achatamento fracos ou assimetria suave a moderada e achatamento moderado, o enviesamento obtido na estimativa média da covariância e da média do declive não pode ser negligenciável, independentemente do número de momentos temporais do modelo que gera os dados. Esta mesma situação repete-se se os dados apresentam omissões resultantes de um *planned missing design*, tal como descrito no estudo apresentado por Enders [40]. Também como no estudo referido, os maiores valores de enviesamento não ocorreram para dados com assimetria e

achatamento severos.

Os valores do erro quadrático médio de estimação são mais elevados em dados com distribuição não normal e aumentam à medida que se acentua a assimetria e o achatamento, assim como se os dados apresentam omissões. O uso de um método de estimação robusto é desejável quando os dados têm distribuição não normal, permitindo obter valores de *coverage* acima do aceitável para o conjunto de parâmetros do modelo. Em particular, a covariância e as variâncias do intercepto e do declive são os parâmetros que apresentam valores de *coverage* inferiores ao limite aceitável, se o estimador utilizado não é robusto, mesmo em amostras grandes.

Assim, o estudo de simulação realizado permitiu perceber que a média do declive, o parâmetro para o qual existia uma contradição nos resultados dos estudos 4, 5, 6 e 7, na presença de dados com distribuição não normal, cujos valores de assimetria e achatamento são fracos a moderados, apresenta um enviesamento na sua estimativa que não pode ser negligenciada. Esta situação persiste se nos dados existem omissões resultantes de um *planned missing design*, tal como acontece no painel ICOR.

6.2 Principais limitações

Na realização dos vários estudos que compõem esta tese o atraso verificado na disponibilização dos dados longitudinais da participação portuguesa no EU-SILC (ICOR) por parte do INE, constituiu uma primeira limitação. Como consequência, e porque na altura não estavam disponíveis dados longitudinais, no primeiro estudo foi realizada uma análise longitudinal da privação material utilizando as duas mais recentes bases de dados transversais que se encontravam disponíveis para os utilizadores. Todavia, este atraso na entrega dos dados acabou por criar desafios metodológicos, tendo sido possível utilizar outros modelos na modelação longitudinal da privação material.

Por outro lado, a utilização de dados longitudinais de 2006 a 2009, quando estamos em 2014 pode parecer uma limitação desta tese. Todavia, aquando da participação na 3rd EU-SILC Users Conference que se realizou em março de 2013, o estudo aí apresentado (estudo 7) era o que, de todos os participantes na conferência, utilizava dados longitudinais mais recentes. De facto, existe um espaçamento temporal elevado entre a recolha dos dados e a sua disponibilização aos investigadores, em particular das bases de dados longitudinais. De referir ainda que, por razões legais no EU-SILC (e consequentemente no ICOR) não é possível cruzar dados das bases transversais por forma a obter as bases de dados longitudinais.

A utilização de um indicador de privação material calculado como uma soma de itens, e não o considerar como uma contagem, pode também ser uma limitação desta tese, mas este tem sido o procedimento usual no estudo desta temática, Guio [60] e Whelan e Maître [167]. De salientar ainda que, do ponto de vista da comparação das condições de vida dos

agregados dos diferentes países que compõem a União Europeia, a existência e utilização de um indicador de privação material mostra-se de grande utilidade.

Do ponto de vista metodológico, uma limitação desta tese é a utilização de um modelo com trajetória latente linear de primeira ordem em vez de um modelo mais complexo, tal como, um modelo com trajetória latente de segunda ordem, ou um modelo com trajetória não linear, no estudo de simulação e nos estudos em que se pretendeu lidar com as não respostas que o inquérito ICOR apresenta (estudos 4 a 7). Tal facto, decorreu do objetivo de modelar um indicador (único) de privação material, que é usualmente calculado como uma soma de itens. Por outro lado, a existência de apenas quatro momentos temporais para a modelação longitudinal da privação material para os agregados que constituíam o “caso completo” (considerado como referência), tornou difícil o recurso a um modelo com trajetória latente não linear.

A condução de um estudo de simulação recorrendo à utilização de um *software* comercial, o Mplus, que tem as suas próprias limitações, pode ainda constituir uma limitação deste trabalho. A sua utilização deveu-se às potencialidades que este pacote estatístico apresenta na modelação de dados longitudinais recorrendo aos modelos com trajetória latente, bem como à sua crescente popularidade na comunidade estatística internacional.

6.3 Perspetivas para trabalho futuro

Esta tese, embora responda a algumas questões, também deixa algumas pistas em aberto para trabalho futuro.

Assim, em termos de trabalho futuro, será de continuar a tirar partido das potencialidades que o pacote estatístico Mplus apresenta quanto à geração de dados, a partir de um modelo com trajetória latente, que apresentem outro tipo de omissões, para além das que configuram um *planned missing design*. No estudo de simulação apresentado foram geradas amostras de dados com omissões, que constituíam um *planned missing design*. Todavia, o ICOR apresenta outro tipo de omissões para as quais um mecanismo MNAR não pode ser descartado. Deste modo, gerar amostras de dados com omissões cujo mecanismo é não ignorável, e posterior análise e avaliação das estimativas obtidas para os vários parâmetros do modelo, é uma questão de interesse.

Ainda, fazendo uso das potencialidades do *software* Mplus, deve ser considerada a geração de amostras de dados com distribuição não normal, com diferentes graus de assimetria e achatamento, obtidas fazendo variar a percentagem de observações em cada uma das classes no modelo de mistura utilizado.

Uma outra pista a ser seguida em trabalho futuro é a utilização de um modelo com trajetória latente mais complexo no estudo de simulação, como por exemplo, um modelo com trajetória latente não linear, considerando dados longitudinais com um maior

número de momentos temporais, ou um modelo com trajetória latente de segunda ordem, considerando três indicadores em cada momento temporal, tal como no estudo 3, em que foram considerados três indicadores da privação material (correspondendo cada um a uma diferente dimensão deste conceito).

Por outro lado, a utilização de dados mais recentes pode contribuir para evidenciar uma mudança na taxa média de privação material que os agregados experimentam. Também o cálculo do indicador de privação material deve ser revisto, e outras variáveis devem ser consideradas no cálculo deste indicador, uma vez que alguns dos itens que compõem este indicador correspondem a bens que conseguem ser assegurados por uma percentagem muito elevada da população, como por exemplo, a posse de televisão ou telefone.

Bibliografia

- [1] Albert, P.S. e Follmann, D.A.(2009), Shared-parameter models. Em Fitzmaurice, G., Davidian, M., Verbeek, G. and Molenberghs, G.(eds.), *Longitudinal Data Analysis*, 433-452. Chapman and Hall.
- [2] Allison, P.D.(1987), Estimation of linear models with incomplete data. *Sociological Methodology*, 17, 71-103.
- [3] Allison, P.D.(2002), *Missing Data*. Sage.
- [4] Amemiya, T.(1985), *Advance econometrics*. Harvard University Press.
- [5] Arbuckle, J.L.(1996), Full information estimation in the presence of incomplete data. Em Marcoulides, G.A. e Schumacker, R.E.(eds.), *Advanced structural equation modeling*, 243-277. Erlbaum.
- [6] Arbuckle, J.L.(2012): *AMOS 21.0*. Chicago: SPSS, Inc..
- [7] Azar, B.(2002), Finding a solution for missing data. *Monitor on Psychology*, 33, 70.
- [8] Baker, G.A.(1954), Factor analysis of relative growth. *Growth*, 18, 137-143.
- [9] Bandalos, D.L. e Leite, W.(2013), Use of Monte Carlo studies in Structural Equation Modeling Research, em *Structural Equation Modeling: A Second Course*, 2nd edition. Information Age Publishing, 625-666.
- [10] Baraldi, A.N. e Enders, C.K.(2010), An introduction to modern missing data analysis. *Journal of School Psychology*, 48, 5-37.
- [11] Blalock, H.M.(1971), *Causal Models in the Social Sciences*. Aldine-Atherton.
- [12] Bentler, P.M.(1980), Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, 31, 419-456.
- [13] Bentler, P.M.(1986), Structural modeling and Psychometrika: An historical perspective on growth and achievements. *Psychometrika*, 51, 35-51.
- [14] Bentler, P.M. e Wu, E.(2005), *EQS 6*. Encino, CA: Multivariate Software, Inc.
- [15] Bentler, P.M. e Yuan, K.H.(1999), Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, 34, 181-197.

-
- [16] Betti, G. e Verma, V.(2004), A methodology for the study of multi-dimensional and longitudinal aspects of poverty and deprivation. *Proceedings do IAOS-IASS Joint Conference*.
- [17] Bertchold, A. e Jeannin, A.(2008), Imputation in data fusion of heterogeneous data sets a model-based numerical experiment. *Communications in Statistics - Simulation and Computation*, 37, 1316-1328.
- [18] Bollen, K.(1989), *Structural Equations with Latent Variables*. Wiley, New York.
- [19] Bollen, K.(2002), Latent Variables in Psychology and the Social Sciences. *Annual Review of Psycology*, 53, 605-634.
- [20] Bollen, K.A. e Curran, P.J.(2004), Autoregressive latent trajectory models: a synthesis of two traditions. *Sociological Methods and Research*, 32, 336-383.
- [21] Bollen, K.A. e Curran, P.J.(2006), *Latent Curve Models - A Structural Equation Perspective*. John Wiley & Sons, New Jersey.
- [22] Boomsma, A.(2013), Reporting Monte Carlo Studies in Structural Equation Modeling. *Structural Equation Modeling*, 20(3), 518-540.
- [23] Browne, M.W.(1984), Assymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37(1), 62-83.
- [24] Byrne, B.M.(2011), *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming*. Routledge Academic.
- [25] Collins, L.M.(2006), Analysis of Longitudinal Data: The Integration of Theoretical Model, Temporal Design and Statistical Model. *Annual Review of Psychology*, 57, 505-528.
- [26] Collins, L.M. e Horn, J.L.(eds)(1991), *Best Methods for the Analysis of Change*. American Psychologist Association.
- [27] Collins, L.M., Schafer, J.L. e Kam, C-M.(2001), A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6, 330-351.
- [28] Davidov, E., Thörner, S., Schmidt, P., Gosen, S. e Wolf, C.(2011), Level and change of group-focused enmity in Germany: unconditional and conditional latent growth curve models with four panel waves, *Adv. Stat. Analysis* , 95, 481-500.
- [29] Dekkers, G.J.M.(2008), Are you unhappy? Then you are poor! Multi-dimensional poverty in Belgium, *International Journal of Sociology* , 11, 502-515.
- [30] Demirtas, H.(2004), Modeling incomplete longitudinal data. *Journal of Modern Applied Statistical Methods*, 3, 305-321.

-
- [31] Demirtas, H. e Schaffer, J.L.(2003), On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine*, 22, 2553-2575.
- [32] Dempster, A., Laird, N. e Rubin, D.(1977), Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39, 1-38.
- [33] Diggle, P. e Kenward, M.G.(1994), Informative drop-out in longitudinal data analysis. *Applied Statistics*, 43, 49-93.
- [34] Duncan, T., Duncan, S. e Strycker, L.(2006a), *An introduction to latent variable growth curve modeling: concepts, issues and applications*. Lawrence Erlbaum Associates, Inc.
- [35] Duncan, T., Duncan, S. e Strycker, L.(2006b), Alcohol Use from Ages 9-16: A Cohort-Sequential Latent Growth Model. *Drug Alcohol Depend.*, 81, 71-81.
- [36] Duncan, S.C., Duncan, T.E. e Hops, H.(1994), The effect of family cohesiveness and peer encouragement on the development of adolescent alcohol use: A cohort-sequential approach to the analysis of longitudinal data. *Journal of Studies on Alcohol*, 55, 588-599.
- [37] Duncan, S.C., Duncan, T.E. e Hops, H.(1996), Analysis of longitudinal data within accelerated longitudinal designs. *Psychological Methods*, 1, 236-248.
- [38] Duncan, S.C., Duncan, T.E. e Stoolmiller, M.(1994), Modeling developmental processes using latent growth structural equation methodology. *Appl. Psychol. Meas.*, 18, 343-354.
- [39] Efron, B. e Tibshirani, R.J.(2002), *An introduction to the Bootstrap*. Chapman & Hall.
- [40] Enders, C.K.(2001), A primer on maximum likelihood algorithms for use with missing data. *Structural Equation Modeling*, 8, 128-141.
- [41] Enders, C.K.(2010), *Applied Missing Data*. The Guilford Press.
- [42] Enders, C.K.(2011), Missing not at random models for latent growth curve analysis. *Psychological Methods*, 16, 17-33.
- [43] Enders, C.K. e Bandalos, D.L.(2001), The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 8, 430-457.
- [44] Eroglu, S.(2007), Developing an index of deprivation which integrates objective and subjective dimensions: Extending the work of Townsend, Mack and Lansley, and Halleröd. *Social Indicators Research*, 80, 493-510.

-
- [45] EUROSTAT(2006), Comparative EU statistics on Income and Living Conditions: Issues and Challenges. *Proceedings of the EU-SILC conference*.
- [46] Fan, X., Thompson, B. e Wang, L.(1999), Effects of sample size, estimation methods and model specification on SEM fit indices. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 56-83.
- [47] Ferrer, E., Balluerka, N. e Widaman, K.(2008), Factorial Invariance and the Specification of Second-Order Latent Growth Models, *Methodology*, 4(1), 22-36.
- [48] Finkel, S.(1995), *Causal Analysis with Panel Data*. CA: Sage, Thousand Oaks.
- [49] Finkbeiner, C.(1979), Estimation for the multiple group factor model when data are missing. *Psychometrika*, 44, 409-420.
- [50] Golderberg, A.S e Duncan, O.D.(1973), *Structural equation models in the social sciences*. Academic Press.
- [51] Gompertz, B.(1820), A sketch of an analysis and notation applicable to the estimation of the value of life contingencies. *Philosophical Transactions of the Royal Society*, 110, 214-294.
- [52] Gompertz, B.(1825), On the nature of the function expressive of the law of human mortality. *Philosophical Transactions of the Royal Society*, 115, 513-583.
- [53] Graham, J.(2009), Missing Data Analysis: Making it Work in the Real World. *Annual Review of Psychology*, 60, 549-576.
- [54] Graham, J. e Donaldson, S.(1993), Evaluation interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data. *Journal of Applied Psychology*, 78, 119-128.
- [55] Graham, J., Hofer, S. e Mackinnon, D.(1996), Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, 31, 197-218.
- [56] Graham, J., Taylor, B, Olchowski, A. e Cumsille, P.(2006), Planned missing data designs in psychological research. *Psychological Methods*, 11, 323-343.
- [57] Graham, J., Olchowski, A.E. e Gilreath, T.D.(2007), How many imputations are really needed? Some practical clarification of multiple imputation theory. *Prev.Sci.*, 8, 206-213.
- [58] Griliches, Z.(1957), Hybrid corn: an exploration in the economics of technical change. *Econometrica*, 48, 501-522.
- [59] Grover, R. e Vriens, M.(2006), *The handbook of marketing research: uses, misuses and future advances*. Sage Publications, Inc.

-
- [60] Guio, A-C.(2009), What can be learned from material deprivation indicators in Belgium and in its regions. *Institut Wallon de L' Evaluation de la Prospective et de la Statistique* 901.
- [61] Guio, A-C.(2005), Material Deprivation in the EU. *Statistics in Focus, Population and Social Conditions and Welfare*, 21, EUROSTAR.
- [62] Halleröd, B., Gordon, D., Larsson, D. e Ritakallio, V.M.(2006), Relative deprivation: a comparative analysis of Britain, Finland and Sweden. *Journal of European Social Policy*, 16, 328-345.
- [63] Hair, J., Anderson, R., Tatham, R. e Black, W.(2009), *Multivariate Data Analysis*, 7th edition. Prentice-Hall.
- [64] Harwell, M.R.(1997), Analyzing the results of Monte Carlo studies in Item Response Theory. *Educational and Psychological Measurement*, 57, 226-279.
- [65] Heckman, J.T.(1976), The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *The Annals of Economic and Social Measurement*, 5, 475-492.
- [66] Heckman, J.T.(1979), Sample selection bias as a specification error. *Econometrica*, 47, 153-161.
- [67] Hedeker, D. e Gibbons, R.D.(1997), Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2, 64-78.
- [68] Hoogland, J.J. e Boomsma, A.(1998), Robustness studies in covariance structure modelling: An overview and meta-analysis. *Sociological Methods and Research*, 26, 329-367.
- [69] Hsiao, C.(2003), *Analysis of Panel Data*. Cambridge University Press.
- [70] Humphreys, K. e Janson, H.(2000), Latent Transition Analysis with Covariates, Nonresponse, Summary Statistics and Diagnostics: Modelling Children's Drawing Development. *Multivariate Behavioral Research*, 35, 89-118.
- [71] Ibrahim, J.G. e Molenberghs, G.(2009), Missing data methods in longitudinal studies: a review. *Test*, 18, 1-43.
- [72] INE(2009), Inquérito às Condições de Vida e Rendimento-ICOR. *Documento Metodológico*.
- [73] Johnson, E.G.(1992), The Design of the National Assessment of educational Progress. *Journal of Educational Measurement*, 29, 95-110.
- [74] Jöreskog, K.G.(1969), A general approach to confirmatory maximum-likelihood factor analysis. *Psychometrika*, 34, 183-202.

-
- [75] Jöreskog, K.G.(1973), A general method for estimating a linear structural equation system, em A.S Goldergee O.D. Duncan (eds), *Structural equation models in the social sciences*. Academic Press, 85-112.
- [76] Jöreskog, K.G.(1979), Statistical estimation of structural models in longitudinal developmental investigation. Em J.R Nesselrode e P.B. Baltes (eds), *Longitudinal Research in the Study of Behaviorand Development*. Academic Press, 303-352.
- [77] Jöreskog, K.G. e Sörbom, D.(2004), LISREL 8. Chicago, IL: Scientific Software International, Inc.
- [78] Kamakura, W.A. e Wedel, M.(2000), Factor analysis and missing data. *Journal of Marketing Research*, 37, 490-498.
- [79] Kaplan, D.(2000), *Structural equation modelling: Foundations and extensions*. CA:Sage.
- [80] Keesling, J.W.(1972), *Maximum Likelihood Approaches to Causal Analysis*. Ph.D. dissertation. Department of Education.
- [81] Kline, R.B.(1998), *Principles and pratices of structural equation modeling*. Guilford.
- [82] Lavori, P., Dawson, R. e Shera, D.(1995), A Multiple Imputation Strategy for Clinical trials with Truncation of Patient Data. *Statistics in Medicine*, 14, 1913-1925.
- [83] Lee, S.Y.(1986), Estimation for structural equation models with missing data. *Psychometrika*, 45, 309-324.
- [84] Little, R. e Rubin, D.B.(1987), *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc..
- [85] MacCallum, R. e Austin, J.(2000), Applications of Structural Equation Modeling in Psychological Research. *Annual Review of Psychology*, 51, 201-226.
- [86] Mack, J. e Lansley, S.(1985), *Poor Britain*. George Allen & Unwin, ltd.
- [87] McArdle, J.J.(1994), Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research*, 29, 409-454.
- [88] McArdle, J.J.(2009), Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577-605.
- [89] McArdle, J.J. e Anderson, E.(1990), Latent variable growth models for research on aging. Em J.E.Birren, K.W.Schaie, *Handbook of the Psychology of Aging*, 21-44.
- [90] McArdle, J.J. e Epstein, D.(1987), Latent growth curves within developmental structural equation models. *Child Development*, 58, 110-133.

-
- [91] McArdle, J.J. e Hamagami, F.(1992), Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. *Experimental Aging Research*, 18, 145-166.
- [92] McArdle, J.J. e McDonald, R.P.(1984), Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology*, 37, 234-251.
- [93] Meredith, W. e Tisak, J.(1984), On “Tuckerizing” curves. *Annual meeting of the Psychometric Society*, Santa Barbara, CA.
- [94] Meredith, W. e Tisak, J.(1990), Latent Curve Analysis. *Psychometrika*, 55, 107-122.
- [95] Molenberghs, G. e Fitzmaurice, G.(2009), Incomplete data: Introduction and overview. Em Fitzmaurice, G., Davidian, M., Verbeek, G. and Molenberghs, G.(eds.), *Longitudinal Data Analysis*. Chapman and Hall.
- [96] Molenberghs, G., Michiels, B., Kenward, M.G. e Diggle, P.J.(1998), Monotone missing data and pattern-mixture models. *Statistica Neerlandica*, 52, 153-161.
- [97] Molenberghs, G., Michiels, B. e Lipsitz, S.R.(1999), Selection models and pattern-mixture models for incomplete categorical data with covariates. *Biometrics*, 55, 978-983.
- [98] Mooney, C.Z.(1997), *Monte Carlo simulation*. Thousand Oaks, CA.
- [99] Muthén, B.O.(1998-2004), *Mplus Technical Appendices*. Los Angeles, CA: Muthén & Muthén.
- [100] Muthén, B.O. e Asparouhov, T.(2002a), *Using Mplus Monte Carlo simulations in practice: A note on assessing estimation quality and power in latent variable models*. Mplus Web Notes, n^o1, version 2.
- [101] Muthén, B.O. e Asparouhov, T.(2002b), *Using Mplus Monte Carlo simulations in practice: A note on non-normal missing data in latent variable models*. Mplus Web Notes, n^o2, version 2.
- [102] Muthén, B.O., Asparouhov, T., Hunter, A. e Leuchter, A.(2011), Growth modeling with STAR*D Antidepressant Trial. *Psychological Methods*, 16, 17-33.
- [103] Muthén, B.O. e Kaplan, D.(1985), A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171-189.
- [104] Muthén, B.O., Kaplan, D. e Hollis, M.(1987), On structural equation modelling with data that are not missing completely at random. *Psychometrika*, 52, 431-462.
- [105] Muthén, L.K. e Muthén, B.O.(1998-2010), *Mplus user's guide*, 6th edition. Los Angeles, CA: Muthén & Muthén.

-
- [106] Muthén, L.K. e Muthén, B.O.(2002), How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599-620.
- [107] Nesselroad, J.R.(1991), Interindividual differences in intraindividual change. Em Collins, L.M. e Horn, J.L.,(eds), *Best Methods for the Analysis of Change*. Am. Psychol.Association.
- [108] Nesselroad, J.R. e Baltes, P.B.(1979), *Longitudinal research in the study of behavior and development*. Academic Press.
- [109] Newman, D.A.(2003), Longitudinal modeling with randomly and systematically missing data. *Organizational Research Methods*, 6, 328-362.
- [110] Nolan, B. e Whelan, C.T.(2010), Using Non-Monetary deprivation Indicators to Analyze Poverty and Social Exclusion: Lesson from Europe?. *Journal of Policy Analysis and Management*, 29, 305-325.
- [111] Nylund, K.L., Asparouhov, T. e Muthén, B.(2007), Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modelling*, 7, 557-595.
- [112] Olsson, U.H., Troye, S.V. e Howell, R.D.(1999), Theoretical fit and empirical fit: The performance of maximum likelihood versus generalized least squares estimation in structural equation models. *Multivariate Behavioral Research*, 34, 31-58.
- [113] Paxton, P., Curran, P., Bollen, K., Kirby, J. e Chen, F.(2001), Monte Carlo Experiments: Design and Implementation. *Structural Equation Modeling*, 8, 287-312.
- [114] Peters, C.O. e Enders, C.(2002), A primer for the estimation of structural equation models in the presence of missing data: Maximum likelihood algorithms. *Journal of Targeting, Measurement and Analysis for Marketing*, 11, 81-95.
- [115] Peugh, J.L. e Enders, C.K.(2005), Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74, 525-556.
- [116] Powell, D.A. e Schafer, W.D.(2001), The robustness of the likelihood ratio chi-square test for structural equation models. A meta-analysis. *Journal of Educational and Behavioral Statistics*, 26, 105-132.
- [117] Preacher, K.J., Wichman, A.L., MacCallum, R.C. e Briggs, N.E.(2008), *Latent Growth Curve Modelling*. Sage Publications, Inc.
- [118] Raudenbush, S.W. & Bryck, A.S.(2002), *Hierarchical Linear Models*. CA:Sage Publications.
- [119] Raghunathan, T.(2004), What Do We Do With Missing Data? Some Options For Analysis Of Incomplete Data. *Annu. Rev.Public Health*, 25, 201-226.

- [120] Raghunathan, T. e Grizzle, J.(1995), A Split Questionnaire Survey Design. *Journal of the American Statistical Association*, 90, 54-63.
- [121] Raghunathan, T., Lepkowski, J., Van Hoewyk, J. e Solenberger, P.(2001), A multi-variate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27, 85-95.
- [122] Rao, C.R.(1958), Some statistical methods for comparison of growth curves. *Biometrika*, 51, 83-90.
- [123] Rässler, S., Koller, F. e Mäenpää, C.(2002), A split questionnaire survey design applied to german media and consumer surveys. *Universität Erlanger-Nürnberg-Diskussionpapier*, 42a.
- [124] Raykov, T.(1992), Structural models for studying correlates and predictors of change. *Australian Journal of Psychology*, 44, 101-112.
- [125] Reinecke, J.(2012), Growth curve models and panel dropouts: Applications with criminological panel data. *Netherlands Journal of Psychology*, 67, 122-131.
- [126] Robson, C.(1993), *Real World Research: A Resource for Social Scientists and Practitioner-Researchers*. Blackweel.
- [127] Rogosa, D.R., Brandt, D. e Zimowski, M.(1982), A growth curve approach to the measurement of change. *Psychological Bulletin*, 90, 726-748.
- [128] Rosenbaum, P. e Rubin, D.(1983), The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70, 41-55.
- [129] Roth, P.L.(1994), Missing data: a conceptual review for applied psychologists. *Personnel Psychology*, 43, 537-560.
- [130] Rovine, M.J.(1994), Latent variable models and missing data analysis. Em Von Eye, A. e Clogg, C.(eds), *Latent variables analysis: applications for developmental research*, 181-225.
- [131] Rubin, D.B.(1976), Inference and missing data. *Biometrika*, 63, 581-592.
- [132] Rubin, D.B.(1987), *Multiple imputation of nonresponse in surveys*. Wiley.
- [133] Rubinstein, R.Y.(1981), *Simulation and the Monte Carlo method*. Wiley.
- [134] Salgueiro, M.F.(2010), *Latent Growth Curve Models: A Structural Equation Approach to Modelling Trajectories of Subjective Well-Being*. Provas de Agregação em Estatística e Análise de Dados, Departamento de Métodos Quantitativos, Instituto Universitário de Lisboa (ISCTE-IUL).
- [135] Salgueiro, M.F.(2012), *Modelos com equações estruturais*. Sociedade Portuguesa de Estatística.

-
- [136] Salgueiro, M.F., Smith, P.W.F. e Vieira, M.D.T.(2013), A Multi-Process Second-Order Latent Growth Curve Model for Subjective Well-Being. *Quality and Quantity*, 47(2), 735-752.
- [137] Sayer, A. e Collins, L.M.(eds)(1999), New Methods for the Analysis of Change. *American Psychologist Association*.
- [138] Sewall Wright(1918), On the Nature of Size Factors. *Genetics*, 3, 367-374.
- [139] Sewall Wright(1921), Correlations and Causation. *Journal of Agricultural Research*, 7, 557-585.
- [140] Sewall Wright(1934), The Method of Path Coefficients. *Annals of Mathematical Statistics*, 5, 161-215.
- [141] Sewall Wright(1960), Path Coefficients and Path Regression: Alternative or Complementary Concepts?. *Biometrics*, 16, 189-202.
- [142] Schaffer,J.L.(1997), *Analysis of incomplete multivariate data*. Chapman and Hall.
- [143] Schaffer, J.L. e Graham, J.(2002), Missing Data: Our view of the state of the art. *Psychological Methods*, 7, 147-177.
- [144] Schaffer, J.L. e Olsen, M.K.(1998), Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33, 545-571.
- [145] Singer, J.D. e Willet, J.B.(2003), *Applied Longitudinal data analysis: Modelling change and event occurrence*. Oxford University Press.
- [146] Schoemaker, D.(1973), *Principles and procedures of multiple matrix sampling*. Bellinger Publishing.
- [147] Schumacker, R.E. e Lomax, R.G.(2010), *A beginner's guide to Structural Equation Modeling*, 2nd edition. Lawrence Erlbaum Associates, Inc..
- [148] Skrondal, A. e Rabe-Hesketh, S.(2007), Latent Variable Modelling: A survey. *Scandinavian Journal of Statistics*, 34, 712-745.
- [149] Smith, V.K.(1973), *Monte Carlo methods: Their role for econometrics*. Lexington Books.
- [150] Spearman, C.(1904), General Intelligence Objectively Determined and Measured. *American Journal of Psychology*, 15, 201-293.
- [151] Stoel, R.D., Wittenboer, G.V. e Hox, J.J.(2003), Analyzing longitudinal data using multilevel regression and latent growth curve analysis. *Metodologia de las Ciencias del Comportamiento*, 5, 21-42.
- [152] Taris, T.W.(2000), *A Primer in Longitudinal Data Analysis*. Sage Publications ltd.

-
- [153] Tanner, M.A. e Wong, W.H.(1987), The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528-540.
- [154] Thayer, D.T.(1983), Maximum Likelihood Estimation of the Joint Covariance Matrix for Sections of Tests Given to Distinct Samples with Application to Test Equating. *Psychometrika*, 48, 293-297.
- [155] Thijs, H., Molenberghs, G., Michiels, B. e Curran, D.(2002), Strategies to fit pattern-mixture models. *Biostatistics*, 3, 245-265.
- [156] Thomas, N., Raghunathan, T.E., Schenker, N., Katzoff, M.J. e Johnson, C.L.(2006), An Evaluation of matrix sampling methods using data from the national health and nutrition examination survey. *Survey Methodology*, 32, 217-231.
- [157] Tolvanen, A.(2007), *Latent growth mixture modeling: A simulation study*. Ph.D. Thesis, Department of Mathematics and Statistics, University of Jyväskylä.
- [158] Townsend, P.(1979), *Poverty in the United Kingdom*. Penguin Books.
- [159] Tucker, L.R.(1958), Determination of Parameters of a Functional Relationship by Factor Analysis. *Psychometrika*, 23, 19-23.
- [160] Verbeek,M.(2008), *A guide to modern econometrics*. John Wiley & Sons ltd.
- [161] Verbeek,G., Lesaffre, E. e Spiessens, B.(2001), The practical use of different strategies to handle dropout in longitudinal studies. *Drug Information Journal*, 35, 419-439.
- [162] Vicente, P.C.R. e Salgueiro, M.F.(2013), Cross-lagged structural equation models vs latent growth curve models: a study of material deprivation in Portugal with ICOR. Em Oliveira, P.E, Temido, M.G., Henriques, C. e Vichi, M.(eds.), *Recent Developments in Modeling and Applications in Statistics*, 13-20. Springer.
- [163] Vriens, M., Wedel, M. e Sandor, Z.(2001), Split questionnaire designs: A new tool in survey design and panel management. *Marketing Research*, Summer, 14-19.
- [164] Werts, C.E., Rock, D.A. e Grandy, J.(1979), Confirmatory factor analysis applications: Missing data problems and comparison of path models between populations. *Multivariate Behavioral Research*, 14, 199-213.
- [165] Wiley, D.E.(1973), The identification problem for structural equation models with unmeasured variables. Em A.S Goldergee O.D. Duncan (eds), *Structural equation models in the social sciences*. Academic Press, 69-83.
- [166] Whelan, C.T., Layte, R., Maître, B. e Nolan, B.(2001), Income, Deprivation and Economic Strain. An analysis of the ECHP. *European Sociological Review*, 40, 115-139.
- [167] Whelan, C.T. e Maître, B.(2007), Measuring material deprivation with EU-SILC: Lessons from the Irish survey. *European Societies*, 9, 147-173.

- [168] Willet, J.B. e Sayer, A.G.(1994), Using Covariance Structure Analysis to Detect Correlate and Predictors of individual Change Over Time. *Psychological Bulletin*, 116, 363-381.
- [169] Wishart, J.(1938), Growth-rate determination in nutrition studies with the bacon pig, and their analysis. *Biometrika*, 30(1-2), 16-28.
- [170] Wu, M.C. e Carroll, R.J.(1988), Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process. *Biometrics*, 44, 175-188.
- [171] Yu, C.Y. e Muthén, B.O.(2002), *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes*. Technical report.
- [172] Yuan, K.H. e Bentler, P.M.(2000), Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. Em Von Sobel, M.E. e Berg, M.P.(eds), *Sociological Methodology*, 165-200.

Anexo A

Fórmulas das medidas de qualidade do ajustamento modelo-dados disponíveis no Mplus

O pacote estatístico Mplus calcula os índices TLI (*Tucker-Lewis Index*) e CFI (*Comparative Fit Index*) do seguinte forma

$$TLI = \frac{\chi_{nulo}^2/d_{nulo} - \chi_{H_0}^2/d_{H_0}}{\chi_{nulo}^2/d_{nulo} - 1} \quad (A.1)$$

e

$$CFI = 1 - \frac{\chi_{H_0}^2 - d_{H_0}}{\chi_{nulo}^2 - d_{nulo}}, \quad (A.2)$$

onde d_{nulo} e d_{H_0} correspondem aos graus de liberdade do modelo nulo (modelo que não postula quaisquer relações entre as variáveis) e do modelo em análise (H_0), respetivamente.

Quando é utilizado o método de estimação de Máxima Verosimilhança (ML), o Mplus calcula os critérios de parcimónia AIC (*Akaike Information Criteria*) e BIC (*Bayesian Information Criteria*) do seguinte modo

$$AIC = -2 \log L + 2r \quad (A.3)$$

e

$$BIC = -2 \log L + r \ln N, \quad (A.4)$$

onde r é o número de parâmetros livres do modelo e N a dimensão da amostra.

Sendo $\hat{\theta}$ a estimativa de ML sob a hipótese H_0 , o valor do teste de χ^2 de ajustamento do modelo para H_0 contra H_1 (o modelo em H_1 considera um vetor de médias μ e uma matriz de variâncias-covariâncias Σ não restritos) é obtido como

$$2NF_{ML}(\hat{\theta}), \quad (A.5)$$

sendo F_{ML} a função de ajustamento e N a dimensão da amostra. De salientar que, a utilização desta estatística de χ^2 pressupõe que a amostra é suficientemente grande e que existe normalidade da distribuição dos dados.

Na presença de dados que apresentam um desvio face à distribuição normal pode ser considerado o método de estimação WLS e neste caso o teste de χ^2 de ajustamento do modelo é dado pela estatística

$$2NF_{WLS}(\hat{\boldsymbol{\theta}}), \quad (\text{A.6})$$

sendo necessária uma amostra de grande dimensão .

Assim, caso estas condições (normalidade da distribuição dos dados e amostras de grande dimensão) não se verifiquem deve ser considerado um estimador robusto, disponibilizando o Mplus um teste de qualidade de ajustamento robusto (*Satorra-Bentler scaled chi-square statistic*). Neste teste a estatística de qui-quadrado é ajustada sendo dividida por um fator de correção do seguinte modo

$$G_M = 2NF(\hat{\boldsymbol{\theta}})/c, \quad (\text{A.7})$$

em que $c = \text{tr}[\mathbf{U}\boldsymbol{\Gamma}]$ com $\mathbf{U} = (\mathbf{W}^{-1} - \mathbf{W}^{-1}\boldsymbol{\Delta}(\boldsymbol{\Delta}^T\mathbf{W}^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^T\mathbf{W}^{-1})$, sendo \mathbf{W} uma matriz de pesos, $\boldsymbol{\Gamma}$ uma matriz de variâncias-covariâncias assintóticas para as estatísticas amostrais (estimada por um estimador ADF para estruturas de médias e covariâncias), e $\boldsymbol{\Delta} = \partial\boldsymbol{\sigma}(\boldsymbol{\theta})/\partial(\boldsymbol{\theta})$. Esta estatística é designada por *Satorra-Bentler (mean adjusted chi-square)*.

Também disponível no Mplus o teste de ajustamento com médias e variâncias ajustadas é definida como

$$G_{MV} = 2NF(\hat{\boldsymbol{\theta}})/c_2 \quad (\text{A.8})$$

onde $c_2 = [\text{tr}[\mathbf{U}\boldsymbol{\Gamma}]/d^a]$, em que d^a é calculado como o inteiro mais próximo de d^b , $d^b = (\text{tr}(\mathbf{U}\boldsymbol{\Gamma}))^2/\text{tr}((\mathbf{U}\boldsymbol{\Gamma})^2)$.

A medida de ajustamento absoluto RMSEA (*Root Mean Square Error of Approximation*) é dada por

$$RMSEA = \sqrt{2F_{ML}(\hat{\boldsymbol{\theta}})/d - 1/N}, \quad (\text{A.9})$$

sendo d o número de graus de liberdade do modelo.

A medida de ajustamento WRMR (*Weighted Root Mean Square Residual*) é calculada do seguinte modo

$$WRMR = \sqrt{\sum_{r=1}^e \frac{(s_r - \hat{\sigma}_r)^2}{v_r}}/e, \quad (\text{A.10})$$

sendo s_r um elemento do vetor de estatísticas amostrais, $\hat{\sigma}_r$ é a correspondente estimativa calculada pelo modelo, v_r é uma estimativa da variância assintótica de s_r , e e é o número de estatísticas amostrais.

A medida de ajustamento SRMR (*Standardized Root Mean Square Residual*) é calculada como

$$SRMR = \sqrt{\sum_{j=1}^p \sum_{k=1}^{j-1} r_{jk}^2}/e, \quad (\text{A.11})$$

sendo $e = p(p + 1)/2$, em que p corresponde ao número de variáveis contínuas observadas e r_{jk} é dado por

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}} - \frac{\hat{\sigma}_{jk}}{\sqrt{\hat{\sigma}_{jj}}\sqrt{\hat{\sigma}_{kk}}}, \quad (\text{A.12})$$

onde s_{jk} e $\hat{\sigma}_{jk}$ representam as covariâncias amostrais e estimadas pelo modelo entre variáveis contínuas observadas y_j e y_k , respetivamente.

Anexo B

Comandos de sintaxe em Mplus e LISREL

Modelo de Análise Fatorial Confirmatória Multigrupos (em LISREL)

```
Group: 1
TI analise multigrupos 07-08 (privação 3 dimensões)
DA NI=13 NO=4304 MA=KM NG=2
LA
HX040 DB100 x60 ind107 ind207 ind307 db1001 db1002 ind407 ind1t07 ind2t07 ind3t07
ind4t07
KM FI='joc07.cor' SY
AC FI='joc07.acm'
SE
10 11 12 /
MO NY=3 NE=1 LY=FU,FI TE=DI,FR PS=DI,FR
VA 1 LY(1,1)
PD
OU ME=WLS AD=OFF
Group: 2
DA NI=13 NO=4450 MA=KM
LA
hx04008 DB10008 x6008 ind108 ind208 ind308 db100108 db100208 ind408 ind1t08 ind2t08
ind3t08 ind4t08
KM FI='joc08.cor' SY
AC FI='joc08.acm'
SE
10 11 12 /
MO NY=3 NE=1 LY=IN TE=DI,FR PS=DI,FR
PD
OU ME=WLS AD=OFF
```

Modelo com equações estruturais com efeitos cruzados e desfasados (em Mplus)

TITLE: *multi-process* privação material 2004 a 2007 e var.categoriais

DATA:

File is dadosind_t.dat ;

Variable:

Names are

hb030 DB090_05 DB100_05 hx040_05 DB090_04 DB100_04 hx040_04 DB090_06
DB100_06 hx040_06 DB090_07 DB100_07 hx040_07 ind1_04 ind1_05 ind1_06
ind1_07 ind2_04 ind2_05 ind2_06 ind2_07 ind3_04 ind3_05 ind3_06 ind3_07
ind4_04 ind4_05 ind4_06 ind4_07 indt_04 indt_05 indt_06 indt_07;

Missing are all (-9999);

Usevariables are ind1_04 ind1_05 ind1_06 ind1_07
 ind2_04 ind2_05 ind2_06 ind2_07;

Categorical are ind1_04 ind1_05 ind1_06 ind1_07
 ind2_04 ind2_05 ind2_06 ind2_07;

Model:

ind1_05 ON ind1_04 ind2_04;

ind2_05 ON ind1_04 ind2_04;

ind1_06 ON ind1_05 ind2_05;

ind2_06 ON ind1_05 ind2_05;

ind1_07 ON ind1_06 ind2_06;

ind2_07 ON ind1_06 ind2_06;

ind1_04 WITH ind2_04;

ind1_05 WITH ind2_05;

ind1_06 WITH ind2_06;

ind1_07 WITH ind2_07;

Output: STANDARDIZED;

Modelo com trajetória latente bivariada (em Mplus)

```
TITLE: modelo com trajetoria latente bivariada com indicadores 2004 a 2007
DATA:
File is dadosind_t.dat ;
Variable:
Names are
hb030 DB090_05 DB100_05 hx040_05 DB090_04 DB100_04 hx040_04 DB090_06
DB100_06 hx040_06 DB090_07 DB100_07 hx040_07 ind1_04 ind1_05 ind1_06
ind1_07 ind2_04 ind2_05 ind2_06 ind2_07 ind3_04 ind3_05 ind3_06 ind3_07
ind4_04 ind4_05 ind4_06 ind4_07 indt_04 indt_05 indt_06 indt_07;
Missing are all (-9999) ;
Usevariables are ind1_04 ind1_05 ind1_06 ind1_07
                ind2_04 ind2_05 ind2_06 ind2_07;
Categorical are ind1_04 ind1_05 ind1_06 ind1_07
                ind2_04 ind2_05 ind2_06 ind2_07;
Model:
INT_CE SLOP_CE|ind1_04@0 ind1_05@1 ind1_06@2 ind1_07@3;
INT_B SLOP_B|ind2_04@0 ind2_05@1 ind2_06@2 ind2_07@3;
INT_CE WITH SLOP_CE;
INT_B WITH SLOP_B;
INT_CE WITH INT_B;
SLOP_CE WITH SLOP_B;
Output: STANDARDIZED residual tech4;
```

Modelo com trajetória latente de segunda ordem com indicadores ordinais (em Mplus)

Title: lgcm com multiplos indicadores ordinais

Data:

File is agreg0710.dat;

Variable:

Names are

hy02007 hx04007 hy02008 hx04008 hy02009 hx04009 hy02010 hx04010
ind107 ind207 ind307 ind108 ind208 ind308 ind109 ind209 ind309
ind110 ind210 ind310 db101 db102 loghy2007 loghy2008 loghy2009
loghy2010;

Missing are all (-9999);

Usevariables are ind107 ind207 ind307
ind108 ind208 ind308
ind109 ind209 ind309
ind110 ind210 ind310;

Categorical are ind107 ind207 ind307
ind108 ind208 ind308
ind109 ind209 ind309
ind110 ind210 ind310;

Model:

PM07 BY ind107
ind207-ind307 (1-2);
PM08 BY ind108
ind208-ind308 (1-2);
PM09 BY ind109
ind209-ind309 (1-2);
PM10 BY ind110
ind210-ind310 (1-2);
[ind107\$1 ind108\$1 ind109\$1 ind110\$1] (3);
[ind107\$2 ind108\$2 ind109\$2 ind110\$2] (4);
[ind107\$3 ind108\$3 ind109\$3 ind110\$3] (5);
[ind107\$4 ind108\$4 ind109\$4 ind110\$4] (6);
[ind107\$5 ind108\$5 ind109\$5 ind110\$5] (7);
[ind207\$1 ind208\$1 ind209\$1 ind210\$1] (8);
[ind207\$2 ind208\$2 ind209\$2 ind210\$2] (9);
[ind207\$3 ind208\$3 ind209\$3 ind210\$3] (10);
[ind207\$4 ind208\$4] (11);
[ind307\$1 ind308\$1 ind309\$1 ind310\$1] (12);
[ind307\$2 ind308\$2 ind309\$2 ind310\$2] (13);
[ind307\$3 ind308\$3 ind309\$3 ind310\$3] (14);
{ind107-ind307@1 ind108-ind310};
IntPM SloPM | PM07@0 PM08@1 PM09@2 PM10@3;

ind107 WITH ind108;
ind107 WITH ind109;
ind107 WITH ind110;
ind108 WITH ind109;
ind108 WITH ind110;
ind109 WITH ind110;
ind207 WITH ind208;
ind207 WITH ind209;
ind207 WITH ind210;
ind208 WITH ind209;
ind208 WITH ind210;
ind209 WITH ind210;
ind307 WITH ind308;
ind307 WITH ind309;
ind307 WITH ind310;
ind308 WITH ind309;
ind308 WITH ind310;
ind309 WITH ind310;
OUTPUT:standardized tech4

Modelo com trajetória latente de segunda ordem com indicadores ordinais e todas as variáveis explicativas invariantes (em Mplus)

Title: lgcm com multiplos indicadores ordinais com explicativas invariantes no tempo

Data:

File is agreg_0710.dat ;

Variable:

Names are

hy020_07 hx040_07 hy020_08 hx040_08 hy020_09 hx040_09 hy020_10 hx040_10
ind1_07 ind2_07 ind3_07 ind1_08 ind2_08 ind3_08 ind1_09 ind2_09 ind3_09
ind1_10 ind2_10 ind3_10 db101 db102 log_hy20_07 log_hy20_08 log_hy20_09
log_hy20_10 dhx040_07 dhx040_08 dhx040_09 dhx040_10 d1hx040_07 d1hx040_08
d1hx040_09 d1hx040_10;

Missing are all (-9999);

Usevariables are ind1_07 ind2_07 ind3_07

ind1_08 ind2_08 ind3_08

ind1_09 ind2_09 ind3_09

ind1_10 ind2_10 ind3_10

log_hy20_07 dhx040_07 db101 db102;

Categorical are ind1_07 ind2_07 ind3_07

ind1_08 ind2_08 ind3_08

ind1_09 ind2_09 ind3_09

ind1_10 ind2_10 ind3_10;

Model: PM_07 BY ind1_07

ind2_07-ind3_07 (1-2);

PM_08 BY ind1_08

ind2_08-ind3_08 (1-2);

PM_09 BY ind1_09

ind2_09-ind3_09 (1-2);

PM_10 BY ind1_10

ind2_10-ind3_10 (1-2);

[ind1_07\$1 ind1_08\$1 ind1_09\$1 ind1_10\$1] (3);

[ind1_07\$2 ind1_08\$2 ind1_09\$2 ind1_10\$2] (4);

[ind1_07\$3 ind1_08\$3 ind1_09\$3 ind1_10\$3] (5);

[ind1_07\$4 ind1_08\$4 ind1_09\$4 ind1_10\$4] (6);

[ind1_07\$5 ind1_08\$5 ind1_09\$5 ind1_10\$5] (7);

[ind2_07\$1 ind2_08\$1 ind2_09\$1 ind2_10\$1] (8);

[ind2_07\$2 ind2_08\$2 ind2_09\$2 ind2_10\$2] (9);

[ind2_07\$3 ind2_08\$3 ind2_09\$3 ind2_10\$3] (10);

[ind2_07\$4 ind2_08\$4] (11);

[ind3_07\$1 ind3_08\$1 ind3_09\$1 ind3_10\$1] (12);

[ind3_07\$2 ind3_08\$2 ind3_09\$2 ind3_10\$2] (13);

[ind3_07\$3 ind3_08\$3 ind3_09\$3 ind3_10\$3] (14);

{ind1_07-ind3_07@1 ind1_08-ind3_10};

```
Int_PM Slo_PM | PM_07@0 PM_08@1 PM_09@2 PM_10@3 ;
ind1_07 WITH ind1_08;
ind1_07 WITH ind1_09;
ind1_07 WITH ind1_10;
ind1_08 WITH ind1_09;
ind1_08 WITH ind1_10;
ind1_09 WITH ind1_10;
ind2_07 WITH ind2_08;
ind2_07 WITH ind2_09;
ind2_07 WITH ind2_10;
ind2_08 WITH ind2_09;
ind2_08 WITH ind2_10;
ind2_09 WITH ind2_10;
ind3_07 WITH ind3_08;
ind3_07 WITH ind3_09;
ind3_07 WITH ind3_10;
ind3_08 WITH ind3_09;
ind3_08 WITH ind3_10;
ind3_09 WITH ind3_10;
Int_PM Slo_PM ON log_hy20_07 dhx040_07 db101 db102;
OUTPUT:standardized tech4
```

Modelo com trajetória latente de segunda ordem com indicadores ordinais e variável explicativa rendimento variante e variáveis explicativas dimensão do agregado e área de urbanização invariantes (em Mplus)

Title: lgcm com multiplos indicadores ordinais com explicativas variante e invariantes no tempo

Data:

File is agreg_0710.dat ;

Variable:

Names are

hy020_07 hx040_07 hy020_08 hx040_08 hy020_09 hx040_09 hy020_10 hx040_10
ind1_07 ind2_07 ind3_07 ind1_08 ind2_08 ind3_08 ind1_09 ind2_09 ind3_09
ind1_10 ind2_10 ind3_10 db101 db102 log_hy20_07 log_hy20_08 log_hy20_09
log_hy20_10 dhx040_07 dhx040_08 dhx040_09 dhx040_10 d1hx040_07 d1hx040_08
d1hx040_09 d1hx040_10;

Missing are all (-9999);

Usevariables are ind1_07 ind2_07 ind3_07

ind1_08 ind2_08 ind3_08

ind1_09 ind2_09 ind3_09

ind1_10 ind2_10 ind3_10

db101 db102 d1hx040_07

log_hy20_07 log_hy20_08 log_hy20_09 log_hy20_10;

Categorical are ind1_07 ind2_07 ind3_07

ind1_08 ind2_08 ind3_08

ind1_09 ind2_09 ind3_09

ind1_10 ind2_10 ind3_10;

Model:

PM_07 BY ind1_07

ind2_07-ind3_07 (1-2);

PM_08 BY ind1_08

ind2_08-ind3_08 (1-2);

PM_09 BY ind1_09

ind2_09-ind3_09 (1-2);

PM_10 BY ind1_10

ind2_10-ind3_10 (1-2);

[ind1_07\$1 ind1_08\$1 ind1_09\$1 ind1_10\$1] (3);

[ind1_07\$2 ind1_08\$2 ind1_09\$2 ind1_10\$2] (4);

[ind1_07\$3 ind1_08\$3 ind1_09\$3 ind1_10\$3] (5);

[ind1_07\$4 ind1_08\$4 ind1_09\$4 ind1_10\$4] (6);

[ind1_07\$5 ind1_08\$5 ind1_09\$5 ind1_10\$5] (7);

[ind2_07\$1 ind2_08\$1 ind2_09\$1 ind2_10\$1] (8);

[ind2_07\$2 ind2_08\$2 ind2_09\$2 ind2_10\$2] (9);

[ind2_07\$3 ind2_08\$3 ind2_09\$3 ind2_10\$3] (10);

[ind2_07\$4 ind2_08\$4] (11);

```
[ind3_07$1 ind3_08$1 ind3_09$1 ind3_10$1] (12);
[ind3_07$2 ind3_08$2 ind3_09$2 ind3_10$2] (13);
[ind3_07$3 ind3_08$3 ind3_09$3 ind3_10$3] (14);
{ind1_07-ind3_07@1 ind1_08-ind3_10};
Int_PM Slo_PM | PM_07@0 PM_08@1 PM_09@2 PM_10@3 ;
ind1_07 WITH ind1_08;
ind1_07 WITH ind1_09;
ind1_07 WITH ind1_10;
ind1_08 WITH ind1_09;
ind1_08 WITH ind1_10;
ind1_09 WITH ind1_10;
ind2_07 WITH ind2_08;
ind2_07 WITH ind2_09;
ind2_07 WITH ind2_10;
ind2_08 WITH ind2_09;
ind2_08 WITH ind2_10;
ind2_09 WITH ind2_10;
ind3_07 WITH ind3_08;
ind3_07 WITH ind3_09;
ind3_07 WITH ind3_10;
ind3_08 WITH ind3_09;
ind3_08 WITH ind3_10;
ind3_09 WITH ind3_10;
PM_07 ON log_hy20_07;
PM_08 ON log_hy20_08;
PM_09 ON log_hy20_09;
PM_10 ON log_hy20_10;
Int_PM Slo_PM ON d1hx040_07 db101 db102;
OUTPUT:standardized tech4
```

Modelo com trajetória latente e abordagem FIML para lidar com as observações omissas (em Mplus)

Title: latent growth entre 2006 e 2009 e FIML

Data:

File is jocl_12.dat ;

Variable:

Names are

hx040_06 db060_06 db100_06 hx040_07 db060_07 db100_07 hx040_08 db060_08

db100_08 ind1_05 ind21_05 ind3_05 ind1_06 ind21_06 ind3_06 ind1_07

ind21_07 ind3_07 ind1_08 ind21_08 ind3_08 ind1_09 ind21_09 ind3_09

indt_05 indt_06 indt_07 indt_08 indt_09;

Usevariables are indt_06 indt_07 indt_08 indt_09;

Missing are all (-9999);

model: i s | indt_06@0 indt_07@1 indt_08@2 indt_09@3;

output: tech1 standardized;

Modelo com trajetória latente para dados completos (em Mplus)

Title: latent growth entre 2006 e 2009

Data:

File is jocl_12.dat ;

listwise = on;

Variable:

Names are

hx040_06 db060_06 db100_06 hx040_07 db060_07 db100_07 hx040_08 db060_08
db100_08 ind1_05 ind21_05 ind3_05 ind1_06 ind21_06 ind3_06 ind1_07
ind21_07 ind3_07 ind1_08 ind21_08 ind3_08 ind1_09 ind21_09 ind3_09
indt_05 indt_06 indt_07 indt_08 indt_09;

Usevariables are indt_06 indt_07 indt_08 indt_09;

Missing are all (-9999);

model: i s | indt_06@0 indt_07@1 indt_08@2 indt_09@3;

output: tech1;

Modelo com trajetória latente e os valores estimados dos parâmetros são guardados para um estudo de simulação de Monte Carlo (em Mplus)

Title: Modelo com trajetória latente e os valores estimados dos parâmetros são guardados para um estudo de simulação de Monte Carlo

Data:

File is joc_12_mc.dat ;

Variable:

Names are

```

    hx040_06 db060_06 db100_06 hx040_07 db060_07 db100_07 hx040_08 db060_08
    db100_08 ind1_06 ind21_06 ind3_06 ind1_07 ind21_07 ind3_07 ind1_08
    ind21_08 ind3_08 ind1_09 ind21_09 ind3_09 indt_06 indt_07 indt_08
    indt_09;

```

Usevariables are indt_06 indt_07 indt_08 indt_09;

Missing are all (-9999);

```

    model: i s | indt_06@0 indt_07@1 indt_08@2 indt_09@3;
    [i@2.500 s@-0.106];
    s*-0.106; i with s*-0.415;
    output: tech1;
    savedata: estimates = joc12_m_estimates.dat;

```

Estudo de simulação de Monte Carlo para um modelo com trajetória latente com dados reais (em Mplus)

Title: Estudo de simulação de Monte Carlo para um modelo com trajetória latente com dados reais

Montecarlo: Names = indt_06 indt_07 indt_08 indt_09;

Nobservations =845;

Nreps = 10000;

Seed = 45335;

Population = joc12_m_estimates.dat;

coverage = joc12_m_estimates.dat;

Patmiss = indt_06(0) indt_07(0) indt_08(0) indt_09(.5);

Patprobs = 1;

Model population:

```

    i s | indt_06@0 indt_07@1 indt_08@2 indt_09@3;
    Model: i s | indt_06@0 indt_07@1 indt_08@2 indt_09@3;
    Output: tech9;

```

Modelo com trajetória latente e imputação múltipla para lidar com dados omissos - 1ºPasso (em Mplus)

```
Title: LGCM com dados omissos + imputação múltipla
Data:
File is ecm2012_1.dat ;
Variable:
Names are
  DB100_06 hx040_06 DB100_07 hx040_07 DB100_08 hx040_08 DB100_09 hx040_09
  indt_06 indt_07 indt_08 indt_09;
Usevariables are indt_06 indt_07 indt_08 indt_09;
Missing are all (-9999);
Analysis: type = BASIC;
bseed = 59765;
bconvergence = 0.05;
Data imputation:
Impute = indt_06 indt_07 indt_08 indt_09;
Ndatasets = 5;
Save = ecm12l1imp*.dat;
thin = 200;
Output: TECH8;
```

Modelo com trajetória latente e imputação múltipla para lidar com dados omissos - 2ºPasso (em Mplus)

```
Title: LGCM usando imputação múltipla;
Data:
File is ecm12l1implist.dat;
type = imputation;
Variable:
Names are
  indt_06 indt_07 indt_08 indt_09;
Missing are all (-9999);
Usevariables are indt_06 indt_07 indt_08 indt_09;
Analysis: Estimator = MLR;
Model: i s | indt_06@0 indt_07@1 indt_08@2 indt_09@3;
Output: TECH1 TECH4;
```

Modelo com trajetória latente e *pattern mixture model* para lidar com dados omissos - abordagem convencional(4 padrões de omissão) (em Mplus)

```
Title: pattern-mixture model com indicadores de padrão de omissão;  
Data:  
File is conf13.dat;  
Variable:  
Names are  
DB100_06 hx040_06 DB100_07 hx040_07 DB100_08 hx040_08 DB100_09 hx040_09  
indt_06 indt_07 indt_08 indt_09 r1 r2 r3;  
Missing are all (-9999);  
Usevariables are indt_06 indt_07 indt_08 indt_09 r1 r2 r3  
model:  
i s | indt_06@0 indt_07@1 indt_08@2 indt_09@3;  
i on r1 r2 r3;  
s on r1 r2 r3;  
output: tech1 sampstat;
```

Modelo com trajetória latente e *pattern mixture model* para lidar com dados omissos - abordagem multigrupos (em Mplus)

Title: *pattern mixture model* com 4 grupos (completo,entradas tardias, saidas antes do tempo, outros);

Data:

File is spe12.dat;

Variable:

Names are

```
DB100_05 hx040_05 DB100_06 hx040_06 DB100_07 hx040_07 DB100_08 hx040_08
DB100_09 hx040_09 indt_05 indt_06 indt_07 indt_08 indt_09 filter_;
```

Usevariables are indt_06 indt_07 indt_08 indt_09 mdpatt;

Missing are all (-9999);

classes = pattern(4);

knownclass = pattern(mdpatt = 1 mdpatt = 2 mdpatt = 3 mdpatt = 4);

define:

!completo

```
if (indt_09 ne _missing and indt_06 ne _missing and
indt_07 ne _missing and indt_08 ne _missing) then mdpatt=1;
```

!intermitentes

```
if (indt_09 eq _missing and indt_06 ne _missing and
indt_07 eq _missing and indt_08 ne _missing) then mdpatt=4;
if (indt_09 ne _missing and indt_06 ne _missing and
indt_07 ne _missing and indt_08 eq _missing) then mdpatt=4;
if (indt_09 ne _missing and indt_06 ne _missing and
indt_07 eq _missing and indt_08 ne _missing) then mdpatt=4;
if (indt_09 ne _missing and indt_06 eq _missing and
indt_07 ne _missing and indt_08 eq _missing) then mdpatt=4;
```

!entradas tardias e saidas antes

```
if (indt_09 eq _missing and indt_06 eq _missing and
indt_07 ne _missing and indt_08 ne _missing) then mdpatt=4;
if (indt_09 eq _missing and indt_06 eq _missing and
indt_07 ne _missing and indt_08 eq _missing) then mdpatt=4;
if (indt_09 eq _missing and indt_06 eq _missing and
indt_07 eq _missing and indt_08 ne _missing) then mdpatt=4;
```

!entradas tardias

```
if (indt_09 ne _missing and indt_06 eq _missing and
indt_07 ne _missing and indt_08 ne _missing) then mdpatt=2;
if (indt_09 ne _missing and indt_06 eq _missing and
indt_07 eq _missing and indt_08 ne _missing) then mdpatt=2;
if (indt_09 ne _missing and indt_06 eq _missing and
indt_07 eq _missing and indt_08 eq _missing) then mdpatt=2;
```

!saidas antes tempo

```
if (indt_09 eq _missing and indt_06 ne _missing and
```

```
    indt_07 ne _missing and indt_08 eq _missing) then mdpatt=3;
    if (indt_09 eq _missing and indt_06 ne _missing and
    indt_07 ne _missing and indt_08 ne _missing) then mdpatt=3;
    if (indt_09 eq _missing and indt_06 ne _missing and
    indt_07 eq _missing and indt_08 eq _missing) then mdpatt=3;
analysis:
    type=mixture;
model:
    i s | indt_06@0 indt_07@1 indt_08@2 indt_09@3;
    [pattern#1] (p1logit);
    [pattern#2] (p2logit);
    [pattern#3] (p3logit);
    [i] (p1i);
    [s] (p1s);
    [i] (p2i);
    [s] (p2s);
    [i] (p3i);
    [s] (p3s);
    [i] (p4i);
    [s] (p4s);
model constraint:
new(c1prop c2prop c3prop c4prop iavg savg);
c1prop = exp(p1logit)/(exp(0)exp(p1logit)+exp(p2logit)+exp(p3logit));+
c2prop = exp(p2logit)/(exp(0)exp(p1logit)+exp(p2logit)+exp(p3logit));+
c3prop = exp(p3logit)/(exp(0)exp(p1logit)+exp(p2logit)+exp(p3logit));+
c4prop = exp(0)/(exp(0)exp(p1logit)+exp(p2logit)+exp(p3logit));+
iavg = c1prop*p1ic2prop*p2i+c3prop*p3i+c4prop*p4i;+
savg = c1prop*p1sc2prop*p2s+c3prop*p3s+c4prop*p4s;+
Plot:
    type is plot3;
    series is indt_06-indt_09 (s);
output: tech1 sampstat;
```

Modelo com trajetória latente para gerar dados com distribuição normal e posterior estimação (em Mplus)

Title: gerar e analisar dados com distribuição normal;

Montecarlo:

Names are y1 y2 y3 y4 y5;

Nobservations = 500;

Nreps = 1000;

Seed = 53487;

Classes = c(1);

Genclasses = c(1);

patmiss =

y1(0) y2(0) y3(0) y4(0) y5(1)|

y1(1) y2(0) y3(0) y4(0) y5(0)|

y1(1) y2(1) y3(0) y4(0) y5(0)|

y1(1) y2(1) y3(1) y4(0) y5(0);

patprobs = .25|.25|.25|.25;

analysis:

type = mixture;

estimator = ml;

model montecarlo:

 i by y1-y5@1;

 s by y1@0 y2@1 y3@2 y4@3 y5@4;

[y1-y5@0];

[i*0 s*0];

i*1;

s*4;

i with s*0;

y1*1;

y2*5;

y3*17;

y4*37;

y5*65;

[i*0 s*0];

model:

 i by y1-y5@1;

 s by y1@0 y2@1 y3@2 y4@3 y5@4;

[y1-y5@0];

[i*0 s*0];

i*1

 s*4;

 i with s*0;

 y1*1;

```
y2*5;  
y3*17;  
y4*37;  
y5*65;  
[i*0 s*0];  
output: tech9;
```

Modelo de mistura para gerar dados com distribuição não normal e estimação de um modelo com trajetória latente aos dados gerados (em Mplus)

Title: gerar e analisar dados com distribuição não normal;

Montecarlo:

```
Names are y1 y2 y3 y4;
Nobservations = 500;
Nreps = 1000;
Seed = 53487;
Classes = c(1);
Genclasses = c(2);
```

analysis:

```
type = mixture;
```

```
estimator = ml;
```

model montecarlo:

```
  i by y1-y4@1;
  s by y1@0 y2@1 y3@2 y4@3;
[y1-y4@0];
[c#1@-2];
[i*15 s*4];
i*35; s*.2; i with s*0;
y1*1;
y2*1.2;
y3*1.8;
y4*2.8;
[i*0 s*0];
i*1; s*.2; i with s*0;
y1*1;
y2*1.2;
y3*1.8;
y4*2.8;
```

model:

```
  i by y1-y4@1;
  s by y1@0 y2@1 y3@2 y4@3;
[y1-y4@0];
[i s];
i*29.312; s*1.915; i with s*6.4124;
y1*1.0017;
y2*1.1830;
y3*1.8058;
y4*2.7973;
[i*1.8226 s*.4869];
```

```
output: tech9;
```


Anexo C

Diagrama e tabela de Análise Fatorial Confirmatória (secção 3.4.4)

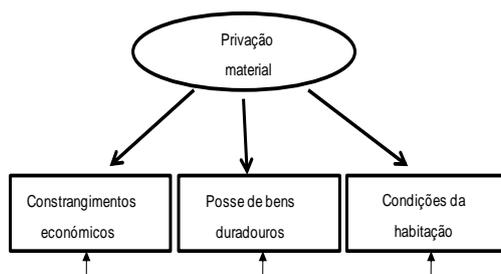


Figura C.1: Diagrama de um modelo de análise fatorial confirmatória com a privação material medida por 3 indicadores (CE, PBD e CH)

	2007	2008	2009	2010
Constrangimentos Económicos	0.765	0.764	0.777	0.844
Posse de bens	0.758	0.684	0.712	0.675
Condições da habitação	0.502	0.501	0.533	0.355

Tabela C.1: Estimativas, numa solução estandardizada, dos pesos fatoriais, para os quatro modelos de AFC

Anexo D

Tabelas do estudo de simulação: Dados com o desenho das subamostras 2 a 7 (secção 4.3.4)

Parâmetros	“população”	média	desvio padrão	média erros padrão	proporção de intervalos de confiança a 95%	proporção de réplicas com coeficiente significativo
μ_α	2.355	2.3555	0.0604	0.0610	0.951	1.000
μ_β	-0.008	-0.0081	0.0168	0.0172	0.953	0.072
$\psi_{\alpha\alpha}$	2.441	2.4381	0.1556	0.1562	0.949	1.000
$\psi_{\beta\beta}$	0.089	0.0885	0.0162	0.0160	0.948	1.000
$\psi_{\alpha\beta}$	-0.196	-0.1951	0.0378	0.0378	0.950	0.999

Tabela D.1: Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 2

Parâmetros	“população”	média	desvio padrão	média erros padrão	proporção de intervalos de confiança a 95%	proporção de réplicas com coeficiente significativo
μ_α	2.355	2.3552	0.0570	0.0573	0.949	1.000
μ_β	-0.008	-0.0080	0.0187	0.0187	0.952	0.072
$\psi_{\alpha\alpha}$	2.441	2.4382	0.1404	0.1402	0.947	1.000
$\psi_{\beta\beta}$	0.089	0.0884	0.0171	0.0170	0.946	1.000
$\psi_{\alpha\beta}$	-0.196	-0.1953	0.0399	0.0396	0.951	0.999

Tabela D.2: Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 3

Parâmetros	“população”	média	desvio padrão	média erros padrão	proporção de intervalos de confiança a 95%	proporção de réplicas com coeficiente significativo
μ_α	2.355	2.3556	0.0684	0.0692	0.948	1.000
μ_β	-0.008	-0.0082	0.0204	0.0206	0.950	0.066
$\psi_{\alpha\alpha}$	2.441	2.4370	0.1877	0.1866	0.951	1.000
$\psi_{\beta\beta}$	0.089	0.0881	0.0187	0.0186	0.944	0.999
$\psi_{\alpha\beta}$	-0.196	-0.1945	0.0448	0.0448	0.948	0.996

Tabela D.3: Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 4

Parâmetros	“população”	média	desvio padrão	média erros padrão	proporção de intervalos de confiança a 95%	proporção de réplicas com coeficiente significativo
μ_α	2.355	2.3552	0.0572	0.0574	0.948	1.000
μ_β	-0.008	-0.0079	0.0198	0.0194	0.944	0.073
$\psi_{\alpha\alpha}$	2.441	2.4386	0.1416	0.1415	0.948	1.000
$\psi_{\beta\beta}$	0.089	0.0888	0.0196	0.0194	0.947	0.997
$\psi_{\alpha\beta}$	-0.196	-0.1955	0.0428	0.0423	0.946	0.998

Tabela D.4: Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 5

Parâmetros	“população”	média	desvio padrão	média erros padrão	proporção de intervalos de confiança a 95%	proporção de réplicas com coeficiente significativo
μ_α	2.355	2.3550	0.0682	0.0682	0.949	1.000
μ_β	-0.008	-0.0080	0.0205	0.0205	0.951	0.070
$\psi_{\alpha\alpha}$	2.441	2.4368	0.1860	0.1858	0.946	1.000
$\psi_{\beta\beta}$	0.089	0.0884	0.0202	0.0202	0.948	0.995
$\psi_{\alpha\beta}$	-0.196	-0.1948	0.0469	0.0469	0.946	0.993

Tabela D.5: Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 6

Parâmetros	“população”	média	desvio padrão	média erros padrão	proporção de intervalos de confiança a 95%	proporção de réplicas com coeficiente significativo
μ_α	2.355	2.3552	0.0600	0.0598	0.948	1.000
μ_β	-0.008	-0.0079	0.0180	0.0179	0.950	0.077
$\psi_{\alpha\alpha}$	2.441	2.4385	0.1508	0.1507	0.949	1.000
$\psi_{\beta\beta}$	0.089	0.0887	0.0176	0.0175	0.947	0.999
$\psi_{\alpha\beta}$	-0.196	-0.1955	0.0394	0.0391	0.949	0.999

Tabela D.6: Resultados obtidos no estudo de simulação gerando amostras com o desenho da subamostra 7

Anexo E

Diagrama de um procedimento *Bootstrap*

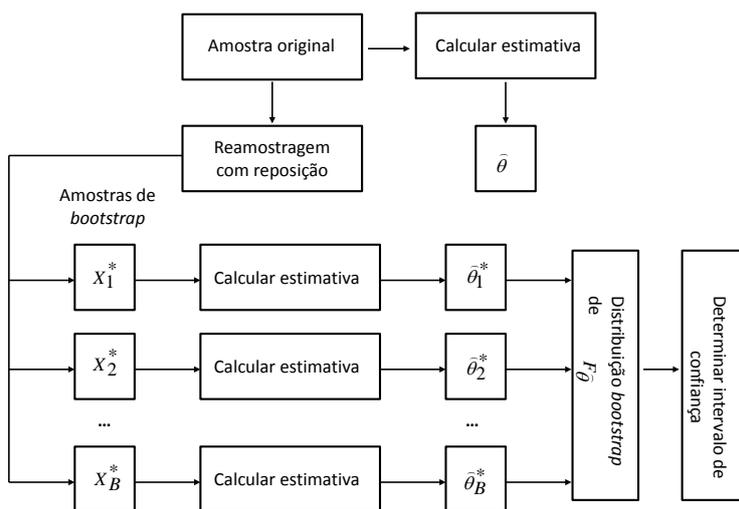


Figura E.1: Diagrama de um procedimento de *Bootstrap*

Anexo F

Tabelas do estudo de simulação: Dados com distribuição normal e $T=3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.0087 (-0.0218)	0.0052 (-0.0130)	0.0013 (-0.0033)	0.0009 (-0.0023)
	4	0.0331 (-0.0184)	0.0121 (-0.0067)	-0.0006 (0.0003)	0.0023 (-0.0013)
μ_α	0.2	-0.0014	-0.0029	-0.0003	-0.0022
	4	-0.0035	-0.0034	-0.0002	-0.0017
μ_β	0.2	0.0028	0.0021	0.0006	0.0012
	4	0.0155	0.0064	0.0008	0.0008
$\psi_{\alpha\alpha}$	0.2	-0.0222 (-0.0222)	-0.0113 (-0.0113)	-0.0030 (-0.0030)	-0.0020 (-0.0020)
	4	-0.0104 (-0.0104)	-0.0143 (-0.0143)	-0.0019 (-0.0019)	-0.0026 (-0.0026)
$\psi_{\beta\beta}$	0.2	-0.0033 (-0.0165)	-0.0025 (-0.0125)	-0.0006 (-0.0030)	-0.0007 (-0.0035)
	4	-0.0821 (-0.0205)	-0.0143 (-0.0036)	0.0027 (0.0007)	-0.0052 (-0.0013)

Tabela F.1: Enviesamento bruto(enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = -0.9 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.0040 (-0.0182)	0.0054 (-0.0245)	0.0012 (-0.0055)	0.0010 (-0.0045)
	4	0.0116 (-0.0116)	0.0093 (-0.0093)	-0.0011 (0.0011)	0.0017 (-0.0017)
μ_α	0.2	-0.0024	-0.0033	-0.0003	-0.0021
	4	-0.0032	-0.0033	-0.0001	-0.0017
μ_β	0.2	0.0057	0.0028	0.0009	0.0010
	4	0.0193	0.0062	0.0013	-0.0002
$\psi_{\alpha\alpha}$	0.2	-0.0197 (-0.0197)	-0.0132 (-0.0132)	-0.0034 (-0.0034)	-0.0024 (-0.0024)
	4	-0.0016 (-0.0016)	-0.0176 (-0.0176)	-0.0029 (-0.0029)	-0.0031 (-0.0031)
$\psi_{\beta\beta}$	0.2	-0.0021 (-0.0105)	-0.0030 (-0.0150)	-0.0008 (-0.0040)	-0.0012 (-0.0060)
	4	-0.0709 (-0.0177)	-0.0097 (-0.0024)	0.0013 (0.0003)	-0.0074 (-0.0019)

Tabela F.2: Enviesamento bruto(enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = -0.5 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	-0.0022	0.0053	0.0012	0.0010
	4	-0.0150	0.0064	0.0011	0.0012
μ_α	0.2	-0.0026	-0.0033	-0.0003	-0.0020
	4	-0.0030	-0.0033	-0.0001	-0.0016
μ_β	0.2	0.0073	0.0029	0.0011	0.0007
	4	0.0212	0.0056	0.0018	-0.0009
$\psi_{\alpha\alpha}$	0.2	-0.0165 (-0.0165)	-0.0148 (-0.0148)	-0.0038 (-0.0038)	-0.0026 (-0.0026)
	4	0.0077 (0.0077)	-0.0208 (-0.0208)	-0.0039 (-0.0039)	-0.0036 (-0.0036)
$\psi_{\beta\beta}$	0.2	-0.0002 (-0.0010)	-0.0037 (-0.0185)	-0.0013 (-0.0065)	-0.0017 (-0.0085)
	4	-0.0578 (-0.0145)	-0.0134 (0.0034)	-0.0032 (-0.0008)	-0.0103 (-0.0026)

Tabela F.3: Enviesamento bruto(enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = 0 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	-0.0086 (-0.0391)	0.0051 (0.0232)	0.0012 (0.0055)	0.0009 (0.0041)
	4	-0.0415 (-0.0415)	0.0042 (0.0042)	-0.0006 (-0.0006)	0.0008 (0.0008)
μ_α	0.2	-0.0027	-0.0033	-0.0002	-0.0019
	4	-0.0029	-0.0033	-0.0001	-0.0016
μ_β	0.2	0.0083	0.0029	0.0012	0.0005
	4	0.0217	0.0046	0.0022	-0.0014
$\psi_{\alpha\alpha}$	0.2	-0.0136 (-0.0136)	-0.0162 (-0.0162)	-0.0042 (-0.0042)	-0.0029 (-0.0029)
	4	0.0147 (0.0147)	-0.0239 (-0.0239)	-0.0050 (-0.0050)	-0.0041 (-0.0041)
$\psi_{\beta\beta}$	0.2	0.0015 (0.0075)	-0.0046 (-0.0230)	-0.0020 (-0.0100)	-0.0023 (-0.0115)
	4	-0.0512 (-0.0128)	-0.0214 (-0.0054)	-0.0091 (-0.0023)	-0.0130 (-0.0033)

Tabela F.4: Enviesamento bruto(enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = 0.5 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	-0.0137 (-0.0343)	0.0048 (0.0120)	0.0013 (0.0033)	0.0009 (0.0023)
	4	-0.0627 (-0.0348)	0.0028 (0.0016)	0.0001 (0)	0.0005 (0.0003)
μ_α	0.2	-0.0027	-0.0033	-0.0002	-0.0019
	4	-0.0028	-0.0033	0	-0.0016
μ_β	0.2	0.0087	0.0028	0.0013	0.0003
	4	0.0201	0.0036	0.0027	-0.0017
$\psi_{\alpha\alpha}$	0.2	-0.0114 (-0.0114)	-0.0171 (-0.0171)	-0.0045 (-0.0045)	-0.0030 (-0.0030)
	4	0.0201 (0.0201)	0.0263 (0.0263)	-0.0060 (-0.0060)	0.0045 (0.0045)
$\psi_{\beta\beta}$	0.2	0.0027 (0.0135)	-0.0056 (-0.0280)	-0.0027 (-0.0135)	-0.0027 (-0.0135)
	4	-0.0529 (-0.0132)	-0.0300 (-0.0075)	-0.0140 (-0.0035)	-0.0149 (-0.0037)

Tabela F.5: Enviesamento bruto(enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = 0.9 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.2196	0.0426	0.0227	0.0112
	4	1.4975	0.2894	0.1551	0.0737
μ_{α}	0.2	0.0389	0.0078	0.0038	0.0018
	4	0.0396	0.0080	0.0039	0.0018
μ_{β}	0.2	0.0253	0.0047	0.0024	0.0011
	4	0.1718	0.0311	0.0162	0.0076
$\psi_{\alpha\alpha}$	0.2	0.5109	0.0987	0.0505	0.0269
	4	2.3732	0.4556	0.2373	0.1237
$\psi_{\beta\beta}$	0.2	0.2237	0.0461	0.0222	0.0110
	4	4.0596	0.8631	0.4005	0.2098

Tabela F.6: Erro quadrático médio dos modelos A1 e A4, com correlação(α, β) = 0.9 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.1764	0.0343	0.0183	0.0090
	4	1.2786	0.2477	0.1332	0.0630
μ_{α}	0.2	0.0386	0.0077	0.0038	0.0017
	4	0.0396	0.0080	0.0039	0.0018
μ_{β}	0.2	0.0225	0.0042	0.0021	0.0010
	4	0.1627	0.0295	0.0152	0.0071
$\psi_{\alpha\alpha}$	0.2	0.4263	0.0824	0.0422	0.0224
	4	2.0025	0.3839	0.2010	0.1038
$\psi_{\beta\beta}$	0.2	0.1679	0.0345	0.0167	0.0082
	4	3.2729	0.7020	0.3244	0.1661

Tabela F.7: Erro quadrático médio dos modelos A1 e A4, com correlação(α, β) = 0.5 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.1249	0.0243	0.0130	0.0064
	4	1.0347	0.2005	0.1082	0.0513
μ_{α}	0.2	0.0381	0.0075	0.0037	0.0017
	4	0.0395	0.0080	0.0039	0.0018
μ_{β}	0.2	0.0187	0.0036	0.0018	0.0008
	4	0.1488	0.0275	0.0138	0.0065
$\psi_{\alpha\alpha}$	0.2	0.3225	0.0624	0.0319	0.0170
	4	1.5346	0.2939	0.1545	0.0791
$\psi_{\beta\beta}$	0.2	0.1066	0.0218	0.0106	0.0052
	4	2.4386	0.5259	0.2443	0.1219

Tabela F.8: Erro quadrático médio dos modelos A1 e A4, com correlação(α, β) = 0 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.749	0.0146	0.0077	0.0039
	4	0.8262	0.1594	0.0861	0.0413
μ_{α}	0.2	0.0369	0.0073	0.0036	0.0017
	4	0.0393	0.0079	0.0039	0.0018
μ_{β}	0.2	0.0144	0.0028	0.0014	0.0006
	4	0.1317	0.0253	0.0123	0.0057
$\psi_{\alpha\alpha}$	0.2	0.2184	0.0423	0.0215	0.0114
	4	1.0627	0.2034	0.1076	0.0544
$\psi_{\beta\beta}$	0.2	0.0532	0.0108	0.0053	0.0026
	4	1.7991	0.3803	0.1817	0.0906

Tabela F.9: Erro quadrático médio dos modelos A1 e A4, com correlação(α, β) = -0.5 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.0351	0.0068	0.0035	0.0018
	4	0.6868	0.1313	0.0706	0.0344
μ_{α}	0.2	0.0341	0.0066	0.0033	0.0015
	4	0.0389	0.0078	0.0038	0.0018
μ_{β}	0.2	0.0093	0.0018	0.0009	0.0004
	4	0.1152	0.0230	0.0110	0.0051
$\psi_{\alpha\alpha}$	0.2	0.1392	0.0258	0.0130	0.0069
	4	0.6826	0.1305	0.0695	0.0346
$\psi_{\beta\beta}$	0.2	0.0152	0.0030	0.0015	0.0008
	4	1.4639	0.2913	0.1466	0.0740

Tabela F.10: Erro quadrático médio dos modelos A1 e A4, com correlação(α, β) = -0.9 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.953	0.958	0.951	0.954
	4	0.949	0.962	0.948	0.956
μ_{α}	0.2	0.947	0.953	0.954	0.961
	4	0.946	0.953	0.959	0.959
μ_{β}	0.2	0.936	0.949	0.954	0.955
	4	0.936	0.957	0.943	0.954
$\psi_{\alpha\alpha}$	0.2	0.957	0.955	0.950	0.945
	4	0.958	0.957	0.947	0.944
$\psi_{\beta\beta}$	0.2	0.956	0.950	0.953	0.955
	4	0.948	0.941	0.956	0.955

Tabela F.11: *Coverage* dos modelos A1 e A4, com correlação(α, β) = 0.9 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.953	0.959	0.954	0.954
	4	0.953	0.954	0.949	0.959
μ_{α}	0.2	0.948	0.955	0.954	0.960
	4	0.948	0.953	0.958	0.960
μ_{β}	0.2	0.929	0.947	0.957	0.955
	4	0.934	0.950	0.946	0.952
$\psi_{\alpha\alpha}$	0.2	0.956	0.955	0.953	0.946
	4	0.955	0.958	0.947	0.943
$\psi_{\beta\beta}$	0.2	0.955	0.947	0.955	0.950
	4	0.946	0.939	0.959	0.951

Tabela F.12: *Coverage* dos modelos A1 e A4, com correlação(α, β) = 0.5 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.955	0.960	0.947	0.950
	4	0.958	0.967	0.946	0.956
μ_{α}	0.2	0.949	0.951	0.953	0.960
	4	0.949	0.952	0.959	0.960
μ_{β}	0.2	0.938	0.942	0.954	0.954
	4	0.934	0.946	0.951	0.958
$\psi_{\alpha\alpha}$	0.2	0.951	0.954	0.951	0.945
	4	0.952	0.958	0.945	0.945
$\psi_{\beta\beta}$	0.2	0.951	0.952	0.953	0.957
	4	0.950	0.945	0.954	0.952

Tabela F.13: *Coverage* dos modelos A1 e A4, com correlação(α, β) = 0 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.954	0.953	0.946	0.942
	4	0.954	0.958	0.942	0.952
μ_{α}	0.2	0.946	0.947	0.949	0.959
	4	0.948	0.954	0.957	0.960
μ_{β}	0.2	0.932	0.942	0.951	0.953
	4	0.931	0.944	0.947	0.960
$\psi_{\alpha\alpha}$	0.2	0.946	0.953	0.949	0.945
	4	0.954	0.953	0.945	0.941
$\psi_{\beta\beta}$	0.2	0.955	0.947	0.952	0.952
	4	0.958	0.949	0.956	0.948

Tabela F.14: *Coverage* dos modelos A1 e A4, com correlação(α, β) = -0.5 e $T = 3$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.947	0.951	0.949	0.946
	4	0.943	0.956	0.940	0.954
μ_α	0.2	0.948	0.948	0.947	0.957
	4	0.952	0.953	0.952	0.959
μ_β	0.2	0.947	0.947	0.948	0.955
	4	0.931	0.954	0.953	0.949
$\psi_{\alpha\alpha}$	0.2	0.933	0.952	0.943	0.949
	4	0.951	0.952	0.942	0.943
$\psi_{\beta\beta}$	0.2	0.955	0.951	0.950	0.951
	4	0.932	0.967	0.957	0.949

Tabela F.15: *Coverage* dos modelos A1 e A4, com correlação(α, β) = -0.9 e $T = 3$

	$\psi_{\beta\beta}$	Enviesamento		EQM		<i>Coverage</i>	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
$\psi_{\alpha\beta}$	0.2	0.0151 (0.0686)	0.0037 (0.0168)	0.0615	0.0309	0.959	0.950
	4	0.0364 (0.0364)	0.0074 (0.0074)	0.4936	0.2462	0.950	0.941
μ_α	0.2	-0.0042	-0.0002	0.0093	0.0048	0.946	0.947
	4	-0.0043	-0.0001	0.0101	0.0052	0.944	0.942
μ_β	0.2	0.0034	0.0008	0.0060	0.0029	0.950	0.956
	4	0.0051	0.0014	0.0342	0.0172	0.952	0.952
$\psi_{\alpha\alpha}$	0.2	-0.0244 (-0.0244)	-0.0072 (-0.0072)	0.1186	0.0599	0.952	0.950
	4	-0.0500 (-0.0500)	-0.0122 (-0.0122)	0.6539	0.3277	0.959	0.952
$\psi_{\beta\beta}$	0.2	-0.0168 (-0.0840)	-0.0054 (-0.0270)	0.0593	0.0280	0.941	0.952
	4	-0.0671 (-0.0168)	-0.0250 (-0.0063)	1.0053	0.4707	0.942	0.947

Tabela F.16: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos A1 e A4, com correlação(α, β) = 0.5 , $T = 3$ e omissões

	$\psi_{\beta\beta}$	Enviesamento		EQM		<i>Coverage</i>	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
$\psi_{\alpha\beta}$	0.2	0.0144	0.0037	0.0444	0.0224	0.959	0.946
	4	0.0418	0.0098	0.4361	0.2156	0.941	0.937
μ_α	0.2	-0.0042	-0.0003	0.0091	0.0047	0.945	0.945
	4	-0.0045	-0.0003	0.0104	0.0053	0.943	0.941
μ_β	0.2	0.0035	0.0008	0.0051	0.0025	0.942	0.959
	4	0.0062	0.0013	0.0315	0.0156	0.948	0.956
$\psi_{\alpha\alpha}$	0.2	0.0222 (0.0222)	0.0067 (0.0067)	0.0895	0.0454	0.947	0.949
	4	-0.0492 (-0.0492)	-0.0121 (-0.0121)	0.5382	0.2696	0.953	0.952
$\psi_{\beta\beta}$	0.2	0.0146 (0.0730)	-0.0045 (-0.0225)	0.0401	0.0191	0.944	0.957
	4	-0.0629 (-0.0157)	-0.0219 (-0.0055)	0.8187	0.3783	0.938	0.950

Tabela F.17: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos A1 e A4, com correlação(α, β) = 0 , $T = 3$ e omissões

	$\psi_{\beta\beta}$	Enviesamento		EQM		Coverage	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
$\psi_{\alpha\beta}$	0.2	0.0124 (-0.0564)	0.0044 (-0.0200)	0.0260	0.0132	0.955	0.946
	4	0.0427 (-0.0427)	0.0105 (-0.0105)	0.3454	0.1691	0.936	0.944
μ_{α}	0.2	-0.0041	-0.0005	0.0086	0.0045	0.945	0.946
	4	-0.0045	-0.0005	0.0103	0.0052	0.945	0.941
μ_{β}	0.2	0.0033	0.0008	0.0039	0.0019	0.945	0.952
	4	0.0070	0.0011	0.0284	0.0137	0.947	0.948
$\psi_{\alpha\alpha}$	0.2	-0.0188 (-0.0188)	-0.0059 (-0.0059)	0.0587	0.0300	0.949	0.952
	4	-0.0452 (-0.0452)	-0.0108 (-0.0108)	0.3865	0.1937	0.959	0.952
$\psi_{\beta\beta}$	0.2	-0.0112 (-0.0560)	-0.0034 (-0.0170)	0.0210	0.0101	0.940	0.950
	4	-0.0554 (-0.0139)	-0.0168 (-0.0042)	0.6210	0.2827	0.940	0.953

Tabela F.18: Enviesamento bruto (enviesamento relativo), erro quadrático médio e coverage para os parâmetros, dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.5$, $T = 3$ e omissões

Anexo G

Tabelas do estudo de simulação: Dados com distribuição mistura de normais e $T=3$

Modelo	Parâmetro	N=50	N=250	N=500	N = 1 000
H1	$\psi_{\alpha\beta}$	0.0016 (0.6957)	0.0088 (3.8261)	0.0005 (0.2174)	0.0008 (0.3478)
	μ_{α}	-0.0023 (-0.0077)	-0.0020 (-0.0067)	-0.0011 (-0.0037)	-0.0027 (-0.0090)
	μ_{β}	0.0002 (0.0282)	-0.0042 (-0.5915)	-0.0060 (-0.8451)	-0.0064 (-0.9014)
	$\psi_{\alpha\alpha}$	-0.0417 (-0.0250)	-0.0237 (-0.0142)	-0.0181 (-0.0109)	-0.0181 (-0.0109)
	$\psi_{\beta\beta}$	-0.0049 (-0.0243)	-0.0096 (-0.0477)	-0.0047 (-0.0233)	-0.0051 (-0.0253)
H2	$\psi_{\alpha\beta}$	-0.0368 (-0.0342)	0.0018 (0.0017)	-0.0224 (-0.0208)	-0.0239 (-0.0222)
	μ_{α}	-0.0030 (-0.0100)	-0.0018 (-0.0060)	-0.0018 (-0.0060)	-0.0039 (-0.0130)
	μ_{β}	-0.0040 (-0.0081)	-0.0069 (-0.0140)	-0.0123 (-0.0250)	-0.0126 (-0.0256)
	$\psi_{\alpha\alpha}$	-0.0596 (-0.0277)	-0.0305 (-0.0142)	-0.0295 (-0.0137)	-0.0311 (-0.0144)
	$\psi_{\beta\beta}$	-0.0484 (-0.0254)	-0.0301 (-0.0158)	-0.0351 (-0.0184)	-0.0326 (-0.0171)
H3	$\psi_{\alpha\beta}$	0.0135 (0.6459)	0.0191 (0.9139)	-0.0100 (-0.4785)	-0.0072 (-0.3445)
	μ_{α}	-0.0155 (-0.0085)	-0.0102 (-0.0056)	-0.0200 (-0.0110)	-0.0211 (-0.0116)
	μ_{β}	0.0002 (0.0282)	-0.0042 (-0.5915)	-0.0060 (-0.8451)	-0.0064 (-0.9014)
	$\psi_{\alpha\alpha}$	-0.7078 (-0.0283)	-0.2382 (-0.0095)	-0.3425 (-0.0137)	-0.3213 (-0.0129)
	$\psi_{\beta\beta}$	-0.0104 (-0.0548)	-0.0211 (-0.1112)	-0.0035 (-0.0184)	-0.0039 (-0.0205)
H4	$\psi_{\alpha\beta}$	-0.1691 (-0.0263)	-0.0343 (-0.0053)	-0.1019 (-0.0159)	-0.0938 (-0.0146)
	μ_{α}	-0.0158 (-0.0087)	-0.0100 (-0.0055)	-0.0207 (-0.0114)	-0.0222 (-0.0122)
	μ_{β}	-0.0040 (-0.0081)	-0.0069 (-0.0140)	-0.0123 (-0.0250)	-0.0126 (-0.0256)
	$\psi_{\alpha\alpha}$	-0.7436 (-0.0292)	-0.2409 (-0.0095)	-0.3703 (-0.0145)	-0.3613 (-0.0142)
	$\psi_{\beta\beta}$	-0.0551 (-0.0291)	-0.0415 (-0.0219)	-0.0351 (-0.0185)	-0.0334 (-0.0176)
H5	$\psi_{\alpha\beta}$	-0.1887 (-0.0293)	-0.0311 (-0.0048)	-0.1165 (-0.0181)	-0.1183 (-0.0184)
	μ_{α}	-0.0178 (-0.0098)	-0.0092 (-0.0051)	-0.0228 (-0.0125)	-0.0258 (-0.0142)
	μ_{β}	-0.0040 (-0.0081)	-0.0069 (-0.0140)	-0.0125 (-0.0254)	-0.0128 (-0.0260)
	$\psi_{\alpha\alpha}$	-0.9295 (-0.0319)	-0.2745 (-0.0094)	-0.4900 (-0.0168)	-0.5150 (-0.0177)
	$\psi_{\beta\beta}$	-0.0509 (-0.0269)	-0.0440 (-0.0232)	-0.0382 (-0.0202)	-0.0382 (-0.0202)

Tabela G.1: Enviesamento bruto (enviesamento relativo) na estimativa média dos parâmetros dos modelos H1 a H5, e $T = 3$

Modelo	Parâmetro	N=50	N=250	N=500	$N = 1\ 000$
H1	$\psi_{\alpha\beta}$	0.1695	0.0350	0.0176	0.0082
	μ_{α}	0.0509	0.0101	0.0049	0.0024
	μ_{β}	0.0187	0.0036	0.0018	0.0009
	$\psi_{\alpha\alpha}$	0.5143	0.1077	0.0507	0.0264
	$\psi_{\beta\beta}$	0.1327	0.0289	0.0138	0.0064
H2	$\psi_{\alpha\beta}$	0.4851	0.1083	0.0559	0.0265
	μ_{α}	0.0602	0.0124	0.0060	0.0029
	μ_{β}	0.0533	0.0107	0.0053	0.0027
	$\psi_{\alpha\alpha}$	0.9817	0.2098	0.0984	0.0517
	$\psi_{\beta\beta}$	0.6918	0.1420	0.0677	0.0321
H3	$\psi_{\alpha\beta}$	1.7183	0.3707	0.1728	0.0786
	μ_{α}	0.5026	0.1017	0.0491	0.0259
	μ_{β}	0.0187	0.0036	0.0018	0.0009
	$\psi_{\alpha\alpha}$	64.0482	14.1086	6.5744	3.4136
	$\psi_{\beta\beta}$	1.0556	0.2317	0.1044	0.0484
H4	$\psi_{\alpha\beta}$	6.0007	1.2957	0.6668	0.3294
	μ_{α}	0.5103	0.1049	0.0505	0.0265
	μ_{β}	0.0533	0.0107	0.0053	0.0027
	$\psi_{\alpha\alpha}$	72.1611	16.4020	7.6280	3.9414
	$\psi_{\beta\beta}$	1.9655	0.4149	0.1823	0.0859
H5	$\psi_{\alpha\beta}$	7.0309	1.5706	0.8067	0.3992
	μ_{α}	0.5741	0.1223	0.0590	0.0307
	μ_{β}	0.0532	0.0107	0.0054	0.0027
	$\psi_{\alpha\alpha}$	143.2563	33.8362	15.8607	8.1521
	$\psi_{\beta\beta}$	2.1226	0.4431	0.1933	0.0913

Tabela G.2: Erro quadrático médio de estimação para os parâmetros, dos modelos H1 a H5, e $T = 3$

Modelo	Parâmetro	N=50	N=250	N=500	$N = 1\ 000$
H1	$\psi_{\alpha\beta}$	0.955	0.949	0.944	0.950
	μ_{α}	0.949	0.941	0.948	0.958
	μ_{β}	0.934	0.945	0.952	0.950
	$\psi_{\alpha\alpha}$	0.914	0.937	0.955	0.938
	$\psi_{\beta\beta}$	0.942	0.938	0.943	0.952
H2	$\psi_{\alpha\beta}$	0.902	0.938	0.927	0.928
	μ_{α}	0.946	0.939	0.943	0.949
	μ_{β}	0.928	0.936	0.948	0.948
	$\psi_{\alpha\alpha}$	0.889	0.919	0.939	0.929
	$\psi_{\beta\beta}$	0.902	0.932	0.944	0.955
H3	$\psi_{\alpha\beta}$	0.950	0.940	0.951	0.954
	μ_{α}	0.914	0.938	0.954	0.950
	μ_{β}	0.934	0.945	0.952	0.950
	$\psi_{\alpha\alpha}$	0.919	0.934	0.945	0.943
	$\psi_{\beta\beta}$	0.942	0.933	0.948	0.949
H4	$\psi_{\alpha\beta}$	0.902	0.935	0.929	0.939
	μ_{α}	0.911	0.936	0.948	0.951
	μ_{β}	0.928	0.936	0.948	0.948
	$\psi_{\alpha\alpha}$	0.918	0.931	0.945	0.938
	$\psi_{\beta\beta}$	0.894	0.935	0.944	0.956
H5	$\psi_{\alpha\beta}$	0.885	0.936	0.925	0.926
	μ_{α}	0.906	0.936	0.947	0.941
	μ_{β}	0.927	0.936	0.948	0.948
	$\psi_{\alpha\alpha}$	0.874	0.916	0.932	0.937
	$\psi_{\beta\beta}$	0.900	0.939	0.946	0.953

Tabela G.3: *Coverage* para os parâmetros, dos modelos H1 a H5, e $T = 3$

Modelo		Enviesamento		EQM		Coverage	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
H1	$\psi_{\alpha\beta}$	0.0139 (6.0435)	0.0041 (1.7826)	0.0526	0.0274	0.955	0.954
	μ_{α}	-0.0059 (-0.0197)	-0.0026 (-0.0087)	0.0120	0.0062	0.949	0.934
	μ_{β}	-0.0032 (-0.4507)	-0.0060 (-0.8451)	0.0052	0.0026	0.945	0.954
	$\psi_{\alpha\alpha}$	-0.0395 (-0.0237)	-0.0197 (-0.0118)	0.1449	0.0737	0.939	0.934
	$\psi_{\beta\beta}$	-0.0173 (-0.0859)	-0.0075 (-0.0373)	0.0438	0.0214	0.948	0.956
H2	$\psi_{\alpha\beta}$	-0.0146 (-0.0136)	-0.0161 (-0.0150)	0.1423	0.0739	0.943	0.950
	μ_{α}	-0.0058 (-0.0193)	-0.0018 (-0.0060)	0.0141	0.0073	0.941	0.936
	μ_{β}	-0.0109 (-0.0221)	-0.0149 (-0.0303)	0.0127	0.0066	0.943	0.939
	$\psi_{\alpha\alpha}$	-0.0526 (-0.0244)	-0.0285 (-0.0132)	0.2764	0.1392	0.924	0.934
	$\psi_{\beta\beta}$	-0.0504 (-0.0264)	-0.0393 (-0.0206)	0.1804	0.0952	0.936	0.940
H3	$\psi_{\alpha\beta}$	0.0069 (0.3301)	-0.0006 (-0.0287)	0.3984	0.2058	0.959	0.959
	μ_{α}	-0.0252 (-0.0139)	-0.0261 (-0.0144)	0.1137	0.0580	0.940	0.923
	μ_{β}	-0.0031 (-0.4366)	-0.0060 (-0.8451)	0.0054	0.0027	0.947	0.953
	$\psi_{\alpha\alpha}$	-0.4392 (-0.0176)	-0.3975 (-0.0159)	15.4460	7.8317	0.924	0.924
	$\psi_{\beta\beta}$	-0.0109 (-0.0574)	-0.0047 (-0.0248)	0.2242	0.1185	0.963	0.961
H4	$\psi_{\alpha\beta}$	-0.0987 (-0.0154)	-0.1039 (-0.0162)	1.3722	0.7323	0.934	0.932
	μ_{α}	-0.0255 (-0.0140)	-0.0257 (-0.0142)	0.1153	0.0589	0.936	0.924
	μ_{β}	-0.0099 (-0.0201)	-0.0141 (-0.0287)	0.0126	0.0066	0.939	0.936
	$\psi_{\alpha\alpha}$	-0.4737 (-0.0186)	-0.4031 (-0.0158)	17.4917	8.8014	0.926	0.921
	$\psi_{\beta\beta}$	-0.0463 (-0.0244)	-0.0352 (-0.0186)	0.4184	0.2294	0.951	0.945
H5	$\psi_{\alpha\beta}$	-0.1021 (-0.0159)	-0.1038 (-0.0161)	1.6514	0.8784	0.932	0.935
	μ_{α}	-0.0294 (-0.0162)	-0.0259 (-0.0143)	0.1302	0.0660	0.932	0.925
	μ_{β}	-0.0097 (-0.0197)	-0.0140 (-0.0284)	0.0127	0.0066	0.941	0.938
	$\psi_{\alpha\alpha}$	-0.6789 (-0.0233)	-0.4930 (-0.0169)	34.7017	17.1841	0.911	0.930
	$\psi_{\beta\beta}$	-0.0395 (-0.0208)	-0.0300 (-0.0158)	0.5159	0.2764	0.952	0.945

Tabela G.4: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos H1 e H5, com $T = 3$ e omissões

Anexo H

Tabelas do estudo de simulação: Dados com distribuição normal e $T=5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.0118 (-0.0295)	0.0024 (-0.0060)	0.0003 (-0.0008)	-0.0001 (0.0003)
	4	0.0422 (-0.0234)	0.0139 (-0.0077)	0.0106 (-0.0059)	0.0014 (-0.0008)
μ_{α}	0.2	0.0041	0.0014	-0.0010	-0.0004
	4	0.0042	0.0013	-0.0010	0
μ_{β}	0.2	-0.0014	-0.0004	0.0004	0.0001
	4	0.0005	0.0006	0.0016	-0.0009
$\psi_{\alpha\alpha}$	0.2	-0.0284 (-0.0284)	-0.0049 (-0.0049)	0.0007 (0.0007)	0.0009 (0.0009)
	4	-0.0390 (-0.0390)	-0.0054 (-0.0054)	-0.0030 (-0.0030)	0.0006 (0.0006)
$\psi_{\beta\beta}$	0.2	-0.0057 (-0.0285)	-0.0015 (-0.0075)	-0.0005 (-0.0025)	-0.0003 (-0.0015)
	4	-0.0613 (-0.0153)	-0.0289 (-0.0073)	-0.0259 (-0.0065)	-0.0082 (-0.0021)

Tabela H.1: Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = -0.9 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.0108 (-0.0491)	0.0027 (-0.0123)	0.0010 (-0.0046)	0.0005 (-0.0023)
	4	0.0301 (-0.0301)	0.0120 (-0.0120)	0.0113 (-0.0113)	0.0038 (-0.0038)
μ_{α}	0.2	0.0043	0.0014	-0.0011	-0.0003
	4	0.0049	0.0013	-0.0011	0.0001
μ_{β}	0.2	-0.0001	0	0.0004	0
	4	0.0041	0.0019	0.0013	-0.0010
$\psi_{\alpha\alpha}$	0.2	-0.0315 (-0.0315)	-0.0053 (-0.0053)	0.0015 (0.0015)	0.0010 (0.0010)
	4	-0.0408 (-0.0408)	-0.0053 (-0.0053)	-0.0011 (-0.0011)	-0.0007 (-0.0007)
$\psi_{\beta\beta}$	0.2	-0.0061 (-0.0305)	-0.0023 (-0.0115)	-0.0013 (-0.0065)	-0.0012 (-0.0060)
	4	-0.0493 (-0.0123)	-0.0259 (-0.0065)	-0.0261 (-0.0065)	-0.0146 (-0.0037)

Tabela H.2: Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = -0.5 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.0093	0.0026	0.0013	0.0011
	4	0.0162	0.0093	0.0106	0.0063
μ_α	0.2	0.0043	0.0014	-0.0011	-0.0002
	4	0.0049	0.0013	-0.0011	0.0001
μ_β	0.2	0.0009	0.0003	0.0004	0
	4	0.0070	0.0027	0.0008	-0.0010
$\psi_{\alpha\alpha}$	0.2	-0.0346 (-0.0346)	-0.0060 (-0.0060)	0.0023 (0.0023)	0.0011 (0.0011)
	4	-0.0429 (-0.0429)	-0.0075 (-0.0075)	0.0009 (0.0009)	-0.0023 (-0.0023)
$\psi_{\beta\beta}$	0.2	-0.0061 (-0.0305)	-0.0026 (-0.0130)	-0.0014 (-0.0070)	-0.0017 (-0.0085)
	4	-0.0439 (-0.0110)	-0.0215 (-0.0054)	-0.0184 (-0.0046)	-0.0167 (-0.0042)

Tabela H.3: Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = 0 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.0075 (0.0341)	0.0022 (0.0100)	0.0013 (0.0059)	0.0014 (0.0064)
	4	0.0043 (0.0043)	0.0059 (0.0059)	0.0088 (0.0088)	0.0079 (0.0079)
μ_α	0.2	0.0043	0.0014	-0.0011	-0.0002
	4	0.0047	0.0013	-0.0012	0.0001
μ_β	0.2	0.0018	0.0005	0.0003	0
	4	0.0091	0.0031	0.0003	-0.0007
$\psi_{\alpha\alpha}$	0.2	-0.0370 (-0.0370)	-0.0065 (-0.0065)	0.0030 (0.0030)	0.0013 (0.0013)
	4	-0.0460 (-0.0460)	-0.0100 (-0.0100)	0.0029 (0.0029)	-0.0027 (-0.0027)
$\psi_{\beta\beta}$	0.2	-0.0062 (-0.0310)	-0.0027 (-0.0135)	-0.0010 (-0.0050)	-0.0018 (-0.0090)
	4	-0.0467 (-0.0117)	-0.0185 (-0.0046)	-0.0072 (-0.0018)	-0.0138 (-0.0035)

Tabela H.4: Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = 0.5 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.0058 (0.0145)	0.0017 (0.0043)	0.0012 (0.0030)	0.0017 (0.0043)
	4	-0.0041 (-0.0023)	0.0024 (0.0013)	0.0065 (0.0036)	0.0084 (0.0047)
μ_α	0.2	0.0042	0.0014	-0.0011	-0.0001
	4	0.0045	0.0013	-0.0012	0.0001
μ_β	0.2	0.0023	0.0006	0.0002	0
	4	0.0101	0.0032	-0.0001	-0.0004
$\psi_{\alpha\alpha}$	0.2	-0.0386 (-0.0386)	-0.0069 (-0.0069)	0.0034 (0.0034)	0.0016 (0.0016)
	4	-0.0495 (-0.0495)	0.0114 (0.0114)	0.0044 (0.0044)	-0.0019 (-0.0019)
$\psi_{\beta\beta}$	0.2	-0.0063 (-0.0315)	-0.0026 (-0.0130)	-0.0005 (-0.0025)	-0.0016 (-0.0080)
	4	-0.0534 (-0.0134)	-0.0175 (-0.0044)	0.0021 (0.0005)	-0.0082 (-0.0021)

Tabela H.5: Enviesamento bruto (enviesamento relativo) na estimativa média para os parâmetros dos modelos A1 e A4, com correlação(α, β) = 0.9 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.0161	0.0029	0.0015	0.0007
	4	0.3539	0.0685	0.0354	0.0175
μ_α	0.2	0.0308	0.0057	0.0027	0.0014
	4	0.0388	0.0078	0.0037	0.0019
μ_β	0.2	0.0062	0.0012	0.0005	0.0003
	4	0.1007	0.0208	0.0100	0.0050
$\psi_{\alpha\alpha}$	0.2	0.0902	0.0163	0.0085	0.0042
	4	0.2949	0.0529	0.0274	0.0139
$\psi_{\beta\beta}$	0.2	0.0035	0.0006	0.0003	0.0002
	4	0.9955	0.1933	0.0994	0.0509

Tabela H.6: Erro quadrático médio, dos modelos A1 e A4, com correlação(α, β) = -0.9 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.0191	0.0035	0.0017	0.0009
	4	0.3348	0.0633	0.0324	0.0162
μ_α	0.2	0.0349	0.0066	0.0031	0.0016
	4	0.0402	0.0080	0.0038	0.0019
μ_β	0.2	0.0075	0.0014	0.0007	0.0008
	4	0.1049	0.0211	0.0101	0.0052
$\psi_{\alpha\alpha}$	0.2	0.1289	0.0232	0.0121	0.0059
	4	0.4831	0.0844	0.0437	0.0222
$\psi_{\beta\beta}$	0.2	0.0057	0.0011	0.0005	0.0003
	4	1.1077	0.2086	0.1044	0.0545

Tabela H.7: Erro quadrático médio, dos modelos A1 e A4, com correlação(α, β) = -0.5 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.0226	0.0040	0.0020	0.0010
	4	0.3396	0.0623	0.0312	0.0157
μ_α	0.2	0.0369	0.0071	0.0034	0.0017
	4	0.0408	0.0081	0.0039	0.0020
μ_β	0.2	0.0088	0.0017	0.0034	0.0017
	4	0.0408	0.0081	0.0039	0.0020
$\psi_{\alpha\alpha}$	0.2	0.1652	0.0295	0.0155	0.0075
	4	0.6725	0.1168	0.0605	0.0303
$\psi_{\beta\beta}$	0.2	0.0088	0.0016	0.0008	0.0004
	4	1.2309	0.2309	0.1149	0.0588

Tabela H.8: Erro quadrático médio, dos modelos A1 e A4, com correlação(α, β) = 0 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.0269	0.0047	0.0023	0.0011
	4	0.3705	0.0662	0.0327	0.0164
μ_{α}	0.2	0.0379	0.0073	0.0035	0.0018
	4	0.0410	0.0082	0.0039	0.0020
μ_{β}	0.2	0.0099	0.0019	0.0009	0.0005
	4	0.1137	0.0219	0.0105	0.0056
$\psi_{\alpha\alpha}$	0.2	0.1950	0.0345	0.0181	0.0088
	4	0.7629	0.1331	0.0689	0.0338
$\psi_{\beta\beta}$	0.2	0.0119	0.0021	0.0011	0.0005
	4	1.2886	0.2431	0.1256	0.0620

Tabela H.9: Erro quadrático médio, dos modelos A1 e A4, com correlação(α, β) = 0.5 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.0311	0.0054	0.0027	0.0013
	4	0.4135	0.0725	0.0356	0.0177
μ_{α}	0.2	0.0384	0.0075	0.0036	0.0018
	4	0.0409	0.0082	0.0039	0.0020
μ_{β}	0.2	0.0107	0.0020	0.0010	0.0005
	4	0.1166	0.0225	0.0107	0.0057
$\psi_{\alpha\alpha}$	0.2	0.2138	0.0377	0.0197	0.0095
	4	0.7547	0.1320	0.0679	0.0328
$\psi_{\beta\beta}$	0.2	0.0143	0.0026	0.0013	0.0007
	4	1.3357	0.2489	0.1325	0.0643

Tabela H.10: Erro quadrático médio, dos modelos A1 e A4, com correlação(α, β) = 0.9 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.908	0.941	0.941	0.945
	4	0.911	0.938	0.944	0.945
μ_{α}	0.2	0.927	0.942	0.950	0.958
	4	0.936	0.947	0.948	0.955
μ_{β}	0.2	0.933	0.948	0.955	0.952
	4	0.943	0.939	0.941	0.944
$\psi_{\alpha\alpha}$	0.2	0.903	0.936	0.944	0.947
	4	0.933	0.948	0.935	0.943
$\psi_{\beta\beta}$	0.2	0.924	0.946	0.941	0.942
	4	0.918	0.936	0.948	0.953

Tabela H.11: *Coverage*, dos modelos A1 e A4, com correlação(α, β) = -0.9 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.920	0.949	0.948	0.947
	4	0.931	0.954	0.951	0.948
μ_{α}	0.2	0.922	0.943	0.953	0.957
	4	0.935	0.948	0.946	0.953
μ_{β}	0.2	0.938	0.942	0.946	0.942
	4	0.945	0.945	0.950	0.945
$\psi_{\alpha\alpha}$	0.2	0.910	0.944	0.941	0.948
	4	0.935	0.956	0.938	0.951
$\psi_{\beta\beta}$	0.2	0.922	0.954	0.951	0.945
	4	0.910	0.945	0.961	0.944

Tabela H.12: *Coverage*, dos modelos A1 e A4, com correlação(α, β) = -0.5 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.939	0.959	0.955	0.959
	4	0.941	0.961	0.959	0.949
μ_{α}	0.2	0.929	0.939	0.952	0.956
	4	0.935	0.943	0.943	0.952
μ_{β}	0.2	0.942	0.947	0.948	0.945
	4	0.942	0.947	0.956	0.937
$\psi_{\alpha\alpha}$	0.2	0.922	0.937	0.944	0.949
	4	0.940	0.957	0.941	0.950
$\psi_{\beta\beta}$	0.2	0.926	0.947	0.947	0.948
	4	0.916	0.944	0.953	0.945

Tabela H.13: *Coverage*, dos modelos A1 e A4, com correlação(α, β) = 0 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1000$
$\psi_{\alpha\beta}$	0.2	0.946	0.962	0.957	0.963
	4	0.947	0.965	0.952	0.963
μ_{α}	0.2	0.933	0.940	0.952	0.952
	4	0.937	0.944	0.945	0.952
μ_{β}	0.2	0.942	0.951	0.942	0.943
	4	0.948	0.954	0.952	0.938
$\psi_{\alpha\alpha}$	0.2	0.930	0.941	0.941	0.952
	4	0.941	0.954	0.943	0.952
$\psi_{\beta\beta}$	0.2	0.937	0.945	0.948	0.950
	4	0.923	0.956	0.956	0.951

Tabela H.14: *Coverage*, dos modelos A1 e A4, com correlação(α, β) = 0.5 e $T = 5$

	$\psi_{\beta\beta}$	$N = 50$	$N = 250$	$N = 500$	$N = 1\,000$
$\psi_{\alpha\beta}$	0.2	0.951	0.963	0.957	0.966
	4	0.935	0.967	0.958	0.970
μ_α	0.2	0.936	0.939	0.949	0.952
	4	0.940	0.942	0.946	0.952
μ_β	0.2	0.947	0.950	0.943	0.941
	4	0.948	0.956	0.955	0.944
$\psi_{\alpha\alpha}$	0.2	0.924	0.943	0.944	0.950
	4	0.938	0.954	0.948	0.951
$\psi_{\beta\beta}$	0.2	0.936	0.953	0.944	0.953
	4	0.939	0.965	0.959	0.956

Tabela H.15: *Coverage*, dos modelos A1 e A4, com correlação(α, β) = 0.9 e $T = 5$

	$\psi_{\beta\beta}$	Enviesamento		EQM		<i>Coverage</i>	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
$\psi_{\alpha\beta}$	0.2	0.0098 (0.0445)	0.0026 (0.0118)	0.0224	0.0116	0.954	0.960
	4	0.0321 (0.0321)	0.0281 (0.0281)	0.3334	0.1602	0.950	0.958
μ_α	0.2	0.0019	-0.0029	0.0194	0.0089	0.945	0.954
	4	0.0030	-0.0024	0.0282	0.0129	0.936	0.954
μ_β	0.2	0.0005	0.0011	0.0036	0.0017	0.940	0.950
	4	0.0030	0.0011	0.0269	0.0125	0.950	0.946
$\psi_{\alpha\alpha}$	0.2	-0.0152 (-0.0152)	0.0092 (0.0092)	0.1373	0.0732	0.958	0.953
	4	-0.0380 (-0.0380)	0.0041 (0.0041)	0.7019	0.3577	0.949	0.953
$\psi_{\beta\beta}$	0.2	-0.0084 (-0.0420)	-0.0032 (-0.0160)	0.0083	0.0040	0.945	0.959
	4	-0.0476 (-0.0119)	-0.0277 (-0.0069)	0.4224	0.2030	0.940	0.954

Tabela H.16: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos A1 e A4, com correlação(α, β) = 0.5, $T = 5$ e omissões

	$\psi_{\beta\beta}$	Enviesamento		EQM		<i>Coverage</i>	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
$\psi_{\alpha\beta}$	0.2	0.0080	0.0017	0.0173	0.0090	0.953	0.955
	4	0.0298	0.0304	0.3204	0.1526	0.955	0.958
μ_α	0.2	0.0026	-0.0024	0.0180	0.0082	0.947	0.954
	4	0.0049	-0.0014	0.0297	0.0135	0.945	0.951
μ_β	0.2	0.0001	0.0010	0.0031	0.0015	0.943	0.947
	4	0.0019	0.0012	0.0261	0.0122	0.956	0.947
$\psi_{\alpha\alpha}$	0.2	-0.0128 (-0.0128)	0.0078 (0.0078)	0.1100	0.0583	0.955	0.952
	4	-0.0301 (-0.0301)	-0.0030 (-0.0030)	0.5992	0.3041	0.949	0.946
$\psi_{\beta\beta}$	0.2	-0.0066 (-0.0330)	-0.0028 (-0.0140)	0.0055	0.0027	0.946	0.952
	4	-0.0467 (-0.0117)	-0.0385 (-0.0096)	0.3955	0.1815	0.937	0.954

Tabela H.17: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos A1 e A4, com correlação(α, β) = 0, $T = 5$ e omissões

	$\psi_{\beta\beta}$	Enviesamento		EQM		Coverage	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
$\psi_{\alpha\beta}$	0.2	0.0052 (-0.0236)	0.0004 (-0.0018)	0.00117	0.0062	0.953	0.953
	4	0.0243 (-0.0243)	0.0256 (-0.0256)	0.2695	0.1318	0.953	0.948
μ_{α}	0.2	0.0030	-0.0017	0.0149	0.0068	0.949	0.951
	4	0.0059	-0.0003	0.0271	0.0122	0.940	0.952
μ_{β}	0.2	-0.0004	0.0008	0.0025	0.0011	0.933	0.947
	4	0.0007	0.0012	0.0254	0.0118	0.947	0.949
$\psi_{\alpha\alpha}$	0.2	-0.0093 (-0.0093)	0.0058 (0.0058)	0.0756	0.0398	0.951	0.951
	4	-0.0183 (-0.0183)	-0.0082 (-0.0082)	0.4165	0.2103	0.961	0.948
$\psi_{\beta\beta}$	0.2	-0.0042 (-0.0210)	-0.0016 (-0.0080)	0.0030	0.0015	0.951	0.947
	4	-0.0444 (-0.0111)	-0.0407 (-0.0102)	0.3465	0.1582	0.938	0.955

Tabela H.18: Enviesamento bruto (enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos A1 e A4, com correlação $(\alpha, \beta) = -0.5$, $T = 5$ e omissões

Anexo I

Tabelas do estudo de simulação: Dados com distribuição mistura de normais e $T=5$

Modelo	Parâmetro	N=50	N=250	N=500	N = 1 000
H1	$\psi_{\alpha\beta}$	0.0076 (6.9091)	0.0005 (0.4545)	-0.0010 (-0.9091)	-0.0009 (-0.8182)
	μ_{α}	0.0039 (0.0130)	0.0027 (0.0090)	-0.0021 (-0.0070)	-0.0010 (-0.0033)
	μ_{β}	0 (0)	-0.0009 (-0.8182)	-0.0007 (-0.6364)	-0.0011 (-1)
	$\psi_{\alpha\alpha}$	0.0509 (0.0308)	0.0513 (0.0310)	0.0547 (0.0331)	0.0550 (0.0333)
	$\psi_{\beta\beta}$	-0.0073 (-0.0364)	-0.0035 (-0.0175)	-0.0020 (-0.0100)	-0.0023 (-0.0115)
H2	$\psi_{\alpha\beta}$	-0.0139 (-0.0132)	0.0084 (0.0080)	-0.0024 (-0.0023)	0.0002 (0.0002)
	μ_{α}	0.0063 (0.0211)	0.0046 (0.0154)	-0.0012 (-0.0040)	0.0002 (6.7408)
	μ_{β}	-0.0053 (-0.0109)	-0.0033 (-0.0068)	-0.0066 (-0.0136)	-0.0067 (-0.0138)
	$\psi_{\alpha\alpha}$	-0.0557 (-0.0260)	-0.0068 (-0.0032)	-0.0112 (-0.0052)	-0.0079 (-0.0037)
	$\psi_{\beta\beta}$	-0.0553 (-0.0291)	-0.0170 (-0.0090)	-0.0223 (-0.0117)	-0.0193 (-0.0102)
H3	$\psi_{\alpha\beta}$	0.0029 (-0.6304)	-0.0009 (-0.1957)	-0.0015 (0.3261)	0.0001 (-0.0217)
	μ_{α}	-0.0116 (-0.0064)	-0.0056 (-0.0031)	-0.0212 (-0.0117)	-0.0194 (-0.0107)
	μ_{β}	0.0003 (0.2727)	-0.0008 (-0.7273)	-0.0007 (-0.6364)	-0.0011 (-1)
	$\psi_{\alpha\alpha}$	-0.6694 (-0.0269)	-0.1421 (-0.0057)	-0.2485 (-0.0100)	-0.2138 (-0.0086)
	$\psi_{\beta\beta}$	-0.0072 (-0.0360)	-0.0032 (-0.0160)	-0.0014 (-0.0070)	-0.0017 (-0.0085)
H4	$\psi_{\alpha\beta}$	-0.1626 (-0.0255)	-0.0361 (-0.0057)	-0.0686 (-0.0108)	-0.0571 (-0.0090)
	μ_{α}	-0.0094 (-0.0052)	-0.0037(-0.0020)	-0.0203 (-0.0112)	-0.0181 (-0.0100)
	μ_{β}	-0.0046 (-0.0095)	-0.0032 (-0.0066)	-0.0066 (-0.0136)	-0.0067 (-0.0138)
	$\psi_{\alpha\alpha}$	-0.6433 (-0.0254)	-0.1081 (-0.0043)	-0.2432 (-0.0096)	-0.1999 (-0.0079)
	$\psi_{\beta\beta}$	-0.0559 (-0.0294)	-0.0169 (-0.0089)	-0.0220 (-0.0116)	-0.0189 (-0.0100)
H5	$\psi_{\alpha\beta}$	-0.1443 (-0.0227)	-0.0120 (-0.0019)	-0.0548 (-0.0086)	-0.0417 (-0.0066)
	μ_{α}	-0.0040 (-0.0022)	0.0023 (0.0013)	-0.0174 (-0.0096)	-0.0144 (-0.0080)
	μ_{β}	-0.0048 (-0.0099)	-0.0031 (-0.0064)	-0.0065 (-0.0134)	-0.0067 (-0.0138)
	$\psi_{\alpha\alpha}$	-0.5987 (-0.0207)	-0.0648 (-0.0022)	-0.2970 (-0.0103)	-0.2168 (-0.0075)
	$\psi_{\beta\beta}$	-0.0552 (-0.0291)	-0.0156 (-0.0082)	-0.0211 (-0.0111)	-0.0180 (-0.0095)

Tabela I.1: Enviesamento bruto (enviesamento relativo) na estimativa média dos parâmetros dos modelos H1 a H5, e $T = 5$

Modelo	Parâmetro	N=50	N=250	N=500	$N = 1\ 000$
H1	$\psi_{\alpha\beta}$	0.0275	0.0053	0.0026	0.0013
	μ_{α}	0.0486	0.0097	0.0045	0.0024
	μ_{β}	0.0088	0.0017	0.0008	0.0004
	$\psi_{\alpha\alpha}$	0.3057	0.0551	0.0288	0.0147
	$\psi_{\beta\beta}$	0.0089	0.0016	0.0008	0.0004
H2	$\psi_{\alpha\beta}$	0.3137	0.0618	0.0294	0.0156
	μ_{α}	0.0592	0.0116	0.0054	0.0029
	μ_{β}	0.0418	0.0083	0.0040	0.0021
	$\psi_{\alpha\alpha}$	0.7754	0.1434	0.0713	0.0371
	$\psi_{\beta\beta}$	0.3471	0.0709	0.0343	0.0178
H3	$\psi_{\alpha\beta}$	0.2136	0.0431	0.0218	0.0114
	μ_{α}	0.4888	0.1016	0.0484	0.0258
	μ_{β}	0.0088	0.0017	0.0008	0.0004
	$\psi_{\alpha\alpha}$	61.9366	12.7918	6.1912	3.3035
	$\psi_{\beta\beta}$	0.0095	0.0018	0.0008	0.0004
H4	$\psi_{\alpha\beta}$	4.4009	0.8955	0.4372	0.2307
	μ_{α}	0.4984	0.1031	0.0491	0.0263
	μ_{β}	0.0419	0.0083	0.0040	0.0021
	$\psi_{\alpha\alpha}$	70.8978	14.6353	7.0556	3.8011
	$\psi_{\beta\beta}$	0.3467	0.0708	0.0342	0.0178
H5	$\psi_{\alpha\beta}$	5.2842	1.0683	0.5169	0.2763
	μ_{α}	0.5714	0.1168	0.0554	0.0300
	μ_{β}	0.0418	0.0083	0.0040	0.0021
	$\psi_{\alpha\alpha}$	150.3408	30.4141	14.5100	7.9014
	$\psi_{\beta\beta}$	0.3463	0.0708	0.0341	0.0178

Tabela I.2: Erro quadrático médio de estimação para os parâmetros, dos modelos H1 a H5, e $T = 5$

Modelo	Parâmetro	N=50	N=250	N=500	$N = 1\ 000$
H1	$\psi_{\alpha\beta}$	0.933	0.949	0.946	0.944
	μ_{α}	0.942	0.945	0.960	0.950
	μ_{β}	0.944	0.945	0.947	0.945
	$\psi_{\alpha\alpha}$	0.889	0.943	0.949	0.945
	$\psi_{\beta\beta}$	0.909	0.939	0.946	0.947
H2	$\psi_{\alpha\beta}$	0.883	0.943	0.958	0.947
	μ_{α}	0.934	0.946	0.955	0.950
	μ_{β}	0.939	0.949	0.964	0.947
	$\psi_{\alpha\alpha}$	0.844	0.921	0.935	0.937
	$\psi_{\beta\beta}$	0.914	0.947	0.956	0.954
H3	$\psi_{\alpha\beta}$	0.953	0.947	0.949	0.938
	μ_{α}	0.921	0.946	0.946	0.954
	μ_{β}	0.936	0.946	0.946	0.945
	$\psi_{\alpha\alpha}$	0.916	0.948	0.953	0.941
	$\psi_{\beta\beta}$	0.907	0.934	0.944	0.952
H4	$\psi_{\alpha\beta}$	0.911	0.949	0.942	0.949
	μ_{α}	0.921	0.948	0.945	0.948
	μ_{β}	0.937	0.949	0.964	0.947
	$\psi_{\alpha\alpha}$	0.910	0.948	0.955	0.937
	$\psi_{\beta\beta}$	0.917	0.948	0.958	0.953
H5	$\psi_{\alpha\beta}$	0.893	0.947	0.958	0.942
	μ_{α}	0.908	0.953	0.950	0.927
	μ_{β}	0.937	0.948	0.964	0.948
	$\psi_{\alpha\alpha}$	0.865	0.936	0.947	0.936
	$\psi_{\beta\beta}$	0.915	0.949	0.958	0.952

Tabela I.3: *Coverage* para os parâmetros, dos modelos H1 a H5, e $T = 5$

Modelo		Enviesamento		EQM		Coverage	
		$N = 250$	$N = 500$	$N = 250$	$N = 500$	$N = 250$	$N = 500$
H1	$\psi_{\alpha\beta}$	0.0061 (5.4555)	0 (0)	0.0202	0.0104	0.942	0.948
	μ_{α}	-0.0018 (-0.0060)	-0.0060 (0.0200)	0.0218	0.0102	0.942	0.957
	μ_{β}	0.0001 (0.0909)	0.0004 (0.3636)	0.0033	0.0015	0.947	0.949
	$\psi_{\alpha\alpha}$	0.0021 (-0.0013)	0.0145 (0.0088)	0.1731	0.0888	0.941	0.951
	$\psi_{\beta\beta}$	-0.0071 (-0.0354)	-0.0030 (-0.0150)	0.0058	0.0030	0.952	0.938
H2	$\psi_{\alpha\beta}$	0.0190 (0.0181)	0.0041 (0.0039)	0.1019	0.0507	0.936	0.939
	μ_{α}	0.0035 (0.0117)	-0.0036 (-0.0121)	0.0236	0.0113	0.947	0.951
	μ_{β}	-0.0060 (-0.0124)	-0.0071 (-0.0146)	0.0098	0.0047	0.948	0.952
	$\psi_{\alpha\alpha}$	-0.0200 (-0.0093)	-0.0086 (-0.0040)	0.3099	0.1544	0.905	0.935
	$\psi_{\beta\beta}$	-0.0269 (-0.0142)	-0.0232 (-0.0122)	0.0947	0.0465	0.933	0.943
H3	$\psi_{\alpha\beta}$	0.0214 (-4.6522)	0.0093 (-2.0217)	0.1259	0.0531	0.940	0.956
	μ_{α}	-0.0245 (-0.0135)	-0.0333 (-0.0183)	0.1150	0.0577	0.951	0.949
	μ_{β}	0.0016 (1.4545)	0.0015 (1.3636)	0.0039	0.0019	0.949	0.946
	$\psi_{\alpha\alpha}$	-0.2961 (-0.0119)	-0.2837 (-0.0114)	15.2216	7.2523	0.936	0.944
	$\psi_{\beta\beta}$	-0.0082 (-0.0410)	-0.0032 (-0.0160)	0.0074	0.0038	0.933	0.932
H4	$\psi_{\alpha\beta}$	-0.0476 (-0.0075)	-0.0683 (-0.0107)	0.9383	0.4462	0.949	0.941
	μ_{α}	-0.0183 (-0.0101)	-0.0297 (-0.0164)	0.1127	0.0557	0.950	0.947
	μ_{β}	-0.0047 (-0.0097)	-0.0064 (-0.0132)	0.0101	0.0048	0.949	0.945
	$\psi_{\alpha\alpha}$	-0.2133 (-0.0084)	-0.2589 (-0.0102)	16.5417	7.8694	0.932	0.941
	$\psi_{\beta\beta}$	-0.0248 (-0.0131)	-0.0225 (-0.0118)	0.3001	0.0454	0.933	0.948
H5	$\psi_{\alpha\beta}$	-0.0089 (-0.0014)	-0.0430 (-0.0068)	1.1340	0.5384	0.944	0.945
	μ_{α}	-0.0106 (-0.0059)	-0.0252 (-0.0139)	0.1303	0.0639	0.938	0.947
	μ_{β}	-0.0045 (-0.0093)	-0.0061 (-0.0126)	0.0103	0.0050	0.947	0.950
	$\psi_{\alpha\alpha}$	-0.2539 (-0.0088)	-0.3389 (-0.0117)	34.7171	16.4968	0.920	0.932
	$\psi_{\beta\beta}$	-0.0192 (-0.0101)	-0.0185 (-0.0097)	0.1040	0.0503	0.928	0.942

Tabela I.4: Enviesamento bruto(enviesamento relativo), erro quadrático médio e *coverage* para os parâmetros, dos modelos H1 e H5, com $T = 5$ e omissões