



Metodologias de cálculo para estimação das provisões para sinistros em anos de ocorrência futuros

Ana Domingues

Metodologias de cálculo para estimação das provisões para sinistros em anos de ocorrência futuros

Ana Domingues

Este trabalho foi submetido a provas públicas no dia 15 de dezembro de 2023 para obtenção do grau de Mestre em Matemática Atuarial

Sob orientação de:

Professora Maria de Lourdes Belchior Afonso Professora Associada, Universidade NOVA de Lisboa Professor Pedro Alexandre da Rosa Corte Real Professor Auxiliar, Universidade NOVA de Lisboa

RESUMO

O cálculo da provisão para sinistros no ramo não-vida é de extrema relevância para a avaliação da rentabilidade de uma seguradora. O presente estudo pretende calcular provisões para sinistros que irão ocorrer futuramente, usando três modelos: o Loglinear, o Poisson sobre-dispersão e o *Bootstrap* recorrendo ao *software* estatístico R. A metodologia aplicada teve por base a construção de uma nova diagonal do triângulo de desenvolvimento a cada iteração.

Os dados utilizados foram extraídos do *website* da Autoridade de Supervisão de Seguros e Fundos de Pensões relativos ao ramo automóvel e aos períodos de 2001 a 2010 e de 2006 a 2015, para perceção do impacto do uso de várias iterações na construção das novas diagonais do triângulo de desenvolvimento. Também foi considerada a inflação passada e futura como variável neste estudo, de modo a perceber o impacto da incorporação nos dados históricos e nas previsões futuras.

Neste trabalho foram calculadas estimativas da provisão para sinistros que irão ocorrer no futuro por ano de calendário e foram comparados os resultados destas estimativas com os valores reais entre 2016 a 2019. Também se comparou os resultados das estimativas de 2020 a 2028, inferindo o impacto da incorporação da inflação, do modelo utilizado e dos dados utilizados.

Palavras-Chave: Provisões para sinistros, Modelo LogLinear, Modelo de Poisson sobredispersão, *Bootstrap*

ABSTRACT

Claims reserve's computation of non-life insurance is extremely important for assessing the profitability of an insurance company. This thesis proposes a computation of claims reserve for claims that will occur in the future, using three models: Loglinear, Over Dispersed Poisson and Bootstrap using a free software environment for statistical computing, called R. The applied methodology was based on the construction of a new diagonal of the run-off triangle at each iteration.

The data used for the calculations were extracted from the website of the Portuguese Insurance and Pension Funds Supervisory Authority (Autoridade de Supervisão de Seguros e Fundos de Pensões) regarding car insurance. It was used two sets of data, one with the payments from 2001 to 2010 and other with payments from 2006 to 2015, to measure the impact of the use of several iterations in the construction of the new diagonals of the run-off triangle. Past and future inflation was also considered as a variable in this study to understand the impact of it in results.

In this thesis, there were made claims reserve calculations by calendar year for future claims that will occur. These calculations were compared to the actual payments during the period from 2016 to 2019. The claims reserve results from 2020 to 2028 were also compared and gathering the impact of inflation incorporation, the model used and the data used.

Keywords: Claim Reserving, LogLinear, Over-Dispersed Poisson, Bootstrap

Conteúdo

Li	sta de	Tabela	as .	Vi					
Li	st de l	Figuras		ix					
1	Intro	odução		1					
2	Prov	isões p	ara sinistros	3					
	2.1	Conte	xtualização	3					
	2.2	Conce	itos	6					
		2.2.1	Triângulo de desenvolvimento	6					
		2.2.2	Inflação	8					
		2.2.3	Estimativas e intervalos de confiança	9					
	2.3	Metod	lologias	14					
		2.3.1	Modelo Loglinear	14					
		2.3.2	Modelo de Poisson sobre-dispersão	20					
		2.3.3	Bootstrap	26					
3	Resu	ıltados		30					
	3.1	Model	lo Loglinear	35					
	3.2	Model	lo de Poisson sobre-dispersão	47					
	3.3	Bootst	trap	59					
4	Con	clusão		68					
Bi	bliogr	afia		73					
A	Anex	xos		74					
	A.1	Verific	cação dos pressupostos do modelo Loglinear	74					
	A.2	A.2 Verificação dos pressupostos do modelo de Poisson sobre-dispersão							

Lista de Tabelas

2.1	Triângulo (ou matriz) de desenvolvimento	6
2.2	Triângulo (ou matriz) de desenvolvimento estimado	7
2.3	Matriz de desenvolvimento a estimar	8
3.1	Dados das indemnizações do ramo automóvel desde 2001 a 2010 (u.m.: EUR).	30
3.2	Dados das indemnizações do ramo automóvel desde 2006 a 2015 (u.m.: EUR).	31
3.3	Resultado do ajustamento do modelo Loglinear aplicado aos dados incrementais	
	de 2006 a 2015 com inclusão da inflação	37
3.4	Resultados da primeira iteração da metodologia para o ano de calendário de 2016	
	com a aplicação do modelo Loglinear aos dados de 2006 a 2015 considerando a	
	inflação (u.m.: EUR)	39
3.5	Triângulo (ou matriz) de desenvolvimento com a inclusão dos resultados da	
	primeira iteração da aplicação do modelo Loglinear aos dados relativos ao período	
	2006-2015, considerando a inflação (u.m.: EUR)	39
3.6	Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do	
	modelo Loglinear aos dados relativos ao período 2006-2015, considerando a	
	inflação (u.m.: EUR)	40
3.7	Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do	
	modelo Loglinear aos dados relativos ao período 2006-2015, não incorporando a	
	inflação (u.m.: EUR)	41
3.8	Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do	
	modelo Loglinear aos dados relativos ao período 2001-2010, considerando a	
	inflação (u.m.: EUR)	42
3.9	Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do	
	modelo Loglinear aos dados relativos ao período 2001-2010, não incorporando a	
	a inflação (u.m.: EUR)	43
3.10	Comparação dos resultados obtidos através do modelo Loglinear e os valores	
	reais das indemnizações relativas ao período de 2016 a 2019 por ano de calendário.	45
3.11	1	
	2006 a 2015 com inclusão da inflação	47

3.12	Resultados da primeira iteração da metodologia para o ano de calendário de	
	2016 com a aplicação do modelo ODP aos dados de 2006 a 2015 considerando a	
	inflação (u.m.: EUR)	49
3.13	Triângulo (ou matriz) de desenvolvimento com a inclusão dos resultados da	
	primeira iteração da aplicação do modelo ODP aos dados relativos ao período	
	2006-2015, considerando a inflação (u.m.: EUR)	50
3.14	Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do	
	modelo ODP aos dados relativos ao período 2006-2015, considerando a inflação	
	(u.m.: EUR)	51
3.15	Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do	
	modelo ODP aos dados relativos ao período 2006-2015, não incorporando a	
	inflação (u.m.: EUR)	52
3.16	Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do	
	modelo ODP aos dados relativos ao período 2001-2010, considerando a inflação	
	(u.m.: EUR)	53
3.17	Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do	
	modelo ODP aos dados relativos ao período 2001-2010, não incorporando a	
	inflação (u.m.: EUR)	54
3.18	Comparação dos resultados obtidos através do modelo ODP e os valores reais	
	das indemnizações relativas ao período de 2016 a 2019 por ano de calendário	56
3.19	Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do	
	modelo Loglinear e Bootstrap aos dados relativos ao período 2006-2015, consi-	
	derando a inflação (u.m.: EUR).	60
3.20	Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do	
	modelo Loglinear e <i>Bootstrap</i> aos dados relativos ao período 2006-2015, não	
	incorporando a inflação (u.m.: EUR).	61
3.21	Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do	
	modelo Loglinear e Bootstrap aos dados relativos ao período 2001-2010, consi-	
	derando a inflação (u.m.: EUR).	62
3.22	Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do	
	modelo Loglinear e <i>Bootstrap</i> aos dados relativos ao período 2001-2010, não	
	incorporando a inflação (u.m.: EUR).	63
3.23	Resultados da provisão para sinistros por ano de calendário para os anos de 2020	
	a 2024, aplicando os três modelos (u.m.: EUR)	64
3.24	Resultados da provisão para sinistros por ano de calendário para os anos de 2025	
	a 2028, aplicando os três modelos (u.m.: EUR)	65

A.1	Verificação dos pressupostos para aplicação do modelo Loglinear	74
A.2	Verificação dos pressupostos para aplicação do modelo de Poisson sobre-dispersão.	81

Lista de Figuras

2.1	Representação temporal de um processo de sinistro (adaptado de Taylor [1])	3
2.2	Evolução do Índice de Preços no Consumidor (IPC) desde 2000 até 2022	9
2.3	Processo de estimação das provisões para sinistros em anos de ocorrência futuros.	14
2.4	Processo de estimação das provisões para sinistros pelo modelo Loglinear [2]	15
2.5	Técnica de <i>Bootstrap</i> aplicada às provisões para sinistros [3]	27
3.1	Pagamentos incrementais e acumulados entre 2001 e 2010 sem considerar a	
	inflação (u.m.: EUR)	32
3.2	Pagamentos incrementais e acumulados entre 2006 e 2015 sem considerar a	
	inflação (u.m.: EUR)	32
3.3	Pagamentos incrementais e acumulados entre 2001 e 2010 considerando a infla-	
	ção (u.m.: EUR)	33
3.4	Pagamentos incrementais e acumulados entre 2006 e 2015 considerando a infla-	
	ção (u.m.: EUR)	33
3.5	Previsão da taxa de inflação em Portugal no período de 2023 a 2027 [4]	34
3.6	Distribuição por ano de origem dos dados das indemnizações relativas ao período	
	de 2006 a 2015 considerando a inflação	35
3.7	Distribuição por ano de desenvolvimento dos dados das indemnizações relativas	
	ao período de 2006 a 2015 considerando a inflação	35
3.8	Distribuição por ano de calendário dos dados das indemnizações relativas ao	
	período de 2006 a 2015 considerando a inflação	36
3.9	Gráfico dos resíduos do modelo por ano de origem	38
3.10	Gráfico dos resíduos do modelo por ano de desenvolvimento	38
3.11	Gráfico dos resíduos do modelo por ano de calendário	38
3.12	Gráfico dos resíduos <i>versus</i> valores modelizados	38
3.13	Gráfico quantil-quantil dos resíduos do modelo Loglinear aplicado aos dados	
	das indemnizações relativas ao período de 2006 a 2015 considerando a inflação.	38
3.14	Gráficos comparativos entre os valores de pagamento reais e os valores estimados	
	pelo modelo Loglinear no período de 2016 a 2019 por ano de desenvolvimento	
	do sinistro (u.m.: EUR).	46

3.15	Gráfico dos resíduos do modelo por ano de origem	48
3.16	Gráfico dos resíduos do modelo por ano de desenvolvimento	48
3.17	Gráfico dos resíduos do modelo por ano de calendário.	48
3.18	Gráfico dos resíduos <i>versus</i> valores modelizados	48
3.19	Gráfico quantil-quantil dos residuos do modelo ODP aplicado aos dados das	
	indemnizações relativas ao período de 2006 a 2015 considerando a inflação	49
3.20	Gráficos comparativos entre os valores de pagamento reais e os valores estimados	
	pelo modelo ODP no período de 2016 a 2019 por ano de desenvolvimento do	
	sinistro (u.m.: EUR).	57
3.21	Gráficos comparativos entre os valores de pagamento reais e os valores esti-	
	mados pelos modelos Loglinear e ODP no período de 2016 a 2019 por ano de	
	desenvolvimento do sinistro (u.m.:EUR).	58
3.22	Análise gráfica dos resíduos de um triângulo obtido por simulação de <i>Bootstrap</i>	
	em função do ano de origem do sinistro	59
3.23	Análise gráfica dos resíduos de um triângulo obtido por simulação de <i>Bootstrap</i>	
	em função do ano de desenvolvimento do sinistro	59
3.24	Gráficos comparativos entre as estimativas obtidas pelos três modelos por ano	
	de calendário (u.m.: EUR)	67
A.1	Análise de tendência por parte dos resíduos das várias iterações aplicando o mo-	
	delo Loglinear aos dados do triângulo de desenvolvimento com as indemnizações	
	de 2006 a 2015 e considerando a inflação	75
A.2	Análise de tendência por parte dos resíduos das várias iterações aplicando o mo-	
	delo Loglinear aos dados do triângulo de desenvolvimento com as indemnizações	
	de 2006 a 2015 e sem incorporar a inflação	76
A.3	Análise de tendência por parte dos resíduos das várias iterações aplicando o mo-	
	delo Loglinear aos dados do triângulo de desenvolvimento com as indemnizações	
	de 2001 a 2010 e considerando a inflação. Iterações de 1 a 5	77
A.4	Análise de tendência por parte dos resíduos das várias iterações aplicando o mo-	
	delo Loglinear aos dados do triângulo de desenvolvimento com as indemnizações	
	de 2001 a 2010 e considerando a inflação. Iterações de 6 a 9	78
A.5	Análise de tendência por parte dos resíduos das várias iterações aplicando o mo-	
	delo Loglinear aos dados do triângulo de desenvolvimento com as indemnizações	
	de 2001 a 2010 e sem incorporar a inflação. Iterações de 1 a 5	79

A.6	Análise de tendência por parte dos resíduos das várias iterações aplicando o mo-	
	delo Loglinear aos dados do triângulo de desenvolvimento com as indemnizações	
	de 2001 a 2010 e sem incorporar a inflação. Iterações de 6 a 9	80
A.7	Análise de tendência por parte dos resíduos das várias iterações aplicando o	
	modelo ODP aos dados do triângulo de desenvolvimento com as indemnizações	
	de 2006 a 2015 e considerando a inflação	82
A.8	Análise de tendência por parte dos resíduos das várias iterações aplicando o	
	modelo ODP aos dados do triângulo de desenvolvimento com as indemnizações	
	de 2006 a 2015 e sem incorporar a inflação	83
A.9	Análise de tendência por parte dos resíduos das várias iterações aplicando o	
	modelo ODP aos dados do triângulo de desenvolvimento com as indemnizações	
	de 2001 a 2010 e considerando a inflação. Iterações de 1 a 5	84
A.10	Análise de tendência por parte dos resíduos das várias iterações aplicando o	
	modelo ODP aos dados do triângulo de desenvolvimento com as indemnizações	
	de 2001 a 2010 e considerando a inflação. Iterações de 6 a 9	85
A. 11	Análise de tendência por parte dos resíduos das várias iterações aplicando o	
	modelo ODP aos dados do triângulo de desenvolvimento com as indemnizações	
	de 2001 a 2010 e sem incorporar a inflação. Iterações de 1 a 5	86
A.12	Análise de tendência por parte dos resíduos das várias iterações aplicando o	
	modelo ODP aos dados do triângulo de desenvolvimento com as indemnizações	
	de 2001 a 2010 e sem incorporar a inflação. Iterações de 6 a 9	87

1. Introdução

O negócio segurador está assente na determinação da probabilidade de determinado evento ocorrer e no seu custo expectável, pelo que para o ramo não-vida o cálculo das provisões técnicas é de extrema relevância para que a seguradora se mantenha solvente com uma elevada probabilidade, cumprindo o compromisso com os segurados e terceiros. Para que haja equilíbrio entre competitividade da empresa e a sua rentabilidade/solvência é necessário que o cálculo das provisões técnicas correspondam o mais possível à realidade [2].

Das provisões técnicas do ramo não-vida, as provisões para sinistros são as que mais se destacam, pelo seu peso e relevância. Estas provisões refletem o custo que a empresa de seguros terá de suportar para a regularização de todos os sinistros já ocorridos (reportados ou não à seguradora) e que ainda não tenham sido objeto de pagamento [2].

No entanto, tendo em conta as exigências do mercado, nomeadamente com o regime de Solvência II, as seguradoras preocupam-se com a continuidade do negócio, a sua sustentabilidade e rentabilidade. Sendo assim, as seguradoras devem efetuar uma análise de sustentabilidade dos ramos não-vida que têm em carteira durante um determinado horizonte temporal.

As metodologias mais comuns para a estimação das provisões para sinistros são os modelos estocásticos, obtendo-se estimativas da provisão a constituir e intervalos de confiança para essa mesma estimativa. Dos modelos mais conhecidos, existem o Modelo de Thomas Mack, o Modelo Loglinear, os Modelos Lineares Generalizados ou a técnica de *Bootstrap*.

Neste trabalho, pretende-se apresentar a aplicação de algumas metodologias para a projeção das provisões para sinistros a constituir para anos de ocorrência futuros, calculando, iterativamente, as estimativas das diagonais dos triângulos de desenvolvimento até uma pré determinada data futura e posteriormente calculando a parte inferior do triângulo de desenvolvimento. Foram usados dois triângulos de desenvolvimento relativos ao ramo automóvel em Portugal: um com os dados das indemnizações pagas no período de 2001 a 2010 e outro com as indemnizações pagas entre 2006 e 2015, extraídos do *website* da Autoridade de Supervisão de Seguros e Fundos de Pensões. Ao usar diferentes triângulos para efetuar a estimação da provisão para sinistros para a mesma data futura, o número de iterações a usar varia, pelo que também se consegue inferir o impacto de efetuar mais iterações do mesmo modelo. Para determinar o impacto da incorporação da inflação passada e futura na provisão para sinistros, para além dos dados referidos anteriormente, também foram considerados os mesmos dados mas incorporando a inflação passada e na

modelação foi considerada a inflação futura.

Neste sentido, a presente dissertação apresenta-se dividida em quatro capítulos, com a seguinte estrutura: O segundo capítulo abordará aspetos gerais relativos à provisão para sinistros, nomeadamente ao processo de um sinistro e aos triângulos de desenvolvimento, fazendo uma introdução aos conceitos e notações utilizados ao longo deste documento. Ainda neste capítulo, apresenta-se a forma como foi abordada a problemática proposta e descrevem-se as metodologias Loglinear, o modelo linear generalizado de Poisson sobre-dispersão e a metodologia de *Bootstrap* aplicadas ao cálculo das provisões para sinistros. No capítulo seguinte apresenta-se a aplicação prática dos métodos referidos considerando os dois conjuntos de dados, tendo sido considerada a incorporação da inflação ou a sua desconsideração, recorrendo ao uso do software *R* [5]. Por fim, no último capítulo, encontram-se as conclusões da análise dos dados e a comparação entre as várias metodologias aplicadas. Também se propõem trabalhos futuros nesta matéria.

2. Provisões para sinistros

2.1 Contextualização

O processo de sinistro tem uma duração muito variada, dependendo do ramo e da natureza do sinistro. Os sinistros ocorridos num ramo não-vida podem ser comunicados com atraso face à data de ocorrência, o que consequentemente também atrasa o pagamento das indemnizações. Ainda existem alguns casos em que o processo de indemnização se arrasta por envolver decisões do tribunal, por exemplo. Veja-se a figura seguinte que representa a linha temporal de um processo de sinistro.

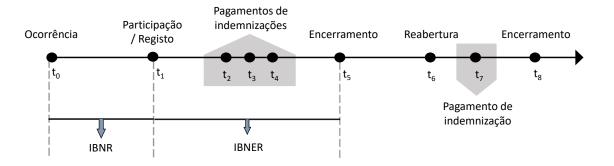


Figura 2.1: Representação temporal de um processo de sinistro (adaptado de Taylor [1]).

Na figura 2.1, o instante t_0 representa o momento em que ocorre o evento que gera o sinistro. No entanto, a seguradora apenas tem conhecimento do facto no momento t_1 e efetua o(s) pagamento(s) depois de todos os processos administrativos, e outros, envolvidos. Os pagamentos podem ser efetuados em vários momentos até que que a seguradora considere que não há lugar a mais pagamentos e encerre o processo (no momento t_5). Posteriormente, o processo pode ter de ser reaberto devido a uma revisão ou uma decisão judicial (no momento t_6) e mais pagamentos podem ocorrer até que o processo seja novamente encerrado (em t_8).

Podemos ainda observar que na figura 2.1 se encontram dois períodos com a designação IBNR e IBNER, entre os instantes t_0 e t_1 e entre t_1 e t_5 , respetivamente. A primeira sigla referese aos sinistros que já ocorreram mas ainda não são do conhecimento da seguradora porque ainda não foram reportados (em inglês, *Incurred But Not Reported*) e a segunda sigla refere-se

aos sinistros que já foram reportados mas a seguradora ainda está a avaliar o valor que deve provisionar para o pagamento das indemnizações (em inglês, *Incurred But Not Enough Reported*). Numa seguradora é fundamental estimar o valor total de indemnizações a pagar tanto de um sinistro *IBNR* como de um sinistro *IBNER*. A este processo de estimação dos pagamentos futuros relativamente a estes sinistros, chama-se de provisões para sinistros [6].

O principal objetivo de um exercício de cálculo de provisão para sinistros é determinar uma estimativa do valor monetário que a seguradora deve ter reservado para pagar os sinistros relativos às apólices subscritas. Esta estimativa tem uma incerteza associada, sendo mais previsível se uma determinada área de negócio for mais bem comportada em termos estatísticos ou se a cauda de pagamentos for mais curta.

As provisões para sinistros têm, na maioria das vezes, o maior peso nos passivos de uma seguradora, tendo impacto tanto a nível de lucro como de solvência da seguradora [7]. Através destas provisões é possível [6]:

- avaliar a condição financeira da seguradora;
- determinar o valor de mercado da seguradora (especialmente em caso de venda ou aquisição);
- avaliar a capacidade da seguradora conseguir assumir os seus passivos;
- fazer o apreçamento da linha de negócio (por extrapolação dos pagamentos e provisões passadas);
- determinar o valor de *run-off* de uma carteira de uma determinada linha de negócio.

O método mais conhecido para a estimação das provisões para sinistros é o método *chain ladder* uma vez que é um método muito simples e não depende de nenhuma distribuição de probabilidade. No entanto, o valor de indemnizações que a seguradora terá de pagar no futuro é uma variável aleatória, pelo que incorporar um modelo probabilístico torna-se fundamental tanto para o cálculo da estimativa como para o erro associado a esta estimativa [8]. Algumas das primeiras publicações que consideravam o erro da estimativa das provisões para sinistros utilizavam a abordagem da regressão dos mínimos quadrados aplicada aos logaritmos dos montantes dos sinistros incrementais: Zehnwirth [9], Renshaw [10], Christofides [11], Verrall [12] ou Verrall [13]. Wright [14] apresentou outra proposta, utilizando os modelos lineares generalizados (cuja sigla em inglês é GLM - *generalized linear model*) e Mack [15] utilizou a distribuição Gama e o estimador de máxima verosimilhança. Este mesmo autor, em 1993, publicou um outro artigo com uma formula mais simples de calcular o erro das provisões para sinistros pelo método de *chain ladder* que não implicava calcular as várias covariâncias entre os estimadores dos parâmetros [8]. England and Verrall [16, 17] introduziram a técnica de *bootstraping* para o método de *chain ladder* onde o erro da estimação é calculado através de

simulações [18]. Mais alguns autores focaram-se nos modelos de GLM para efetuar a estimação das provisões para sinistros: Despeyroux et al. [19], England and Verrall [20] e Taylor and McGuire [21]. Alguns destes autores focaram-se nos dados agregados dos montantes dos sinistros, sejam eles incrementais ou acumulados, e outros autores aplicaram as metodologias aos dados dos sinistros individuais.

Mais recentemente, outras metodologias têm vindo a ser estudadas para a estimação das provisões para sinistros, como por exemplo, a aplicação de metodologias de *machine learning* que se baseiam em dados dos sinistros individuais [22]. Uma vez que neste trabalho utilizaramse os dados da indústria seguradora portuguesa reportado pela Autoridade de Supervisão de Seguros e Fundos de Pensões (ASF), não se tem os dados individuais, pelo que apenas serão implementadas metodologias para dados agrupados. Mais detalhes acerca das metodologias utilizadas na elaboração deste trabalho podem ser consultados na secção 2.3.

Apesar da vasta literatura relativamente à estimação da provisão para sinistros, existe pouca informação para o cálculo desta provisão em anos de ocorrência futuros, uma vez que na maioria dos casos a aplicação das metodologias apenas estima as provisões para o futuro mas relativos a sinistros já ocorridos (que podem ainda não ser do conhecimento da seguradora). O cálculo da provisão para sinistros em anos de ocorrência futuros tem especial relevância para a análise de rentabilidade de um qualquer ramo da área de não-vida. A elaboração deste trabalho pretende aprofundar esta temática de estimação das provisões em anos de ocorrência futuros desde o ano de origem do sinistro e fazer comparações entre algumas metodologias mais comuns.

2.2 Conceitos

2.2.1 Triângulo de desenvolvimento

Para efetuar o cálculo da provisão para sinistros é usual agrupar os dados históricos em classes homogéneas considerando o período de origem e o período do desenvolvimento e distribuindo a informação agregada na forma triangular, a que se chama de triângulo (ou matriz) de desenvolvimento ou run-off. Cada linha representa, na maioria das vezes o ano de ocorrência do sinistro, mas poderá ser o ano de subscrição ou de comunicação bem como pode ser considerada outra unidade de tempo. Ao longo das colunas encontram-se os períodos de desenvolvimento (em anos ou outra unidade), que em função da definição do período de origem, poderá ser o ano de acidente, o ano de comunicação ou de pagamento (tendo em conta o processo de sinistro explicado em 2.1, um sinistro ocorrido em determinado ano i, pode ter pagamentos em j, j+1, j+2, ..., o que corresponde a pagamentos no ano de desenvolvimento 0, 1, 2, ..., respetivamente) [6]. Para ilustrar o triângulo de desenvolvimento veja-se a tabela 2.1.

Período de desenvolvimento 0 1 2 J-2 J J-1 Período de ocorrência $X_{0,2}$... $X_{0,J-2}$ $X_{0,J-1}$ $X_{0,J}$ $X_{0,\infty}$ 0 $X_{0,0}$ $X_{1,1}$ $X_{1,2}$... $X_{1,J-2}$ $X_{1,J-1}$ 1 $X_{2,1}$ $X_{2,2}$... $X_{2,J-2}$ 2 $X_{I-1,0} \quad X_{I-1,1}$ I-1 I

Tabela 2.1: Triângulo (ou matriz) de desenvolvimento

No exemplo do triângulo de desenvolvimento presente na tabela 2.1, os valores $X_{i,j}$, onde $i=0,1,\ldots,I$ e $j=0,1,\ldots,J$, podem referir-se a [23][2]:

- número de sinistros comunicados;
- número de sinistros encerrados;
- pagamentos efetuados;
- indemnizações emitidas;
- pagamentos efetuados acrescidos das provisões.

O valor $X_{0,\infty}$, que se denomina de *ultimate* é relativo às indemnizações de sinistros ocorridos no período 0 e que irão ser pagas após o ano J. Neste trabalho não irá ser incluído este período

ultimate, mas podem ser consultados trabalhos já desenvolvidos de modo a efetuar a sua inclusão, como por exemplo: Mack [24], Boor [25], Merz and Wüthrich [26] ou Evans [27].

O triângulo de desenvolvimento pode considerar os valores incrementais, tendo em conta apenas os valores respeitantes ao período de desenvolvimento para cada período de ocorrência, ou pode considerar os valores acumulados, sendo que neste caso os valores apresentados correspondem à soma dos valores incrementais até ao final de cada período de desenvolvimento para cada ano de ocorrência. Por exemplo, considerando que os valores $X_{i,j}$, onde $i=0,1,\ldots,I$ e $j=0,1,\ldots,J$, do triângulo de desenvolvimento anterior, o triângulo de desenvolvimento com os valores acumulados seria obtido através da equação (2.1):

$$C_{i,j} = \sum_{k=0}^{J-i} X_{i,k} \qquad 0 \le i \le I$$
 (2.1)

O cálculo da provisão para sinistros permite que se consiga completar o triângulo de desenvolvimento, obtendo os valores de forma genérica $\hat{X}_{i,j}$, onde $i=0,1,\ldots,I$ e $j=1,2,\ldots,J,\infty$, da seguinte forma apresentada pela tabela 2.2.

Período de desenvolvimento 0 J-2 J-1 ∞ Período de ocorrência $\hat{X}_{0,\infty}$ $\mathbf{0}$ $\hat{X}_{1,J}$ $\hat{X}_{1,\infty}$ 1 $\hat{X}_{2,J-1}$ $\hat{X}_{2,J}$ $\hat{X}_{2,\infty}$ 2 $\hat{X}_{I-1,2}$... $\hat{X}_{I-1,J-2}$ $\hat{X}_{I-1,J-1}$ $\hat{X}_{I-1,J}$ $\hat{X}_{I-1,\infty}$ $\hat{X}_{I,1}$ $\hat{X}_{I,2}$... $\hat{X}_{I,J-2}$ $\hat{X}_{I,J-1}$ $\hat{X}_{I,J}$ $\hat{X}_{I,\infty}$

I-1 I

Tabela 2.2: Triângulo (ou matriz) de desenvolvimento estimado

No entanto, o problema em questão neste trabalho é determinar uma estimativa para os valores dos sinistros futuros, o que implica a continuação da construção da matriz de desenvolvimento para as linhas inferiores, n períodos para além do ano de exercício (o último conhecido) como mostra a tabela 2.3, na zona sombreada. Tal como referido anteriormente, o cálculo da provisão para a coluna relativa ao *ultimate* ficará de fora do âmbito deste trabalho, razão pela qual na tabela se encontra a divisão em tracejado da matriz de desenvolvimento.

Tabela 2.3: Matriz de desenvolvimento a estimar

Período de desenvolvimento Período de ocorrência	0	1	2	 J-2	J-1	J	∞
0							$\hat{X}_{0,\infty}$
1						$\hat{X}_{1,J}$	$\hat{X}_{1,\infty}$
2					$\hat{X}_{2,J-1}$	$\hat{X}_{2,J}$	$\hat{X}_{2,\infty}$
<u>I-1</u>			$\hat{X}_{I-1,2}$	 $\hat{X}_{I-1,J-2}$	$\hat{X}_{I-1,J-1}$	$\hat{X}_{I-1,J}$	$\hat{X}_{I-1,\infty}$
I		$\hat{X}_{I,1}$	$\hat{X}_{I,2}$	 $\hat{X}_{I,J-2}$	$\hat{X}_{I,J-1}$	$\hat{X}_{I,J}$	$\hat{X}_{I,\infty}$
I+1	$\hat{X}_{I+1,0}$	$\hat{X}_{I+1,1}$	$\hat{X}_{I+1,2}$	 $\hat{X}_{I+1,J-2}$	$\hat{X}_{I+1,J-1}$	$\hat{X}_{I+1,J}$	$\hat{X}_{I+1,\infty}$
I+n	$\hat{X}_{I+n,0}$	$\hat{X}_{I+n,1}$	$\hat{X}_{I+n,2}$	 $\hat{X}_{I+n,J-2}$	$\hat{X}_{I+n,J-1}$	$\hat{X}_{I+n,J}$	$\hat{X}_{I+n,\infty}$

2.2.2 Inflação

A inflação influencia os valores das indemnizações e por essa razão a sua evolução deve ser tida em conta no cálculo da provisão para sinistros, especialmente se existirem períodos de grandes oscilações nos preços. A não inclusão da inflação nas estimativas será razoável caso esta se mantenha estável tanto no passado como expectavelmente no futuro.

Na figura 2.2 está representada a evolução do índice de Preços no Consumidor (IPC)¹ em Portugal [28]. Ao analisar a figura, verifica-se que existiram oscilações abruptas entre 2008 e 2013 e a partir de 2021 e ainda se verificam períodos onde existiu inflação negativa, nomeadamente nos anos de 2009 e 2014, que correspondem ao ano seguinte à falência do banco de investimento *Lehman Brothers Holdings Inc.*, que provocou uma recessão económica por todo o mundo, e ao fim do Programa de Assistência Económica e Financeira a Portugal (PAEF)², respetivamente. O ano de 2020 apresenta inflação zero, e coincide com o ano de declaração da doença COVID-19 como uma pandemia pela Organização Mundial da Saúde.

Neste sentido é necessário efetuar alguns ajustamentos ao triângulo de desenvolvimento de modo a que todas as indemnizações estejam a preços do ano de avaliação. Os valores obtidos pela metodologia escolhida deverão depois ser ajustados para incorporar a inflação futura prevista. Ou seja, o triângulo (incremental) de desenvolvimento³ a considerar deverá ter incorporado na sua

¹Indicador mais usual para a inflação.

²Monitorizado por elementos da Comissão Europeia, do Banco Central Europeu (BCE) e do Fundo Monetário Internacional (FMI).

³Na sua forma genérica dado pela tabela 2.1.

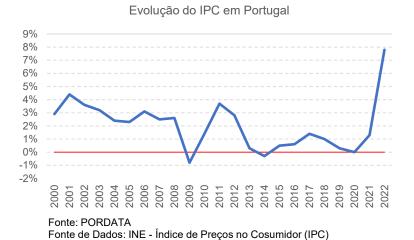


Figura 2.2: Evolução do Índice de Preços no Consumidor (IPC) desde 2000 até 2022.

diagonal a inflação desde o momento 0 (período de ocorrência e desenvolvimento 0) até ao último período de ocorrência conhecido. Em termos matemáticos, o triângulo de desenvolvimento será o constituído pelos valores $Y_{i,j}$, onde $i=0,1,\ldots,I-1$ e $j=0,1,\ldots,J-i-1$, obtidos pela equação (2.2):

$$Y_{i,j} = X_{i,j} \prod_{k=i+j}^{J-1} (1 + \delta_k)$$
(2.2)

onde $0 \le i \le I-1, 0 \le j \le J-i-1$ e δ_k representa a inflação passada do ano k.

Para incorporar a inflação futura aos preços do respetivo ano de exercício (soma do ano de ocorrência com o ano de desenvolvimento), é necessário atualizar cada diagonal do triângulo estimado, presente na tabela 2.3, com a inflação desde o último período conhecido até ao último ano que se pretende estimar, ou seja, a estimativa $\hat{X}_{i,j}$, onde $i=1,\ldots,I+n$ e $j=\max(0,J-i+1),\ldots,J$, é atualizada de acordo com a equação (2.3):

$$\hat{Y}_{i,j} = \hat{X}_{i,j} \prod_{k=I+1}^{i+j} (1 + \delta_{k-1})$$
(2.3)

onde $1 \le i \le I + n$, $\max(0, I - i + 1) \le j \le J$, δ_k representa a inflação prevista para o ano k e n o número de períodos de ocorrência a estimar.

2.2.3 Estimativas e intervalos de confiança

Tendo o valor das estimativas para cada ano de ocorrência e ano de desenvolvimento, o valor as provisões para sinistros é calculado pela soma dos valores individuais, ou seja, para

cada ano de ocorrência a provisão é dada pela soma dos valores da respetiva linha na matriz de desenvolvimento, para cada ano de calendário a provisão é dada pela soma da respetiva diagonal na matriz de desenvolvimento e para a provisão total a provisão é dada pela soma de todos os valores estimados. Reescrevendo em forma matemática, obtemos as seguintes equações (2.4), (2.5) e (2.6), respetivamente.

O valor das provisões por ano de ocorrência i, \hat{R}_i^O , é dado por

$$\hat{R}_{i}^{O} = \begin{cases} \sum_{j=I-i+1}^{J} \hat{Y}_{i,j} & \text{se} \quad 1 \leq i \leq I\\ \sum_{j=0}^{J} \hat{Y}_{i,j} & \text{se} \quad I < i \leq I+n \end{cases}$$
 (2.4)

O valor das provisões por ano de calendário k, \hat{R}_k^C é dado por

$$\hat{R}_k^C = \sum_{j=\max(0,k-I-n)}^J \hat{Y}_{k-j,j} \quad I+1 \le k \le I+n+J$$
 (2.5)

O valor das provisões por totais, \hat{R}^T é dado por

$$\hat{R}^T = \sum_{i=1}^{I+n} \hat{R}_i^O \tag{2.6}$$

Sendo estimativas, devemos aferir da sua qualidade, nomeadamente, medir a precisão dessa estimativa recorrendo ao seu intervalo de confiança. No entanto, na maioria das vezes desconhece-se a distribuição da variável aleatória da provisão total [2]. Na prática, assume-se, por conveniência, que a provisão total R, segue uma distribuição Normal de valor médio μ e desvio-padrão σ , sendo $\mu = \hat{R}$ e $\sigma^2 = \widehat{EQM}(\hat{R})^4$, onde a sigla EQM significa Erro Quadrático Médio⁵, ou seja, $R \sim N(\hat{R}, \widehat{EQM}(\hat{R}))$. Do mesmo modo, aplicamos para a provisão por ano de ocorrência e por ano de calendário: $R_i \sim N(\hat{R}_i, \widehat{EQM}(\hat{R}_i))$ e $R_k \sim N(\hat{R}_k, \widehat{EQM}(\hat{R}_k))$.

A partir das distribuições das várias provisões, consegue-se construir o intervalo de confiança $1-\alpha$ para estas. Ou seja, os intervalos de confiança para as provisões totais, por ano de ocorrência ou por ano de calendário, são dados através das expressões (2.7), (2.8) e (2.9), respetivamente.

 $^{{}^4\}hat{R}$ e $\widehat{EQM}(\hat{R})$ são as estimativas da provisão total e do seu Erro Quadrático Médio, respetivamente, dadas pela aplicação do modelo estocástico.

⁵O EQM é a média dos quadrados dos erros, ou seja, é a diferença quadrática média entre os valores estimados pelo modelo e os valores observados.

O intervalo de confiança para as provisões totais é dado por:

$$\left[\hat{R}^T - \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\widehat{EQM}(\hat{R}^T)} \quad ; \quad \hat{R}^T + \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\widehat{EQM}(\hat{R}^T)}\right] \tag{2.7}$$

O intervalo de confiança por ano de ocorrência é determinado pela equação seguinte:

$$\left[\hat{R}_{i}^{O} - \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\widehat{EQM}(\hat{R}_{i}^{O})} \quad ; \quad \hat{R}_{i}^{O} + \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\widehat{EQM}(\hat{R}_{i}^{O})}\right] \tag{2.8}$$

O intervalo de confiança por ano de calendário é dado através da seguinte expressão:

$$\left[\hat{R}_k^C - \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\widehat{EQM}(\hat{R}_k^C)} \quad ; \quad \hat{R}_k^C + \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\widehat{EQM}(\hat{R}_k^C)}\right] \tag{2.9}$$

sendo que $\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)$ representa o quantil de probabilidade $1-\frac{\alpha}{2}$ da distribuição Normal Padrão, $1\leq i\leq I+n$ e $I+1\leq k\leq I+n+J$.

Note-se que o Erro Quadrático Médio (EQM) que é obtido pela metodologia adotada tem de ser ajustado tendo em conta a inflação futura, pelo que ao ser considerada $\theta_{i,j}$, com $i=1,\ldots,I$ e $j=\max(0,J-i+1),\ldots,J$ como a inflação acumulada, ou seja, $\theta_{i,j}=\prod_{k=I+1}^{i+j}(1+\delta_{k-1})$, então tem-se que o EQM de cada provisão estimada por ano de ocorrência e ano de desenvolvimento é dada pela equação (2.10), o EQM da provisão estimada por ano de ocorrência é dada pela equação (2.11), o EQM da provisão estimada por ano de calendário é dada pela equação (2.12) e o EQM da provisão total estimada é dada pela equação (2.13):

$$\widehat{EQM}\left(\hat{Y}_{i,j}\right) = \widehat{EQM}\left(\hat{X}_{i,j}\prod_{k=I+1}^{i+j} (1+\delta_{k-1})\right) = \widehat{EQM}\left(\hat{X}_{i,j}\theta_{i,j}\right) = \theta_{i,j}^2 \widehat{EQM}\left(\hat{X}_{i,j}\right)$$
onde $1 \le i \le I+n$ e $\max(0, J-i+1) \le j \le J$

$$(2.10)$$

$$\begin{split} \widehat{EQM} \left(\hat{R}_{i}^{O} \right) &= \widehat{EQM} \left(\sum_{j=\max(0,J-i+1)}^{J} \hat{Y}_{i,j} \right) = \widehat{EQM} \left(\sum_{j=\max(0,J-i+1)}^{J} \hat{X}_{i,j} \theta_{i,j} \right) = \\ &= \sum_{\substack{j_{1}=\max(0,J-i+1)\\j_{2}=\max(0,J-i+1)}}^{J} \mathbb{C}\text{ov} \left[\theta_{i,j_{1}} \hat{X}_{i,j_{1}}, \theta_{i,j_{2}} \hat{X}_{i,j_{2}} \right] = \sum_{\substack{j_{1}=\max(0,J-i+1)\\j_{2}=\max(0,J-i+1)}}^{J} \theta_{i,j_{1}} \theta_{i,j_{2}} \mathbb{C}\text{ov} \left[\hat{X}_{i,j_{1}}, \hat{X}_{i,j_{2}} \right] = \\ &= \sum_{j=\max(0,J-i+1)}^{J} \theta_{i,j}^{2} \widehat{EQM} \left(\hat{X}_{i,j} \right) + \sum_{\substack{j_{1}=\max(0,J-i+1)\\j_{2}=\max(0,J-i+1)\\j_{1}\neq j_{2}}}^{J} \theta_{i,j_{1}} \theta_{i,j_{2}} \mathbb{C}\text{ov} \left[\hat{X}_{i,j_{1}}, \hat{X}_{i,j_{2}} \right] \\ &\text{onde } 1 \leq i \leq I+n \end{split}$$

$$\widehat{EQM}\left(\hat{R}_{k}^{C}\right) = \widehat{EQM}\left(\sum_{j=0}^{J} \hat{Y}_{k-j,j}\right) = \widehat{EQM}\left(\sum_{j=0}^{J} \hat{X}_{k-j,j}.\theta_{k-j,j}\right) = \\
= \sum_{\substack{j_{1}=0\\j_{2}=0}}^{J} \mathbb{C}\text{ov}\left[\theta_{k-j_{1},j_{1}}.\hat{X}_{k-j_{1},j_{1}},\theta_{k-j_{2},j_{2}}.\hat{X}_{k-j_{2},j_{2}}\right] = \\
= \sum_{\substack{j_{1}=0\\j_{2}=0}}^{J} \theta_{k-j_{1},j_{1}}.\theta_{k-j_{2},j_{2}}.\mathbb{C}\text{ov}\left[\hat{X}_{k-j_{1},j_{1}},\hat{X}_{k-j_{2},j_{2}}\right] = \\
= \sum_{j=0}^{J} \theta_{k-j,j}^{2} \widehat{EQM}\left(\hat{X}_{k-j,j}\right) + \sum_{\substack{j_{1}=0\\j_{2}=0\\j\neq j_{2}}}^{J} \theta_{k-j_{1},j_{1}}\theta_{k-j_{2},j_{2}}\mathbb{C}\text{ov}\left[\hat{X}_{k-j_{1},j_{1}},\hat{X}_{k-j_{2},j_{2}}\right] \\
\text{onde } I + 1 \leq k \leq I + n + J$$
(2.12)

$$\begin{split} \widehat{EQM}\left(\hat{R}^T\right) &= \widehat{EQM}\left(\sum_{i=1}^{I+n}\hat{R}_i^O\right) = \widehat{EQM}\left(\sum_{i=1}^{I+n}\left(\sum_{j=\max(0,J-i+1)}^{J}\hat{Y}_{i,j}\right)\right) = \\ &= \widehat{EQM}\left(\sum_{i=1}^{I+n}\left(\sum_{j=\max(0,J-i+1)}^{J}\hat{X}_{i,j}\theta_{i,j}\right)\right) = \\ &= \sum_{\substack{(i_1,j_1)\in\Delta\\(i_2,j_2)\in\Delta}}^{J}\mathbb{C}\mathrm{ov}\left[\theta_{i_1,j_1}\hat{X}_{i_1,j_1},\theta_{i_2,j_2}\hat{X}_{i_2,j_2}\right] = \sum_{\substack{(i,j_1)\in\Delta\\(i_2,j_2)\in\Delta}}^{J}\theta_{i_1,j_1}\theta_{i_2,j_2}\mathbb{C}\mathrm{ov}\left[\hat{X}_{i_1,j_1},\hat{X}_{i_2,j_2}\right] = \\ &= \sum_{\substack{(i,j)\in\Delta\\(i_2,j_2)\in\Delta\\(i_1,j_1)\neq(i_2,j_2)}}^{J}\theta_{i_1,j_1}\theta_{i_2,j_2}\mathbb{C}\mathrm{ov}\left[\hat{X}_{i_1,j_1},\hat{X}_{i_2,j_2}\right] \\ \mathrm{onde}\;\Delta = \left\{(i,j):1\leq i\leq I+n;\max(0,J-i+1)\leq j\leq J\right\} \end{split} \tag{2.13}$$

Para obtenção das estimativas para o erro padrão de cada estimador, é necessário calcular a raiz quadrada do respetivo EQM.

2.3 Metodologias

Para obter as estimativas da matriz de desenvolvimento recorreu-se a três métodos: o modelo Loglinear, o modelo linear generalizado de Poisson sobre-dispersão e o *Bootstrap*. Esta última metodologia foi usada em combinação com o modelo Loglinear.

Uma vez que a maioria da bibliografia apenas se restringe a calcular as provisões até o ano de origem I (último conhecido) e neste trabalho pretende-se extrapolar para os anos seguintes, optou-se por usar a metodologia Loglinear ou o modelo de Poisson sobre-dispersão de modo a obter uma diagonal do triângulo de desenvolvimento a cada iteração. Aplicou-se o mesmo método escolhido até atingir as n diagonais calculadas. Seguidamente aplicou-se o mesmo método para completar o triângulo de desenvolvimento. Relativamente à utilização do método de Bootstrap este consistiu na sua aplicação após a utilização do modelo Loglinear para o cálculo das diagonais do triângulo de desenvolvimento, preenchendo assim o resto do triângulo, como se exemplifica na imagem 2.3.

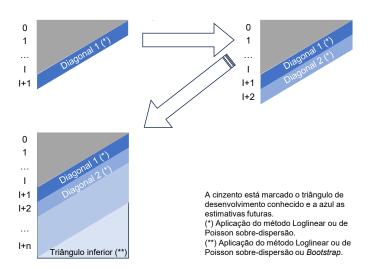


Figura 2.3: Processo de estimação das provisões para sinistros em anos de ocorrência futuros.

2.3.1 Modelo Loglinear

Para a aplicação do modelo Loglinear é necessário primeiro colocar os dados incrementais na escala logarítmica. O modelo usa a regressão múltipla para obter as estimativas das variáveis explicativas. Os resultados da estimação do modelo têm depois de ser transformados da escala logarítmica para a escala inicial, como mostra a figura 2.4.

Neste trabalho, a estimação das provisões para sinistros através do modelo Loglinear baseouse em Christofides [11]. No entanto como teremos que calcular mais estimativas para além do ano já conhecido, I, ter-se-á de fazer umas adaptações. Optou-se por fazer um método iterativo, em que a cada iteração se adiciona um ano de ocorrência, I+t, onde $1 \le t \le n$, sendo n o número de períodos para além do ano de exercício (último conhecido). Para este modelo, tem-se as seguintes definições:

- A_i representa os últimos pagamentos acumulados para o período de acidente $i, 0 \le i \le I$;
- B_j é a percentagem dos sinistros incrementais pagos durante o período de desenvolvimento j, $0 \le j \le J$; e
- $P_{i,j}$ representa o valor incremental pago durante o período de desenvolvimento j relativamente aos sinistros ocorridos no período i, onde $0 \le i \le I$ e $0 \le j \le J$. Esta definição é equivalente ao $X_{i,j}$ (ou $Y_{i,j}$, se for considerada a inclusão da inflação) definido na secção 2.2.

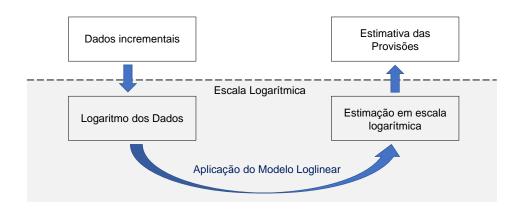


Figura 2.4: Processo de estimação das provisões para sinistros pelo modelo Loglinear [2].

Para calcular o $P_{i,j}$, com $0 \le i \le I$ e $0 \le j \le J$, basta efetuar a multiplicação de A_i por B_j , em que $\sum_{j=0}^J B_j = 1$. Sendo assim, este modelo baseia-se em dois parâmetros: um associado ao período de acidente (linha) e outro ao ano de desenvolvimento (coluna). Para transformar o cálculo de $P_{i,j}$ na escala logarítmica e admitindo que os valores de $P_{i,j}$ são estritamente positivos, chega-se à expressão (2.14):

$$\log P_{i,j} = \log(A_i \times B_j) = \log A_i + \log B_j \tag{2.14}$$

No entanto, aplicando o logaritmo à condição $\sum_{j=0}^J B_j = 1$, não é possível obter uma equação

linear uma vez que $\log \left(\sum_{j=0}^J B_j\right) \neq \sum_{j=0}^J \log B_j$. Deixando cair esta condição, redefinindo os parâmetros $\log A_i = a_i$, $\log B_j = b_j$ e introduzindo um erro $\varepsilon_{i,j}$, obtém-se a equação linear (2.15):

$$\log P_{i,j} = W_{i,j} = a_i + b_j + \varepsilon_{i,j},$$
 onde $\varepsilon_{i,j}$ é o erro estocástico associado. (2.15)

Neste modelo, é usual e conveniente assumir-se que os valores dos erros $\varepsilon_{i,j}$ são independentes e identicamente distribuídos, seguindo uma distribuição Normal de média zero e variância σ^2 , ou seja, $\varepsilon_{i,j} \sim N(0,\sigma^2)$ i.i.d.. A verificação da normalidade poderá ser feita recorrendo aos testes estatísticos como o de *Shapiro-Wilk*, ou a gráficos quantil-quantil.

Sendo assim, tem-se que $W_{i,j}$ segue uma distribuição Normal de média $a_i + b_j$ e variância σ^2 e portanto $P_{i,j}$ segue uma distribuição LogNormal, ou seja:

$$P_{i,j} \sim log N(a_i + b_j, \sigma^2) \tag{2.16}$$

Antes da aplicação do modelo é necessário ver se este se adapta aos dados existentes para poder tirar as conclusões necessárias. Uma forma de efetuar a análise de aplicabilidade do modelo Loglinear é construir os gráficos dos pagamentos incrementais *versus* período de ocorrência/desenvolvimento/pagamento e os mesmos gráficos na escala logarítmica. Os gráficos na escala logarítmica permitem ter uma perceção da escolha dos parâmetros: por ano de ocorrência, ano de desenvolvimento ou ano de calendário [2]. Após esta escolha, determina-se o erro da estimação, dado pela equação (2.17):

$$\hat{s}^2 = \frac{1}{m-p} \sum_{i,j} \left[\left(W_{i,j} - \hat{W}_{i,j} \right)^2 \right], \quad 0 \le i \le I \quad \text{e} \quad 0 \le j \le J - i$$
 (2.17)

onde m o número de observações conhecidas, p o número de parâmetros a estimar, I é último período de ocorrência conhecido e J o último período de desenvolvimento conhecido.

Da análise do erro é possível verificar se este modelo se ajusta aos dados, uma vez que se pretende um valor de erro baixo, que evidencia que existe grande proximidade entre os dados reais e os obtidos pelo modelo.

Outra análise relevante é a análise gráfica dos resíduos *versus* ano de ocorrência/desenvolvimento/ calendário, resíduos *versus* dados modelizados e a normalidade dos resíduos. Se da análise gráfica se verificar uma tendência, os erros podem não ser independentes e falha um dos pressupostos do modelo Loglinear. Note-se que para além da análise gráfica também poderão ser aplicados testes estatísticos de independência. Também é desejável que os resíduos se aproximem de uma distribuição Normal de média zero, que pode ser verificado com o teste *Shapiro-Wilk*.

Neste modelo existem I+J+2 parâmetros a estimar: I+1 parâmetros de linha, a_i , e J+1 parâmetros de coluna, b_j . Por conveniência irá assumir-se que $a_0=b_0=0$. Esta condição introduz uma constante k ao modelo, que representa o padrão seguido por todos os $W_{i,j}^6$. Ou seja, para efetuar as estimativas de $W_{i,j}$, temos a expressão definida na equação (2.18):

$$\hat{W}_{i,j} = \hat{k} + \hat{a}_i + \hat{b}_j + \varepsilon_{i,j} \tag{2.18}$$

Neste sentindo estamos perante um modelo de regressão linear e podemos escrever a equação (2.18) de forma matricial, resultando a equação (2.19):

$$W = X\beta + \varepsilon \tag{2.19}$$

onde W é a matriz coluna com os logaritmos das observações conhecidas $(m \times 1)$, X é a matriz que define o modelo $(m \times p)$, β é a matriz coluna dos parâmetros $(p \times 1)$, sendo m o número de observações conhecidas e p o número de parâmetros a estimar (neste caso, $m = \frac{(I+2)\times (J+1)}{2}$ e p = I + J + 1).

As estimativas para os parâmetros podem ser obtidas através do método dos mínimos quadrados, que tendo em conta o modelo de regressão linear, são dadas pela equação matricial (2.20):

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \mathbf{W}$$
 (2.20)

A verificação dos parâmetros quanto à sua significância é relevante dado que da sua exclusão (por não serem significativamente diferente de zero) pode resultar um modelo com melhor ajustamento. Para comparar os modelos recorre-se ao coeficiente de determinação R^2 , onde o melhor modelo terá o valor mais próximo de 1, e cujo cálculo é dado pela equação (2.21):

$$R^{2} = \frac{\sum_{i,j} (\hat{W}_{i,j} - \overline{W})^{2}}{\sum_{i,j} (W_{i,j} - \overline{W})^{2}} \quad 0 \le i \le I \quad \mathbf{e} \quad 0 \le j \le J - i$$
(2.21)

Depois de calculadas as estimativas para os parâmetros do modelo é necessário calcular as estimativas para os montantes incrementais futuros relativos à primeira diagonal⁷, $\hat{P}_{i,j}$, $1 \le i \le I+1$ e max $(0,J-i+1) \le j \le J$. Tendo em conta que os montantes incrementais seguem a distribuição LogNormal de parâmetros $k+a_i+b_j$ e σ^2 , tem-se que a estimativa é dada pela

⁶Recorde-se que $W_{i,j} = \log(P_{i,j})$, onde $P_{i,j}$ é o valor incremental pago durante o período de desenvolvimento j relativamente aos sinistros ocorridos no período i.

⁷A primeira diagonal é calculada na primeira iteração do processo. As seguintes diagonais são obtidas de forma análoga.

equação (2.22) e a variância de $\hat{P}_{i,j}$ é dada pela equação (2.23):

$$\hat{P}_{i,j} = \mathbb{E}\left[P_{i,j}\right] = \exp\left(\hat{k} + \hat{a}_i + \hat{b}_j + \frac{\widehat{EQM}(\hat{W}_{i,j})}{2}\right)$$
(2.22)

$$\mathbb{V}\big[\hat{P}_{i,j}\big] = \left(\exp(\widehat{EQM}(\hat{W}_{i,j})) - 1\right) \exp\left(2(\hat{k} + \hat{a}_i + \hat{b}_j) + \widehat{EQM}(\hat{W}_{i,j})\right) \\
= \left(\exp(\widehat{EQM}(\hat{W}_{i,j})) - 1\right) \left[\exp\left(\hat{k} + \hat{a}_i + \hat{b}_j + \frac{\widehat{EQM}(\hat{W}_{i,j})}{2}\right)\right]^2 \\
= \hat{P}_{i,j}^2 \left(\exp(\widehat{EQM}(\hat{W}_{i,j})) - 1\right) \tag{2.23}$$

onde $1 \le i \le I + 1$ e $\max(0, J - i + 1) \le j \le J$.

Para completar as equações (2.22) e (2.23) é necessário calcular o $\widehat{EQM}(\hat{W}_{i,j})$, sendo dada pela equação (2.24):

$$\widehat{EQM}(\hat{W}_{i,j}) = \mathbb{E}[(\hat{W}_{i,j} - W_{i,j})^{2}] = \mathbb{E}[(\hat{W}_{i,j} - \mathbb{E}[\hat{W}_{i,j}] + \mathbb{E}[\hat{W}_{i,j}] - W_{i,j})^{2}] = \\
= \mathbb{E}[(\hat{W}_{i,j} - \mathbb{E}[\hat{W}_{i,j}])^{2} + 2(\hat{W}_{i,j} - \mathbb{E}[\hat{W}_{i,j}])(\mathbb{E}[\hat{W}_{i,j}] - W_{i,j}) + \\
+ (\mathbb{E}[\hat{W}_{i,j}] - W_{i,j})^{2}] = \\
= \mathbb{E}[(\hat{W}_{i,j} - \mathbb{E}[\hat{W}_{i,j}])^{2}] + 2\mathbb{E}[(\hat{W}_{i,j} - \mathbb{E}[\hat{W}_{i,j}])(\mathbb{E}[\hat{W}_{i,j}] - W_{i,j})] + \\
+ \mathbb{E}[(\mathbb{E}[\hat{W}_{i,j}] - W_{i,j})^{2}]$$
(2.24)

Uma vez que $\mathbb{E}[\hat{W}_{i,j}] - W_{i,j}$ é constante em relação a $\hat{W}_{i,j}$, então a equação anterior, traduz-se na equação (2.25):

$$\widehat{EQM}(\hat{W}_{i,j}) = \mathbb{E}\left[\left(\hat{W}_{i,j} - \mathbb{E}\left[\hat{W}_{i,j}\right]\right)^{2}\right] + 2\left(\mathbb{E}\left[\hat{W}_{i,j}\right] - W_{i,j}\right)\mathbb{E}\left[\mathbb{E}\left[\hat{W}_{i,j}\right] - \hat{W}_{i,j}\right] + \left(\mathbb{E}\left[\hat{W}_{i,j}\right] - W_{i,j}\right)^{2} =$$

$$= \mathbb{V}\left[\hat{W}_{i,j}\right] + 2\left(\mathbb{E}\left[\hat{W}_{i,j}\right] - W_{i,j}\right)\left(\mathbb{E}\left[\hat{W}_{i,j}\right] - \mathbb{E}\left[\hat{W}_{i,j}\right]\right) + \mathbb{V}\left[W_{i,j}\right]$$

$$= \hat{s}^{2} + \mathbb{V}\left[W_{i,j}\right]$$

$$(2.25)$$

onde \hat{s}^2 é o estimador do erro estocástico σ^2 dado pela equação (2.17), que neste caso se pode reescrever como $\hat{s}^2 = \frac{1}{m-p} \sum_{i,j} \left[W_{i,j} - \left(\hat{k} + \hat{a}_i + \hat{b}_j \right) \right]^2$, para i e j tal que $W_{i,j}$ é conhecido.

Para obter a variância de $W_{i,j}$ é necessário calcular a matriz de variância-covariância, que é

obtida através do produto de matrizes constante na equação (2.26):

$$V[W] = \sigma^2 X_f (X^T X)^{-1} X_f^T$$
(2.26)

onde σ^2 é a variância do modelo e depende dos dados, \boldsymbol{X} é a matriz que define o modelo $(m \times p)$, \boldsymbol{X}_f $(m_f \times p)$ é a matriz que define os valores futuros, sendo m o número de observações conhecidas, m_f o número de observações a estimar e p o número de parâmetros a estimar (neste caso, $m = \frac{(I+2)\times(J+1)}{2}$, $m_f = \frac{(I+2)\times(J+1)}{2}$ e p = I+J+1).

Uma vez que já se tem a estimativa e o erro para o valor das indemnizações incrementais, $P_{i,j}$, é possível efetuar a estimativa da provisão para sinistros por ano de ocorrência, ano de calendário ou o total da matriz de desenvolvimento (através das equações (2.4), (2.5) e (2.6), respetivamente) e calcular o seu erro.

O erro relativo às estimativas das provisões para sinistros por ano de ocorrência, ano de calendário ou as provisões totais, considerando os cálculos da primeira iteração, são definidas pelas equações (2.27), (2.28) e (2.29), respetivamente.

$$\widehat{EQM}(\hat{R}_{i}^{O}) = \sum_{\substack{j_{1} = \max(0, J - i + 1)\\ j_{1} = \max(0, J - i + 1)\\ j_{1} \neq j_{2}}}^{J} \widehat{EQM}(\hat{P}_{i, j}) + \sum_{\substack{j_{1} = \max(0, J - i + 1)\\ j_{1} \neq j_{2}\\ j_{1} \neq j_{2}}}^{J} \mathbb{C}\text{ov}[\hat{P}_{i, j_{1}}, \hat{P}_{i, j_{2}}]$$
(2.27)

onde $1 \le i \le I + 1$

$$\widehat{EQM}(\hat{R}_{k}^{C}) = \sum_{j=0}^{J} \widehat{EQM}(\hat{P}_{k-j,j}) + \sum_{\substack{j_{1}=0\\j_{2}=0\\j\neq j_{2}}}^{J} \mathbb{C}\text{ov}[\hat{P}_{k-j_{1},j_{1}}, \hat{P}_{k-j_{2},j_{2}}]$$
onde $I + 1 \le k \le I + J + 1$ (2.28)

$$\begin{split} \widehat{EQM}\Big(\hat{R}^T\Big) &= \sum_{(i,j)\in\Delta}^J \widehat{EQM}\Big(\hat{P}_{i,j}\Big) + \sum_{\substack{(i_1,j_1)\in\Delta\\(i_2,j_2)\in\Delta\\(i_1,j_1)\neq(i_2,j_2)}}^J \mathbb{C}\mathrm{ov}\Big[\hat{P}_{i_1,j_1},\hat{P}_{i_2,j_2}\Big] \\ \mathrm{onde}\; \Delta &= \{(i,j): 1\leq i\leq I+1; \max(0,J-i+1)\leq j\leq J\} \end{split} \tag{2.29}$$

Tendo como base o resultado de Aitchison and Brown [29], a covariância entre duas estimati-

vas \hat{P}_{i_1,j_1} e \hat{P}_{i_2,j_2} é dada pela equação (2.30):

$$\mathbb{C}\text{ov}\Big[\hat{P}_{i_1,j_1},\hat{P}_{i_2,j_2}\Big] = \hat{P}_{i_1,j_1}\hat{P}_{i_2,j_2}\bigg(\exp\big(\mathbb{C}\text{ov}\Big[\hat{W}_{i_1,j_1},\hat{W}_{i_2,j_2}\Big]\big) - 1\bigg)$$
(2.30)

com

$$\operatorname{Cov}\left[\hat{W}_{i_{1},j_{1}},\hat{W}_{i_{2},j_{2}}\right] = \operatorname{Cov}\left[\hat{k} + \hat{a}_{i_{1}} + \hat{b}_{j_{1}},\hat{k} + \hat{a}_{i_{2}} + \hat{b}_{j_{2}}\right] = \\
= \mathbb{V}\left[\hat{k}\right] + \operatorname{Cov}\left[\hat{k},\hat{a}_{i_{1}}\right] + \operatorname{Cov}\left[\hat{k},\hat{a}_{i_{2}}\right] + \operatorname{Cov}\left[\hat{k},\hat{b}_{j_{1}}\right] + \operatorname{Cov}\left[\hat{k},\hat{b}_{j_{2}}\right] + \\
+ \operatorname{Cov}\left[\hat{a}_{i_{1}},\hat{a}_{i_{2}}\right] + \operatorname{Cov}\left[\hat{a}_{i_{1}},\hat{b}_{j_{2}}\right] + \operatorname{Cov}\left[\hat{b}_{j_{1}},\hat{a}_{i_{2}}\right] + \operatorname{Cov}\left[\hat{b}_{j_{1}},\hat{b}_{j_{2}}\right] \tag{2.31}$$

Após concluída a primeira iteração, extraiu-se a primeira diagonal e juntou-se ao triângulo de desenvolvimento com os dados reais (excluindo a sua primeira linha), aplicando novamente o método a este novo triângulo. Repetiu-se o método até atingir as n diagonais, sendo que na última iteração calcularam-se todos os valores do triângulo inferior.

2.3.2 Modelo de Poisson sobre-dispersão

O modelo de Poisson sobre-dispersão, adiante chamado de ODP proveniente do nome *Over-dispersed Poisson*, é da família dos modelos lineares generalizados. A modelação através dos modelos lineares generalizados é obtida através da relação entre variáveis, generalizando o modelo linear clássico através de pressupostos menos restritivos, e fornecendo uma análise para os dados não normais [30].

Para usar os modelos lineares generalizados necessitamos de ter um preditor linear $\eta = X\beta$ $(m \times 1)$, onde X é a matriz que define o modelo $(m \times p)$, β é a matriz coluna dos parâmetros $(p \times 1)$, sendo m o número de observações conhecidas e p o número de parâmetros a estimar. Considere-se a função de ligação g definida por $g(\mu) = \eta$ e Z a matriz coluna das variáveis dependentes. Tem-se a relação presente na equação (2.32):

$$\mathbb{E}[\boldsymbol{Z}|\boldsymbol{X}] = \mu = g^{-1}(\boldsymbol{\eta}) \tag{2.32}$$

As variáveis Z_i , $0 \le i \le I$ pertencem à família exponencial, tendo como função de densidade de probabilidade a expressão da equação (2.33):

$$f(z_i|\theta_i,\phi) = h(z_i,\phi).exp\left\{w_i \frac{z_i\theta_i - b(\theta_i)}{\phi}\right\}$$
 (2.33)

com b(.) e h(.) funções específicas para cada distribuição (Normal, Exponencial, Gamma, Poisson, etc.), ϕ um parâmetro de escala positivo e constante para cada variável Z_i e w_i um peso conhecido.

Para determinar o valor esperado e a variância da variável aleatória Z_i recorrer-se-á a definição de função de densidade de probabilidade. Sendo assim, partindo da equação (2.33) obtém-se a equação (2.34):

$$\int h(z_{i},\phi).exp\left\{w_{i}\frac{z_{i}\theta_{i}-b(\theta_{i})}{\phi}\right\}dz_{i} = 1 \iff
\iff \int h(z_{i},\phi).exp\left\{w_{i}\frac{z_{i}\theta_{i}}{\phi}\right\}dz_{i} = exp\left\{w_{i}\frac{b(\theta_{i})}{\phi}\right\}$$
(2.34)

Derivando ambos os membros da equação (2.34) em ordem a θ_i , obtém-se a equação (2.35):

$$\int h(z_{i},\phi) \frac{w_{i}z_{i}}{\phi} exp\left\{w_{i} \frac{z_{i}\theta_{i}}{\phi}\right\} dz_{i} = b'(\theta_{i}) \frac{w_{i}}{\phi} exp\left\{w_{i} \frac{b(\theta_{i})}{\phi}\right\} \iff \int z_{i}h(z_{i},\phi) .exp\left\{w_{i} \frac{z_{i}\theta_{i} - b(\theta_{i})}{\phi}\right\} dz_{i} = b'(\theta_{i})$$
(2.35)

Como por definição o valor esperado de uma variável aleatória X é dado por $\mathbb{E}[X] = \int x f(x) dx$, com f(.) a função de densidade de probabilidade, então a equação (2.35) transformase na equação (2.36):

$$\mathbb{E}[Z_i] = b'(\theta_i) \tag{2.36}$$

Derivando a equação (2.35) em ordem a θ_i , obtém-se a equação 2.37:

$$\int z_{i}h(z_{i},\phi).\frac{w_{i}}{\phi}(z_{i}-b'(\theta_{i})).exp\left\{w_{i}\frac{z_{i}\theta_{i}-b(\theta_{i})}{\phi}\right\}dz_{i}=b''(\theta_{i})\iff$$

$$\iff \int z_{i}^{2}h(z_{i},\phi).exp\left\{w_{i}\frac{z_{i}\theta_{i}-b(\theta_{i})}{\phi}\right\}dz_{i}-b'(\theta_{i})\int z_{i}h(z_{i},\phi).exp\left\{w_{i}\frac{z_{i}\theta_{i}-b(\theta_{i})}{\phi}\right\}dz_{i}=\frac{\phi}{w_{i}}b''(\theta_{i})\iff$$

$$\iff \mathbb{E}[Z_{i}^{2}]-\mathbb{E}[Z_{i}]\times\mathbb{E}[Z_{i}]=\frac{\phi}{w_{i}}b''(\theta_{i})\iff$$

$$\iff \mathbb{V}[Z_{i}]=\frac{\phi}{w_{i}}b''(\theta_{i})$$

$$(2.37)$$

Para o exemplo do triângulo de desenvolvimento com os montantes incrementais $P_{i,j}$, considerando i o período de origem e j o período de desenvolvimento, o modelo ODP assume que $P_{i,j}$ são variáveis aleatórias independentes, seguindo uma distribuição de Poisson com uma dispersão dos dados superior à média (sobre-dispersão) e a forma geral deste modelo é caracterizada por pelas condições (2.38) e (2.39):

$$\mathbb{E}(P_{i,j}) = x_i y_j, \tag{2.38}$$

$$\mathbb{V}(P_{i,j}) = \phi x_i y_j, \tag{2.39}$$

onde $\sum_{j=0}^{J} y_j = 1$ e ϕ é o parâmetro de escala, que não é conhecido e é estimado a partir dos dados, x_i representa o valor do último período de desenvolvimento conhecido relativamente ao período de origem i e y_j representa a proporção do valor do último período de desenvolvimento no valor total.

Outra alternativa muito usada para expressar o modelo ODP é usando a forma linear em vez da forma multiplicativa, recorrendo ao uso da função logarítmica como função de ligação. Tem-se então a seguinte formulação da equação (2.40):

$$\mathbb{E}[P_{i,j}] = \mu_{i,j}, \quad \mathbb{V}[P_{i,j}] = \phi \mu_{i,j}$$

$$\eta_{i,j} = \log(\mu_{i,j}) = \alpha_i + \beta_j + \gamma$$
(2.40)

com $0 \le i \le I$ e $0 \le j \le J$, tal que $P_{i,j}$ sejam conhecidos. Ainda é necessário adicionar a condição $\alpha_0 = \beta_0 = 0$ para evitar que o modelo tenha mais parâmetros que os necessários. Sendo assim, para esta definição do modelo, os valores relativos à estimativa nos pontos (i,j) = (0,J) e (i,j) = (I,0) são os valores atuais [31].

Na definição do modelo através das equações (2.40), α_i representa o fator associado às linhas do triângulo de desenvolvimento (ou seja, associado ao período de origem), β_j representa o fator associado às colunas do triângulo de desenvolvimento (ou seja, associado ao período de desenvolvimento) e γ é o parâmetro constante que é aplicado para todas as células do triângulo de desenvolvimento.

O modelo ODP na sua formulação matricial segue a equação (2.41):

$$\log\left(\mathbb{E}[\boldsymbol{P}]\right) = \mathbf{X}\boldsymbol{\beta},\tag{2.41}$$

onde X $(m \times p)$ é a matriz que define o modelo, β $(p \times 1)$ é a matriz coluna com os parâmetros e P $(m \times 1)$ é o vetor coluna com as observações conhecidas, ou seja:

$$\mathbf{P} = [P_{0,0} \quad P_{1,0} \quad \dots \quad P_{I,0} \quad P_{1,0} \quad P_{1,1} \quad \dots \quad P_{I-1,1} \quad \dots \quad P_{i,0} \quad \dots \quad P_{i,j} \quad \dots \quad P_{0,J-1} \quad P_{1,J-1} \quad P_{0,J}]^T$$
(2.42)

onde para o caso em estudo, $m = \frac{(I+2)(J+1)}{2}$ e p = I+J+1.

Para estimar os parâmetros do modelo, usa-se o processo iterativo de mínimos quadrados denominado *Iteratively Reweighed Least Squares (IRLS)*. Para isso, é necessário definir a forma de estimação dos parâmetros, que é dada pela a equação (2.43) e os parâmetros iniciais definidos

em (2.46):

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{Z}, \tag{2.43}$$

com

$$Z = X\beta + \frac{X - \exp(X\beta)}{\exp(X\beta)}$$
 e (2.44)

$$\mathbf{W}_{i,j} = \begin{cases} \exp\left((\mathbf{X}\boldsymbol{\beta})_i\right) & i = j \\ 0 & i \neq j \end{cases}$$
 (2.45)

$$\mathbf{W}_{i,j}^{(0)} = \begin{cases} \mathbf{P}_i & i = j \\ 0 & i \neq j \end{cases}, \quad \boldsymbol{\beta}_i^{(0)} = log(\mathbf{P}_i)$$
 (2.46)

Com os valores iniciais, substitui-se nas equações (2.44) e (2.43). Desta última determina-se o valor de $\beta^{(1)}$ a usar na próxima iteração. Com estes novos valores, recalculam-se as equações (2.44) e (2.45) e assim sucessivamente até que a diferença absoluta entre $\beta^{(t+1)}$ e $\beta^{(t)}$ seja inferior a um valor positivo muito pequeno.

Tendo os valores estimados dos parâmetros, os valores estimados para os dados incrementais do triângulo de desenvolvimento são dados pela equação (2.47):

$$\hat{P}_{i,j} = \mathbb{E}(P_{i,j}) = \exp(\hat{\eta}_{i,j}) = \exp(\hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma})$$
(2.47)

A partir do raciocínio efetuado para obtenção das equações (2.24) e (2.25), sabe-se que o valor do erro da estimativa dos valores incrementais é dado por:

$$\widehat{EQM}(\hat{P}_{i,j}) = \mathbb{V}[\hat{P}_{i,j}] + \mathbb{V}[P_{i,j}]$$
(2.48)

Pela definição do modelo ODP (2.40), tem-se que $\mathbb{V}\big[P_{i,j}\big] = \hat{\phi}\hat{\mu}_{i,j} = \hat{\phi}\exp(\hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma})$ onde a estimativa para o parâmetro de escala pode ser calculada através da soma dos resíduos de *Pearson* enunciada na equação (2.49):

$$\hat{\phi} = \frac{1}{m-p} \sum_{i,j} \left(\frac{P_{i,j} - \hat{\mu}_{i,j}}{\sqrt{\mathbb{V}(\hat{\mu}_{i,j})}} \right)^2, \quad 0 \le i \le I \text{ e } 0 \le j \le J - i$$
 (2.49)

sendo m o número de observações conhecidas e p o número de parâmetros a estimar (neste caso, $m=\frac{(I+2)\times (J+1)}{2}$ e p=I+J+1).

Quanto à componente $\mathbb{V}[\hat{P}_{i,j}]$, aplicar-se-á o modelo Delta [32], que a partir da série de Taylor obtém-se uma aproximação de $\hat{P}_{i,j}$. A expansão da série de Taylor é feita sobre a média

de $\hat{P}_{i,j}$, da forma descrita na equação (2.50):

$$\hat{P}_{i,j} \approx \exp(\hat{\eta}_{i,j}) + (\hat{\eta}_{i,j} - \exp(\hat{\eta}_{i,j})) \frac{\mathrm{d}}{\mathrm{d}\eta_{i,j}} \exp(\eta_{i,j}) = \exp(\hat{\eta}_{i,j}) + (\hat{\eta}_{i,j} - \exp(\hat{\eta}_{i,j})) \exp(\eta_{i,j})$$
(2.50)

O estimador da variância pelo modelo Delta é obtido aplicando a variância à equação anterior [32]. Verifica-se que:

$$\mathbb{V}\big[\hat{P}_{i,j}\big] \approx \mathbb{V}\big[\exp(\hat{\eta}_{i,j})\big] + \mathbb{V}\big[\big(\hat{\eta}_{i,j} - \exp(\hat{\eta}_{i,j})\big) \exp(\eta_{i,j})\big] = 0 + \big(\exp(\eta_{i,j})\big)^2 \big(\mathbb{V}\big[\hat{\eta}_{i,j}\big] - 0\big) =$$

$$= \big(\exp(\eta_{i,j})\big)^2 \mathbb{V}\big[\hat{\eta}_{i,j}\big]$$
(2.51)

Resta então determinar a variância dos preditores lineares $\hat{\eta}_{i,j}$ para obter o EQM de $\hat{P}_{i,j}$. Partindo da matriz de variâncias e covariâncias dos parâmetros estimados $\hat{\beta}$, que é dada pela equação matricial (2.52), e construindo a matriz X_f que define o modelo em termos dos valores futuros, determina-se a matriz de variância e covariância dos preditores lineares na parte inferior do triângulo de desenvolvimento, dada pela equação (2.53). A diagonal desta última matriz é a que contém as variâncias dos preditores lineares.

$$\hat{\mathbf{\Sigma}} = \hat{\phi} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$$
 (2.52)

sendo X a matriz que define o modelo $(m \times p)$ e W a matriz definida em (2.45) $(m \times m)$.

$$\hat{\mathbf{N}} = \mathbf{X}_f \hat{\mathbf{\Sigma}} (\mathbf{X}_f)^T \tag{2.53}$$

Sendo assim, a equação (2.48) pode ser reescrita da seguinte forma:

$$\widehat{EQM}(\hat{P}_{i,j}) \approx (\exp(\eta_{i,j}))^2 \mathbb{V}[\hat{\eta}_{i,j}] + \hat{\phi} \exp(\hat{\eta}_{i,j})$$
(2.54)

com $\hat{\eta}_{i,j} = \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}$, $\hat{\phi}$ dado pela equação 2.49 e $\mathbb{V}[\hat{\eta}_{i,j}]$ pode ser extraída da diagonal da matriz \hat{N} da equação (2.53).

Para chegar a esta expressão foi assumido que os montantes dos sinistros futuros são independentes dos sinistros ocorridos no passado e que $\hat{P}_{i,j}$ só depende dos sinistros passados [6].

Para o cálculo do EQM das provisões para sinistros por ano de origem, ano de calendário ou o valor total, teremos de somar não só o EQM de cada montante a pagar de cada sinistro futuro como a covariância entre dois quaisquer montantes de sinistros futuros, \hat{P}_{i_1,j_1} e \hat{P}_{i_2,j_2} com

 $(i_1, j_1) \neq (i_2, j_2)$. Ou seja, o EQM da provisão por ano de origem i é dado pela equação (2.55):

$$\widehat{EQM}(\hat{R}_{i}) \approx \sum_{j=I-i+1}^{J} \hat{\phi} \exp(\hat{\eta}_{i,j}) + \sum_{j=I-i+1}^{J} (\exp(\hat{\eta}_{i,j}))^{2} \mathbb{V}[\hat{\eta}_{i,j}]$$

$$+ \sum_{\substack{j_{1} = \max(0, J-i+1) \\ j_{2} = \max(0, J-i+1) \\ j_{1} \neq j_{2}}}^{J} \exp(\hat{\eta}_{i,j_{1}}) \exp(\hat{\eta}_{i,j_{2}}) \mathbb{C}\text{ov}[\hat{\eta}_{i,j_{1}}, \hat{\eta}_{i,j_{2}}]$$
onde $1 < i < I + 1$ (2.55)

O EQM da provisão para sinistros por ano de calendário pode ser calculada através da equação (2.56):

$$\widehat{EQM}(\hat{R}_{k}) \approx \sum_{j=0}^{J} \hat{\phi} \exp(\hat{\eta}_{k-j,j}) + \sum_{j=0}^{J} \left(\exp(\hat{\eta}_{\min(k-j,I+n),j}) \right)^{2} \mathbb{V}[\hat{\eta}_{\min(k-j,I+n),j}]$$

$$+ \sum_{\substack{j_{1}=0\\j_{2}=0\\j_{1}\neq j_{2}}}^{J} \exp(\hat{\eta}_{\min(k-j_{1},I+n),j_{1}}) \exp(\hat{\eta}_{\min(k-j_{2},I+n),j_{2}}) \mathbb{C}\text{ov}[\hat{\eta}_{i_{1},j_{1}},\hat{\eta}_{i_{2},j_{2}}]$$
onde $I+1 < k < I+n+J$ (2.56)

Para o cálculo do erro da previsão total tem-se o resultado da equação (2.57):

$$\widehat{EQM}(\hat{R}) \approx \sum_{\substack{(i,j) \in \Delta \\ (i_1,j_1) \in \Delta \\ (i_2,j_2) \in \Delta \\ (i_1,j_1) \neq (i_2,j_2)}}^{J} \exp(\hat{\eta}_{i,j}) + \sum_{\substack{(i,j) \in \Delta \\ (i_2,j_2) \in \Delta \\ (i_1,j_1) \neq (i_2,j_2)}}^{J} \exp(\hat{\eta}_{i_1,j_1}) \exp(\hat{\eta}_{i_2,j_2}) \mathbb{C}\text{ov}[\hat{\eta}_{i_1,j_1}, \hat{\eta}_{i_2,j_2}]$$

$$\text{onde } \Delta = \{(i,j) : 1 \le i \le I+1; \max(0, J-i+1) \le j \le J\}$$

Para analisar a adequabilidade deste modelo recorre-se à análise dos resíduos produzidos por este. Por exemplo, no modelo Normal os resíduos são calculados pela diferença entre os dados observados e os dados modelizados. Na teoria dos modelos lineares generalizados usam-se normalmente os resíduos de Pearson ou os resíduos baseados na deviance. Os primeiros são dados pela fórmula $r_{i,j}^P = \frac{P_{i,j} - \hat{\mu}_{i,j}}{\sqrt{V(\hat{\mu}_{i,j})}}$ e os segundos dados por $r_{i,j}^D = sgn(P_{i,j} - \hat{\mu}_{i,j})\sqrt{D_{i,j}}$ onde sgn(x) é a função que indica o sinal de x (-1, se x é negativo ou 1, se x é positivo) e $D_{i,j}$ é igual ao dobro da diferença das razões log-verosimilhança do modelo proposto comparativamente ao

modelo saturado⁸ com $1 \le i \le I + 1$ e max $(0, I - i + 1) \le j \le J$. No caso do modelo ODP os resíduos são dados pela equação (2.58) [33]:

$$r_{i,j} = \frac{P_{i,j} - \hat{\mu}_{i,j}}{\sqrt{\hat{\phi}\hat{\mu}_{i,j}}}$$
 (2.58)

sendo ϕ o parâmetro de escala referido anteriormente.

Neste modelo é desejável que os resíduos sigam uma distribuição Normal uma vez que a existência de um desvio significante na normalidade poderá indicar que a distribuição escolhida não é a correta [34].

É também frequente representar-se graficamente os resíduos relativamente aos valores ajustados e as variáveis explicativas. No primeiro avalia-se a adequação do modelo sendo aceite caso não se verifique uma tendência dos resíduos em função dos valores estimados. Relativamente aos gráficos com as variáveis explicativas, é desejável que não apresentem nenhum padrão ou tendência.

Após a conclusão da primeira iteração, extraiu-se a primeira diagonal e juntou-se ao triângulo de desenvolvimento com os dados reais (excluindo a sua primeira linha), aplicando novamente o método a este novo triângulo. Repetiu-se o método até atingir as n diagonais, sendo que na última iteração calcularam-se todos os valores do triângulo inferior.

2.3.3 Bootstrap

O *Bootstrap* é uma técnica de simulação por reamostragem, inicialmente proposta por Efron [35]. Esta técnica utiliza uma amostra de dados a partir da qual são produzidas várias subamostras aleatórias, chamadas de pseudo dados cuja distribuição é consistente com a distribuição da amostra original. Este método também permite aferir os erros de previsão e os intervalos de confiança, sendo considerado como uma alternativa aos tradicionais métodos estatísticos e como complemento na análise dos resultados obtidos [36].

De salientar que a metodologia *Bootstrap* não é, por si só, um modelo de avaliação de provisões, sendo utilizado em associação com um modelo estocástico ou determinístico, em que as provisões estimadas são dadas pelo modelo subjacente e as medidas de variabilidade através do *Bootstrap* [2]. Neste caso, iremos apenas aprofundar o caso em que o modelo estocástico subjacente é o modelo de Poisson sobre-dispersão.

Na geração de amostras através da técnica de *Bootstrap* podem-se usar métodos paramétricos ou não-paramétricos, sendo que para os primeiros é necessário ter um conhecimento prévio da distribuição de probabilidade dos dados. Nos métodos paramétricos a técnica de *Bootstrap* é

⁸Um modelo saturado é um modelo que tem tantos parâmetros como observações.

realizada com base nas estimativas dos parâmetros da distribuição de probabilidade (obtidas através do método da máxima verosimilhança), enquanto que nos métodos não-paramétricos a reamostragem é feita tendo conta que o conjunto das variáveis aleatórias são independentes e identicamente distribuídas [23].

A técnica de *Bootstrap* pode ser aplicada aos dados observados ou aos resíduos, sendo a primeira opção mais robusta que a segunda. No entanto, no contexto das provisões para sinistros só é possível aplicar a metodologia aos resíduos dado que existe dependência entre algumas observações e a estimativa dos parâmetros [37]. Sendo assim, para aplicação desta técnica é necessário garantir que os resíduos são independentes e identicamente distribuídos. Veja-se a figura 2.5 que explica graficamente o procedimento na aplicação da técnica de *Bootstrap*.

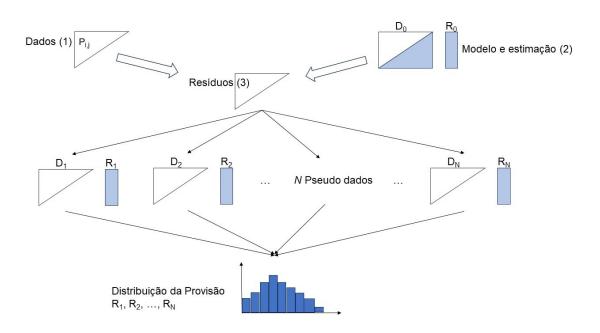


Figura 2.5: Técnica de *Bootstrap* aplicada às provisões para sinistros [3].

Tal como explicado anteriormente, antes da aplicação da técnica de *Bootstrap* aos dados reais extraídos do *website* da ASF, foi aplicado o método Loglinear até calcular n diagonais e juntar ao triângulo de desenvolvimento. Ou seja, a aplicação da técnica de *Bootstrap* será aplicada a um triângulo com I+1 linhas mas que por questão de coerência com a notação mencionada anteriormente corresponderá desde a linha n até à linha I+n.

Para a aplicação da técnica de *Bootstrap* deve-se primeiramente transformar os dados históricos num triângulo de desenvolvimento com dados acumulados, $C_{i,j} = \sum_{k=0}^{J-i} P_{i,k}$, com $i=0,\ldots,I$ e $j=0,\ldots,J$. Depois aplica-se o modelo escolhido ao triângulo de desenvolvimento para proceder à estimação da provisão para sinistros, R_0 , que na figura 2.5 corresponde o número

(2). No ponto (3) da figura referida, encontra-se o passo seguinte que consiste no cálculo dos resíduos que, se for assumido como tendo uma distribuição Normal, é dado pela diferença entre os valores observados e os valores modelizados, ambos incrementais. Considerando como modelo de cálculo das estimativas analíticas das provisões o modelo de Poisson sobre-dispersão, então utilizar-se-á os resíduos de Pearson, $r_{i,j}^P$, obtidos através da seguinte equação:

$$r_{i,j}^{P} = \frac{P_{i,j} - D_{i,j}^{0}}{\sqrt{D_{i,j}^{0}}}$$
 (2.59)

sendo $P_{i,j}$ os montantes incrementais observados e $D_{i,j}^0$ os valores modelizados.

Note-se que existem dois valores nos resíduos que serão obrigatoriamente nulos uma vez que o valor modelizado é igual ao valor observado: $P_{I,0}$ e $P_{0,J}$. Sendo assim a amostra de resíduos a considerar para efetuar a reamostragem tem $\frac{I\times (J+1)}{2}$ elementos. Considere-se $r_{i,j}^b$ é a b-ésima reamostragem ($1 \le b \le N$, sendo N o número de simulações). A criação de um triângulo de pseudo-dados $D_{i,j}^b$ é dada pela inversão da transformação realizada em (2.59), ou seja, pela equação (2.60):

$$D_{i,j}^b = r_{i,j}^b \times \sqrt{D_{i,j}^0} + D_{i,j}^0 \tag{2.60}$$

Para cada matriz de pseudo-dados aplica-se o modelo escolhido para estimação das provisões para sinistros e obtém-se as estimativas para as reservas simuladas. Neste sentido, o estimador para a provisão para sinistros relativa ao ano de origem i e ano de desenvolvimento j é dada pela média dos N valores de pseudo-dados:

$$\hat{P}_{i,j} = \frac{1}{N} \sum_{b=1}^{N} D_{i,j}^{b}$$
(2.61)

Para determinar o Erro Quadrático Médio (EQM) de cada estimador, relembremos a equação (2.48):

$$\widehat{EQM}(\hat{P}_{i,j}) = \mathbb{V}[\hat{P}_{i,j}] + \mathbb{V}[P_{i,j}]$$

O cálculo da variância da estimativa $\hat{P}_{i,j}$, segundo England and Verrall [16] é dado pela equação (2.62).

$$\mathbb{V}[\hat{P}_{i,j}] = \frac{m}{m-p} \frac{1}{N} \sum_{b=1}^{N} \left(\hat{D}_{i,j}^{b} - \hat{P}_{i,j} \right)^{2}, \quad n+1 \le i \le I+n, \quad J-i \le j \le J \quad (2.62)$$

com m o número de observações do triângulo de desenvolvimento inicial e p o número de parâmetros a estimar pelo modelo subjacente à estimativa das provisões.

Este cálculo tem o ajuste da multiplicação pelo factor $\frac{m}{m-p}$ pelo facto de na definição do modelo subjacente à técnica de *Bootstrap* existirem parâmetros a estimar, que não são capturados no cálculo da variância das simulações. Esta correção pode ser menosprezada quando m é suficientemente grande.

Relativamente à parcela relativa à variabilidade do processo estocástico subjacente ao modelo, $\mathbb{V}[P_{i,j}]$, esta é obtida através do produto das estimativas analíticas do modelo de Poisson sobredispersão e o parâmetro de escala definido pela equação (2.63):

$$\hat{\phi} = \frac{1}{m-1} \sum_{i=0}^{m} \sum_{j=0}^{I-i} \left(r_{i,j}^{P} \right)^{2}$$
(2.63)

Neste sentido, o EQM para cada estimador da provisão, é dado pela equação (2.64):

$$\widehat{EQM}(\hat{P}_{i,j}) = \frac{m}{m-p} \frac{1}{N} \sum_{b=1}^{N} \left(\hat{D}_{i,j}^{b} - \hat{P}_{i,j} \right)^{2} + \hat{\phi} \hat{P}_{i,j}, \quad n+1 \le i \le I+n, \quad J-i \le j \le J$$
(2.64)

com m o número de observações do triângulo de desenvolvimento inicial, p o número de parâmetros a estimar pelo modelo subjacente à estimativa das provisões e $\hat{\phi}$ dado pela equação (2.63). Recorde-se que b representa a b-ésima reamostragem da aplicação da metodologia.

Para as estimativas das provisões para sinistros por ano de origem, ano de calendário ou provisão total, o raciocínio será semelhante, pelo que é necessário fazer as devidas adaptações às equações (2.61) e (2.64).

Esta metodologia tem grande aplicabilidade, no entanto pode gerar resultados não satisfatórios devido a alguns fatores, tais como [38]:

- à presença de dados incompletos ou incorretos, uma vez que estes erros serão propagados pelo *Bootstrap*;
- à existência de dados dependentes, pois põe em causa o pressuposto de independência para aplicação da metodologia;
- à presença de *outliers*, que influenciarão os resultados finais, pelo que é aconselhável a sua remoção ou correção.

Neste sentido, antes de aplicar esta metodologia é necessário proceder a uma análise dos dados e fazer os ajustamentos necessários, se for adequado.

3. Resultados

Para a aplicação prática do problema proposto, utilizaram-se os dados históricos relativos ao ramo automóvel em Portugal disponibilizados pela Autoridade de Supervisão de Seguros e Fundos de Pensões (ASF). No entanto, optou-se por não integrar o período de 2020 e 2021 devido à menor taxa de sinistralidade verificada nestes anos condicionada pelas medidas de restrição da circulação decretadas pelo Governo Português para atender à situação epidemiológica provocada pela COVID-19. Segundo o relatório da Autoridade Nacional de Segurança Rodoviária [39], o número de acidentes com vitimas diminuiu 25.8% em 2020 e aumentou 10.2% em 2021 (mas comparativamente com 2019, o número de acidentes com vítimas é 18.2% inferior).

Tabela 3.1: Dados das indemnizações do ramo automóvel desde 2001 a 2010 (u.m.: EUR).

Ano de ocorrência	Ano de envolvimento 0	1	2	3	4	5	6	7	8	9
2001	642,807	300,894	81,270	61,298	34,405	27,771	26,792	18,115	13,800	12,489
2002	691,972	282,517	80,695	53,565	35,697	29,386	23,777	15,973	14,243	
2003	631,916	292,534	76,387	53,402	34,243	24,975	23,432	17,758		
2004	619,485	263,655	71,484	49,144	29,212	25,217	18,942			
2005	584,169	266,338	73,846	48,174	30,728	23,894				
2006	563,413	272,327	68,837	42,771	25,322					
2007	581,787	269,694	61,603	36,534						
2008	618,601	263,429	50,020							
2009	631,741	266,284								
2010	675,985									

Neste sentido, utilizaram-se dois triângulos de desenvolvimento: um com os dados históricos relativos ao ramo automóvel em Portugal desde 2001 a 2010 e outro triângulo de desenvolvimento com os dados históricos desde 2006 a 2015. O objetivo será efetuar comparações entre as estimativas produzidas com a aplicação dos modelos referidos no capítulo 2, bem como efetuar comparações entre as estimativas e o triângulo de desenvolvimento com dados reais desde 2016

a 2019, sendo que para efetuar esta última comparação, usando o primeiro conjunto de dados estimam-se 9 anos e utilizando o segundo conjunto estimam-se 4 anos. Os dados históricos foram obtidos a partir dos relatórios "Estatísticas de Seguros" da Autoridade de Supervisão de Seguros e Fundos de Pensões (ASF), produzidos em 2011, 2016 e 2020, sendo que os dados mais recentes podem ser consultados na *webpage* da ASF. Os dados históricos incrementais utilizados encontram-se nas tabelas 3.1 e 3.2.

Tabela 3.2: Dados das indemnizações do ramo automóvel desde 2006 a 2015 (u.m.: EUR).

Ano de ocorrência	Ano de desenvolvimento	0	1	2	3	4	5	6	7	8	9
	2006	576,491	276,707	69,911	43,789	25,993	18,160	17,504	14,181	9,310	6,339
	2007	594,609	273,732	62,543	36,958	23,937	21,424	17,380	13,260	9,200	
	2008	635,360	268,027	51,725	34,204	19,061	18,318	17,854	11,128		
	2009	648,292	270,349	53,717	34,706	25,182	20,066	12,841			
	2010	692,857	297,691	51,297	31,634	25,066	21,941				
	2011	689,464	256,924	43,047	35,534	19,729					
	2012	639,390	221,623	40,480	30,952						
	2013	613,349	218,607	41,688							
	2014	609,027	217,167								
	2015	617,926									

Numa primeira análise, quando se comparam os dois triângulos de desenvolvimento com os dados históricos, verifica-se que existem pequenas diferenças nos anos em comum (2006 a 2010), num total de cerca de 99 mil euros (valores superiores no triângulo de desenvolvimento mais recente). Isto pode ter acontecido devido a acertos e correções feitas após 2010. Os valores comparados encontram-se marcados a cinzento claro nas tabelas 3.1 e 3.2.

Fazendo uma análise gráfica aos dados incrementais e acumulados, obtém-se os gráficos das figuras 3.1 e 3.2. O gráfico dos pagamentos incrementais parece bastante semelhante ao longo do período de desenvolvimento, verificando-se um decrescimento aparentemente exponencial. Efetuando uma análise aos pagamentos acumulados, verificam-se que os anos entre 2001 a 2003 (linhas 1 a 3 da figura 3.1) apresentam indemnizações acumuladas superiores em relação aos anos seguintes (valores superiores a 1.1 milhões de euros). Isto poderá sugerir que nas estimativas a realizar, as produzidas com os dados de 2001 a 2010 sejam superiores às produzidas com os dados de 2006 a 2015. De salientar que as linhas 6 a 9 da figura 3.1 corresponde aos anos de

2006 a 2009, cujas linhas na figura 3.2 corresponde às linhas 1 a 4.

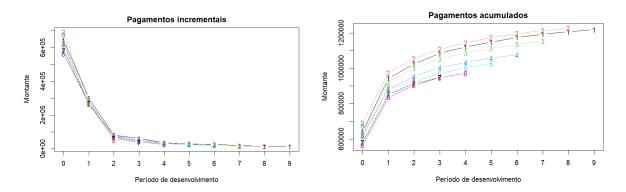


Figura 3.1: Pagamentos incrementais e acumulados entre 2001 e 2010 sem considerar a inflação (u.m.: EUR).

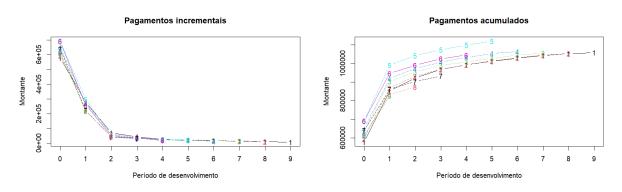


Figura 3.2: Pagamentos incrementais e acumulados entre 2006 e 2015 sem considerar a inflação (u.m.: EUR).

Permitindo a incorporação da inflação passada (refletida no gráfico da figura 2.2 presente no capítulo 2), os gráficos anteriores (figuras 3.1 e 3.2) transformam-se nas figuras 3.3 e 3.4. Verifica-se que ao incorporar a inflação, ou seja, calculando o valor das indemnizações para o último ano conhecido, alterou-se a ordem da evolução das indemnizações em alguns anos de ocorrência. Por exemplo, o ano de 2011 aparentava ser um ano com custos elevados pela figura 3.2 (linha 6), mas ao aplicar a inflação, os custos ficaram ao mesmo nível dos anos de 2007 a 2009 (linhas 2 a 4), como se observa na figura 3.4. No entanto, para outros anos de ocorrência manteve-se o *ranking* que já existia antes da incorporação da inflação, como por exemplo, o ano de 2010 (representado na linha 5 dos gráficos das figuras 3.2 e 3.4) apresenta as indemnizações mais elevadas em ambas as figuras. Ressalva-se que os gráficos das figuras 3.3 e 3.4 não são diretamente comparáveis, uma vez que no primeiro os valores das indemnizações tem a inflação histórica até 2010 e no segundo os valores das indemnizações incorporam a inflação até 2015. No

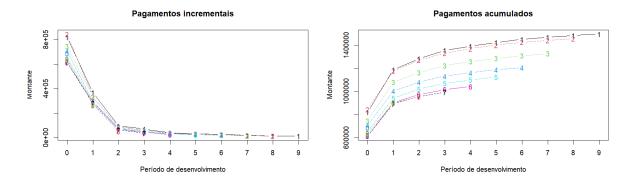


Figura 3.3: Pagamentos incrementais e acumulados entre 2001 e 2010 considerando a inflação (u.m.: EUR).

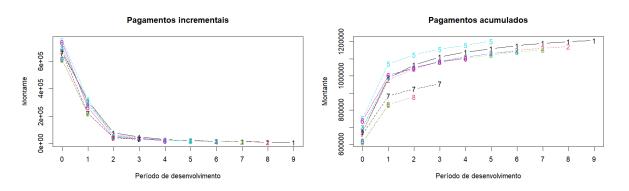


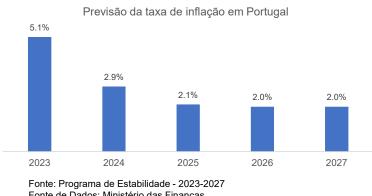
Figura 3.4: Pagamentos incrementais e acumulados entre 2006 e 2015 considerando a inflação (u.m.: EUR).

entanto, como os dois gráficos têm os anos de 2006 a 2009 em comum (linhas 6 a 9 da figura 3.1 e linhas 1 a 4 da figura 3.2) é possível visualizar que os primeiros anos apresentam custos com sinistros superiores, nomeadamente nos anos de 2001 a 2005. Os anos de 2012 a 2014 (linhas 7 a 9 da figura 3.4) apresentam indemnizações inferiores, pelo que é de esperar que ao usar os dados de 2006 a 2015, as estimativas produzidas sejam inferiores às estimativas utilizando os dados de 2001 a 2010.

Tanto da análise dos triângulos de desenvolvimento como da análise dos gráficos anteriores, é possível ver que os dados têm um comportamento idêntico por ano de origem, não existindo valores negativos, nem registos em falta, pelo que não existe impedimento para a aplicação de qualquer um dos métodos mencionados no capítulo 2. No entanto ter-se-á que validar os restantes pressupostos de cada modelo.

Para a produção das estimativas das provisões para sinistros contemplando a inclusão da inflação histórica passada nos dados, também se teve em atenção a inclusão da inflação futura. Neste sentido, utilizaram-se as projeções elaboradas pelo XXIII Governo Constitucional da

República Portuguesa apresentadas no Programa de Estabilidade 2023-2027 [4], tendo como estimativa de longo prazo a inflação a 2%. A figura 3.5 reflete as previsões mencionadas.



Fonte de Dados: Ministério das Finanças

Figura 3.5: Previsão da taxa de inflação em Portugal no período de 2023 a 2027 [4].

Nas próximas secções deste capítulo serão apresentados os resultados da aplicação dos modelos anteriormente mencionados aos dados já referidos, efetuando uma comparação dos resultados:

- usando o mesmo modelo mas tendo como base de projeção diferentes triângulos de desenvolvimento, e;
- usando vários modelos partindo do mesmo triângulo de desenvolvimento.

3.1 Modelo Loglinear

Antes da aplicação da metodologia Loglinear é recomendado efetuar uma análise aos dados por ano de origem, ano de desenvolvimento e ano de calendário para perceber quais as variáveis que irão influenciar mais o modelo. Como a metodologia foi aplicada aos dois triângulos das tabelas 3.1 e 3.2 e várias vezes (pois foi usado um processo iterativo como explicado no capítulo 2) até chegar ao ano de projeção de pagamento de 2019, apenas se irá apresentar o estudo feito para a primeira iteração e para os dados relativos ao período de 2006 a 2015 considerando a inflação, como exemplo ilustrativo da análise feita em cada iteração para cada conjunto de dados (incluindo ou excluindo a inflação). Os gráficos apresentados nas figuras 3.6, 3.7 e 3.8 encontram-se na escala logarítmica.

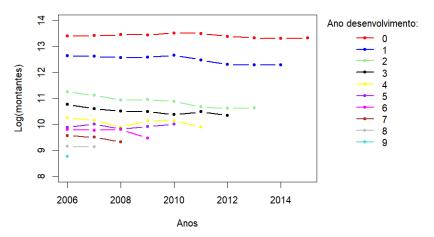


Figura 3.6: Distribuição por ano de origem dos dados das indemnizações relativas ao período de 2006 a 2015 considerando a inflação.

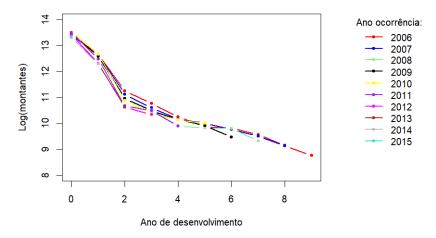


Figura 3.7: Distribuição por ano de desenvolvimento dos dados das indemnizações relativas ao período de 2006 a 2015 considerando a inflação.

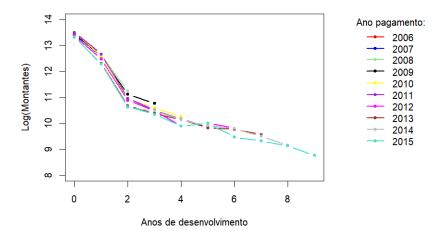


Figura 3.8: Distribuição por ano de calendário dos dados das indemnizações relativas ao período de 2006 a 2015 considerando a inflação.

O gráfico da figura 3.6 mostra que cada ano de desenvolvimento tem um nível diferente, pelo que a inclusão de um parâmetro diferente para cada ano de desenvolvimento deverá ser incluído no modelo. Analisando os gráficos das figuras 3.7 e 3.8 verifica-se que os montantes pagos, em escala logarítmica, têm um comportamento aproximadamente linear por ano de desenvolvimento a partir do segundo o ano, acontecendo o mesmo para os dados por ano de calendário. No entanto, nem todos os anos se caracterizam pelo mesmo fator de linearidade, pelo que possivelmente existirão parâmetros que deverão ser incluídos e outros que poderão ser excluídos.

Aplicando então o modelo de regressão linear múltipla aos dados incrementais na escala logarítmica, obteve-se o modelo presente na tabela 3.3. Ao analisar esta tabela constata-se que o modelo obtido ajusta-se bem aos dados pois o valor de ajustamento R^2 é de 99.49% (o que significa que o modelo explica 99.49% a variabilidade nos dados) e o valor do p-value da estatística F é muito próximo de zero, não existindo assim informação que permita rejeitar o modelo ajustado. Relativamente aos coeficientes do modelo, e considerando como referência o nível de significância $\alpha=10\%$, apenas um coeficiente (a_1) não é considerado significativamente diferente de zero (o p-value da estatística t é superior aos nível de significância de 10%). Uma vez que a estimativa para o coeficiente em causa é tão diminuta, não terá impacto no modelo, pelo que não se optou pela sua exclusão. Recorde-se que neste modelo considera-se a_0 e b_0 são nulos.

Numa primeira análise dos resíduos verifica-se que estes têm média igual a zero. Vejamos se estes cumprem a independência entre si e se seguem uma distribuição Normal. Para isso,

¹Foi explorada a possibilidade de se excluírem as variáveis não significativas. No entanto, quando se utilizam os dados sem incorporar a inflação, os valores estimados para a provisão para sinistros repetiam-se ao longo do triângulo inferior da matriz de desenvolvimento, o que não é realista. Tendo em conta estes resultados e o valor reduzido das estimativas para as variáveis não significativas, optou-se por manter todas as variáveis nos quatro cenários em estudo.

Tabela 3.3: Resultado do ajustamento do modelo Loglinear aplicado aos dados incrementais de 2006 a 2015 com inclusão da inflação.

Resíduos:

Minimo	1º Quartil	Média	Mediana	3º Quartil	Máximo
-0.20407	-0.04075	0	0.01117	0.04838	0.1947

Coeficientes:

Cocheiente					
	Estimativa	Erro padrão	Valor da	Pr(> t)	
			estatística t		
k	13.62447	0.05096	267.380	<2e-16	***
a_1	-0.04609	0.04974	-0.927	0.360279	
a_2	-0.15780	0.05202	-3.034	0.004464	**
a_3	-0.14721	0.05450	-2.701	0.010472	*
a_4	-0.11035	0.05746	-1.920	0.062756	
a_5	-0.23544	0.06121	-3.846	0.000471	***
a_6	-0.34175	0.06631	-5.154	9.38E-06	***
a_7	-0.37000	0.07387	-5.009	1.46E-05	***
a_8	-0.35294	0.08685	-4.064	0.000250	***
a_9	-0.28536	0.11717	-2.435	0.019953	*
b_1	-0.92805	0.04974	-18.660	<2e-16	***
b_2	-2.56373	0.05202	-49.288	<2e-16	***
b_3	-2.96881	0.05450	-54.470	<2e-16	***
b_4	-3.43274	0.05746	-59.739	<2e-16	***
b_5	-3.60993	0.06121	-58.972	<2e-16	***
b_6	-3.82965	0.06631	-57.757	<2e-16	***
b_7	-4.09606	0.07387	-55.450	<2e-16	***
b_8	-4.46504	0.08685	-51.408	<2e-16	***
b_9	-4.86501	0.11717	-41.522	<2e-16	***

Nível de significância: '***'0.1% '**'1% '*'5% '.'10%

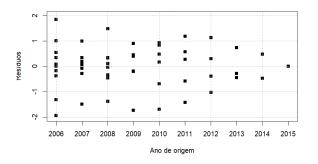
Erro padrão dos resíduos: 0.1055 com 36 graus de liberdade

 R^2 : 0.9966, R^2 ajustado: 0.9949

Estatística F: 584.5 com 18 e 36 graus de liberdade, p-value: <2.2e-16

traçaram-se os gráficos dos resíduos para procurar alguma tendência (figuras 3.9, 3.10, 3.11 e 3.12), onde se inferiu a não existência de tendência nos resíduos. Elaborou-se também o gráfico quantil-quantil (presente na figura 3.13) e realizou-se o teste estatístico de *Kolmogorov–Smirnov*, obtendo-se um valor de *p-value* igual a 0.24 (superior aos nível de significância de 10%), e por isso não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, apesar do gráfico quantil-quantil presente na figura 3.13 evidenciar alguns *outliers* face à distribuição normal.

Neste sentido, pode-se dizer que o modelo Loglinear é um modelo que se ajusta aos dados e sendo assim, calcular-se-á as estimativas para os valores das provisões para sinistros para o primeiro ano após o último conhecido (na primeira iteração da metodologia) ou o último



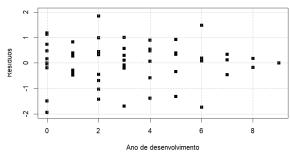
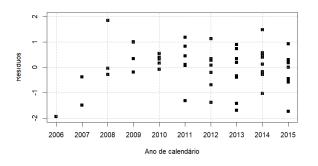


Figura 3.9: Gráfico dos resíduos do modelo por ano de origem.

Figura 3.10: Gráfico dos resíduos do modelo por ano de desenvolvimento.



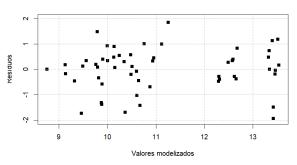


Figura 3.11: Gráfico dos resíduos do modelo por ano de calendário.

Figura 3.12: Gráfico dos resíduos *versus* valores modelizados.

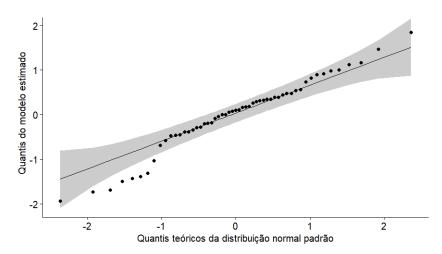


Figura 3.13: Gráfico quantil-quantil dos resíduos do modelo Loglinear aplicado aos dados das indemnizações relativas ao período de 2006 a 2015 considerando a inflação.

calculado (no caso das iterações posteriores à primeira). Uma vez que se está a usar os dados de 2006 a 2015, as estimativas da primeira iteração são os valores da diagonal correspondente ao ano de pagamento em 2016, que se encontram na tabela 3.4. Nesta tabela também está o erro padrão da estimativa.

Tabela 3.4: Resultados da primeira iteração da metodologia para o ano de calendário de 2016 com a aplicação do modelo Loglinear aos dados de 2006 a 2015 considerando a inflação (u.m.: EUR).

Ano de origem	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007
Ano de desenvolvimento	0	1	2	3	4	5	6	7	8	9
Estimativa	836,779	250,048	45,396	29,736	19,227	17,910	16,294	12,036	8,242	6,196
Erro padrão	98,380	39,572	6,251	3,889	2,460	2,277	2,085	1,574	1,135	981

Incorporando os resultados da primeira iteração do processo no triângulo de desenvolvimento, obtém-se um novo triângulo de desenvolvimento apresentado na tabela 3.5, onde os resultados obtidos correspondem à primeira diagonal estimada. Os dados estimados encontram-se na zona sombreada a azul.

Tabela 3.5: Triângulo (ou matriz) de desenvolvimento com a inclusão dos resultados da primeira iteração da aplicação do modelo Loglinear aos dados relativos ao período 2006-2015, considerando a inflação (u.m.: EUR).

Ano de ocorrência	Ano de nvolvimento 0	1	2	3	4	5	6	7	8	9
2006	673,606	313,599	77,299	47,190	28,238	19,456	18,084	14,252	9,328	6,371
2007	673,886	302,661	67,400	40,150	25,645	22,134	17,467	13,286	9,246	6,196 981
2008	702,507	288,843	56,192	36,645	19,693	18,409	17,889	11,184	8,242 1,135	
2009	698,641	293,695	57,550	35,856	25,308	20,106	12,905	12,036 1,574		
2010	752,689	318,933	52,997	31,792	25,116	22,051	16,294 2,085			
2011	738,662	265,436	43,262	35,605	19,828	17,910 2,277				
2012	660,573	222,729	40,560	31,107	19,227 2,460					
2013	616,410	219,041	41,896	29,736 3,889						
2014	610,236	218,253	45,396 6,251							
2015	621,016	250,048 39,572								
2016	836,779 98,380									

Zona sombreada com os valores estimados pelo modelo Loglinear. A azul estão os valores do erro padrão das estimativas.

O processo foi repetido quatro vezes até atingir o ano de calendário de 2019 (sendo que no caso da base de dados desde 2001 a 2010, repetiu-se 9 vezes). Para cada iteração verificaram-se os pressupostos, que se encontram no anexo A.1. Sendo que na última iteração para além dos valores da diagonal da matriz de desenvolvimento do ano de 2019, foram também calculados os valores relativos às provisões para sinistros até ao ano de desenvolvimento 9, por forma a

completar a matriz de desenvolvimento. A tabela 3.6 apresenta o resultado da aplicação da metodologia Loglinear aos dados históricos de 2006 a 2015, incorporando a inflação.

Tabela 3.6: Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do modelo Loglinear aos dados relativos ao período 2006-2015, considerando a inflação (u.m.: EUR).

Ano de desenvolvimento Ano de ocorrência	0	1	2	3	4	5	6	7	8	9
2006	673,606	313,599	77,299	47,190	28,238	19,456	18,084	14,252	9,328	6,371
2007	673,886	302,661	67,400	40,150	25,645	22,134	17,467	13,286	9,246	6,196 981
2008	702,507	288,843	56,192	36,645	19,693	18,409	17,889	11,184	8,242 1,135	5,676 735
2009	698,641	293,695	57,550	35,856	25,308	20,106	12,905	12,036 1,574	8,539 963	5,858 679
2010	752,689	318,933	52,997	31,792	25,116	22,051	16,294 2,085	12,476 1,336	8,906 899	6,136 588
2011	738,662	265,436	43,262	35,605	19,828	17,910 2,277	14,495 1,519	10,970 1,053	7,914 661	5,427 523
2012	660,573	222,729	40,560	31,107	19,227 2,460	16,814 1,750	13,066 1,226	10,064 798	7,182 664	4,989 483
2013	616,410	219,041	41,896	29,736 3,889	18,769 1,966	16,326 1,523	12,139 942	9,737 780	7,039 596	5,203 507
2014	610,236	218,253	45,396 6,251	30,090 3,223	18,996 1,783	16,724 1,289	12,361 972	10,045 814	7,727 662	5,569 547
2015	621,016	250,048 39,572	47,433 5,347	31,908 3,062	20,668 1,604	17,883 1,407	13,390 1,073	11,578 956	8,684 757	6,127 610
2016	836,779 98,380	332,818 43,102	61,517 6,213	41,966 3,329	27,297 2,186	23,926 1,941	19,064 1,575	16,072 1,365	11,802 1,055	8,263 839
2017	818,675 78,838	319,461 37,055	58,608 4,893	40,366 3,394	26,598 2,256	24,809 2,128	19,272 1,680	15,908 1,422	11,590 1,086	8,107 853
2018	741,786 64,002	285,184 27,337	52,571 5,063	36,679 3,551	25,719 2,506	23,388 2,299	17,788 1,770	14,569 1,479	10,604 1,116	7,41′ 858
2019	746,185 53,230	284,712 20,310	53,165 3,803	39,473 2,837	26,984 1,953	24,024 1,757	18,129 1,347	14,833 1,131	10,796 865	7,549 686

Zona sombreada com os valores estimados pelo modelo Loglinear. A azul estão os valores do erro padrão das estimativas.

As tabelas seguintes, 3.7, 3.8 e 3.9, apresentam os resultados da aplicação do modelo Loglinear para o triângulo de desenvolvimento com os dados de 2006 a 2015 não incorporando a inflação e para o triângulo de desenvolvimento com os dados de 2001 a 2010 (considerando e não considerando a inflação), respetivamente.

Ao efetuar as comparações entre as tabelas com os mesmos dados históricos, isto é, comparando a tabela 3.6 com a tabela 3.7 e comparando a tabela 3.8 com a tabela 3.9, é evidente o valor superior tanto nas estimativas das provisões para sinistros, como no erro padrão, nas tabelas considerando a inflação histórica (até 2022) e futura (de 2023 a 2028).

Quando se compara os resultados das tabelas 3.6 com 3.8 e 3.7 com 3.9, ou seja, resultados baseados em triângulos de desenvolvimento diferentes, mas com o mesmo pressuposto face

Tabela 3.7: Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do modelo Loglinear aos dados relativos ao período 2006-2015, não incorporando a inflação (u.m.: EUR).

desenvolvimento Ano de ocorrência	0	1	2	3	4	5	6	7	8	9
2006	576,491	276,707	69,911	43,789	25,993	18,160	17,504	14,181	9,310	6,339
2007	594,609	273,732	62,543	36,958	23,937	21,424	17,380	13,260	9,200	6,236 1,011
2008	635,360	268,027	51,725	34,204	19,061	18,318	17,854	11,128	8,372 1,181	5,722 740
2009	648,292	270,349	53,717	34,706	25,182	20,066	12,841	12,316 1,650	8,662 975	5,924 667
2010	692,857	297,691	51,297	31,634	25,066	21,941	16,663 2,183	12,732 1,361	9,124 894	6,25° 574
2011	689,464	256,924	43,047	35,534	19,729	18,295 2,381	14,906 1,559	11,326 1,054	8,213 656	5,61 518
2012	639,390	221,623	40,480	30,952	19,662 2,576	17,364 1,804	13,608 1,239	10,615 806	7,584 610	5,18 481
2013	613,349	218,607	41,688	30,428 4,075	19,308 2,019	17,061 1,543	12,823 952	10,404 798	7,434 603	5,08 474
2014	609,027	217,167	46,110 6,502	30,633 3,275	19,686 1,792	17,570 1,297	13,126 988	10,651 827	7,610 625	5,20 490
2015	617,926	251,646 40,785	47,852 5,385	32,610 3,035	21,525 1,599	18,839 1,418	14,074 1,080	11,420 903	8,159 681	5,58 532
2016	725,304 87,318	288,677 37,317	53,770 5,268	37,428 2,842	24,696 1,893	21,615 1,678	16,148 1,277	13,102 1,065	9,361 801	6,40 622
2017	719,266 69,138	281,120 31,629	52,333 4,182	36,637 2,948	24,174 1,962	21,158 1,737	15,806 1,319	12,825 1,098	9,163 821	6,269
2018	668,571 55,958	258,556 23,721	48,337 4,455	33,840 3,136	22,328 2,083	19,542 1,839	14,599 1,391	11,846 1,151	8,464 853	5,79 641
2019	689,831 47,103	265,662 18,140	49,664 3,401	34,768 2,392	22,940 1,590	20,077 1,406	14,999 1,067	12,169 889	8,694 667	5,94 517

Zona sombreada com os valores estimados pelo modelo Loglinear. A azul estão os valores do erro padrão das estimativas.

à inflação, existe uma predominância de estimativas superiores quando se usa o triângulo de desenvolvimento com os dados relativos ao período de 2001 a 2010, sendo exceção para os anos de desenvolvimento 0 e 1. O mesmo sucede com o erro padrão.

Destaca-se também que à medida que se vão efetuando novas iterações, o valor do erro padrão diminui, o que se atendermos ao contexto de cálculo de provisões não era expectável uma vez que existe maior incerteza quanto mais distante for o horizonte temporal da estimativa. No entanto, em cada iteração adiciona-se a diagonal calculada na iteração anterior que já tem incorporada o modelo explicativo, razão pela qual o valor do erro padrão do modelo vai diminuindo. Isto pode ser verificado no anexo A.1.

Tabela 3.8: Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do modelo Loglinear aos dados relativos ao período 2001-2010, considerando a inflação (u.m.: EUR).

Ano de desenvolvimento Ano de ocorrência	0	1	2	3	4	5	6	7	8	9
2001	819,739	367,543	95,822	70,033	38,386	30,288	28,342	18,695	13,881	12,664
2002	845,246	333,104	92,194	59,764	38,932	31,086	24,539	16,067	14,442	12,519 1,331
2003	745,065	334,219	85,226	58,242	36,224	25,775	23,570	18,007	13,556 1,255	12,26° 1,217
2004	707,760	294,165	77,963	51,987	30,148	25,365	19,207	15,263 1,342	12,492 1,079	11,01: 1,004
2005	651,769	290,478	78,117	49,718	30,909	24,229	21,242 1,827	15,186 1,247	12,241 971	10,920 932
2006	614,478	288,079	71,043	43,023	25,677	22,765 1,946	18,968 1,523	14,102 1,064	11,224 834	10,04 799
2007	615,439	278,335	61,966	37,045	26,915 2,315	21,614 1,725	17,698 1,306	13,123 927	10,636 737	9,561 709
2008	638,422	264,980	50,720	42,514 3,738	26,411 2,121	20,624 1,512	17,106 1,182	12,935 852	10,449 675	9,497 644
2009	635,461	270,012	70,265 6,503	46,547 3,821	28,560 2,107	23,075 1,585	19,507 1,257	14,491 889	11,792 697	10,72 925
2010	685,449	310,402 32,991	77,245 6,674	50,371 3,799	31,067 2,147	25,549 1,636	21,383 1,284	16,065 902	13,047 888	11,84 243
2011	917,994 72,594	412,259 40,898	99,995 7,938	65,530 4,629	41,426 2,670	33,919 2,024	28,439 1,562	21,318 1,034	17,267 309	15,66 323
2012	858,220 63,356	381,250 34,748	91,817 6,827	60,703 3,999	38,319 2,301	31,634 1,727	26,383 1,252	19,723 336	15,998 288	14,70 305
2013	783,456 53,146	345,407 29,484	83,486 5,788	54,953 3,373	35,643 1,958	28,766 1,356	24,064 400	17,981 309	14,775 269	14,45 302
2014	696,686 44,266	311,505 24,794	73,773 4,765	49,718 2,792	31,722 1,505	25,653 424	21,490 363	16,267 283	14,224 262	13,56 286
2015	693,658 41,101	310,443 23,019	73,455 4,341	50,423 2,445	31,424 523	25,599 432	21,723 374	17,499 310	14,918 279	13,92 297
2016	636,083 35,114	285,335 19,357	67,609 3,451	45,745 778	28,760 494	23,733 413	21,432 380	16,832 307	14,049 269	13,01 283
2017	604,939 30,557	269,892 15,810	65,645 1,175	43,682 788	27,820 506	24,430 449	21,510 402	16,539 317	13,697 275	12,67 286
2018	582,704 25,420	259,859 5,339	63,120 1,303	42,548 883	28,836 602	24,689 520	21,282 454	16,237 353	13,434 303	12,43 308
2019	634,512 9,711	282,058 4,317	69,403 1,065	49,785 768	32,896 511	27,575 433	23,585 376	17,976 294	14,873 256	13,76 268

Zona sombreada com os valores estimados pelo modelo Loglinear. A azul estão os valores do erro padrão das estimativas.

Relativamente à análise da aderência da estimativa à realidade, foi necessário extrair os valores efetivamente pagos desde 2016 a 2019 através da informação disponibilizada no *website* do supervisor de seguros [40] e fazer o cálculo dos intervalos de confiança a 95% para cada valor estimado. Os resultados encontram-se na tabela 3.10 e nos gráficos da figura 3.14. A tabela 3.10 apresenta o valor real de pagamentos por ano de calendário e ano de desenvolvimento e o valor total do ano de pagamento, bem como as várias estimativas e respetivos intervalos de confiança

(considerando o pressuposto de inflação e os dados históricos). Os intervalos de confiança que não contêm o valor real encontram-se marcados a *bordeaux*.

Tabela 3.9: Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do modelo Loglinear aos dados relativos ao período 2001-2010, não incorporando a a inflação (u.m.: EUR).

Ano de desenvolvimento de ocorrência	0	1	2	3	4	5	6	7	8	9
2001	642,807	300,894	81,270	61,298	34,405	27,771	26,792	18,115	13,800	12,489
2002	691,972	282,517	80,695	53,565	35,697	29,386	23,777	15,973	14,243	12,22° 1,428
2003	631,916	292,534	76,387	53,402	34,243	24,975	23,432	17,758	13,458 1,369	11,91- 1,258
2004	619,485	263,655	71,484	49,144	29,212	25,217	18,942	15,224 1,471	12,420 1,142	10,86 1,056
2005	584,169	266,338	73,846	48,174	30,728	23,894	21,223 2,006	15,307 1,337	12,474 1,056	11,01 991
2006	563,413	272,327	68,837	42,771	25,322	22,738 2,136	19,215 1,642	14,558 1,171	11,647 913	10,26 862
2007	581,787	269,694	61,603	36,534	26,891 2,542	21,946 1,863	18,468 1,453	13,829 1,030	11,198 819	9,901 750
2008	618,601	263,429	50,020	42,407 4,098	26,747 2,285	21,553 1,685	18,096 1,318	13,735 954	11,123 734	9,838 670
2009	631,741	266,284	69,758 7,096	46,832 4,091	29,693 2,336	24,418 1,767	20,713 1,408	15,537 974	12,535 743	11,09 644
2010	675,985	303,161 35,415	76,128 6,999	51,204 4,119	32,246 2,349	26,601 1,797	22,544 1,383	16,949 954	13,719 694	12,13 232
2011	716,116 62,227	319,240 33,704	79,538 6,734	53,285 3,968	33,920 2,306	28,053 1,710	23,668 1,304	17,828 857	14,410 240	12,75 244
2012	704,623 55,350	316,933 30,809	77,927 6,107	51,995 3,613	33,103 2,031	27,522 1,507	23,222 1,092	17,501 277	14,143 237	12,51 241
2013	683,359 49,436	306,182 27,550	74,742 5,467	49,709 3,118	32,363 1,784	26,549 1,241	22,472 347	16,903 270	13,659 231	12,09 235
2014	626,417 41,952	282,130 23,689	67,500 4,456	45,530 2,565	29,516 1,388	24,278 373	20,537 322	15,448 250	12,483 213	11,04 216
2015	631,863 39,493	284,011 21,521	67,227 3,986	46,670 2,243	29,682 459	24,480 384	20,708 331	15,576 256	12,588 219	11,14 221
2016	591,927 33,392	263,776 17,953	62,836 3,179	43,301 684	27,663 441	22,815 369	19,299 317	14,517 246	11,731 209	10,38 210
2017	571,457 28,959	253,381 14,713	62,357 1,037	42,117 705	26,907 454	22,191 379	18,772 326	14,120 251	11,411 213	10,10 212
2018	564,101 24,392	250,867 4,787	61,643 1,182	41,635 802	26,599 516	21,938 429	18,557 368	13,958 282	11,280 236	9,98 ² 230
2019	627,348 8,917	4,787 278,944 3,965	68,542 977	46,295 663	29,576 427	24,393 356	20,634 306	15,520 236	12,542 200	11,10 201

Zona sombreada com os valores estimados pelo modelo Loglinear. A azul estão os valores do erro padrão das estimativas.

Da análise dos resultados constata-se que a melhor estimativa para as provisões totais para os anos de 2016 a 2019, é o modelo que não tem em consideração a inflação e que usa os dados menos recentes. Também é possível verificar que as estimativas produzidas pelo modelo considerando a inflação revelam números bastantes superiores aos reais (mesmo que o valor real

esteja contido no intervalo de confiança da estimativa).

Relativamente à comparação entre o valor real e as várias estimativas para cada ano de calendário e ano de desenvolvimento, não existe um padrão. No entanto, verifica-se que para os anos de calendário (ou pagamento) de 2016, 2017 e 2019, o modelo sem considerar a inflação e utilizando os anos mais recentes apresenta melhores estimativas (na maioria dos anos de desenvolvimento). Ressalva-se ainda que o modelo utilizando os dados mais antigos (considerando ou não a inflação) para os vários anos de calendário não se adequa uma vez que na maioria dos anos de desenvolvimento o intervalo de confiança não contempla o valor real pago. Para o ano de calendário de 2018 e ano de desenvolvimento posterior ao ano 2, o modelo que apresenta menores diferenças (na sua maioria) face a cada valor pago por ano de desenvolvimento é o modelo considerando a inflação e os dados mais recentes. Em todo o caso as generalizações aqui apresentadas não coincidem para os anos de desenvolvimento 0 e 1. Relativamente aos últimos anos de desenvolvimento, verifica-se que quanto mais iterações se calcular, menor é o erro padrão (como visto anteriormente), o que provoca que o valor real deixe de estar dentro do intervalo de confiança do modelo Loglinear para os anos mais recentes. Por exemplo, para o ano de calendário de 2019 e anos de desenvolvimento 7, 8 e 9, o valor real não está incluído em nenhum intervalo de confiança das quatro bases de dados em estudo utilizando a metodologia Loglinear.

Ao avançar para a análise dos dados acumulados, presentes nos gráficos da figura 3.14 verificase que os modelos que mais se ajustam são os modelos sem inflação, tendo melhor aderência à
realidade quando se usa a base de dados menos recente. Ao utilizar o triângulo de desenvolvimento
com os dados mais recentes, as estimativas produzidas para a provisão para sinistros são sempre
superiores ao valor real, enquanto que utilizando o triângulo de desenvolvimento com os dados
de 2001 a 2010, as estimativas produzidas pelo modelo Loglinear são normalmente inferiores
aos valores reais nos primeiros anos de desenvolvimento e superiores nos últimos anos de
desenvolvimento (exceção para o ano de 2017 quando não se incorpora a inflação no modelo, em
que os valores estimados são inferiores aos valores reais em todos os anos de desenvolvimento).
No entanto, a tendência de crescimento após o ano 1 de desenvolvimento é melhor capturada
pelo modelo com os dados de 2006 a 2015.

Tabela 3.10: Comparação dos resultados obtidos através do modelo Loglinear e os valores reais das indemnizações relativas ao período de 2016 a 2019 por ano de calendário.

								Ano de desenvo	lvimento					u.m.: EUR
				0	1	2	3	4	5	6	7	8	9	Total
	Val	lor real		645,259	230,506	46,957	33,058	21,340	19,834	18,389	14,800	8,418	6,965	1,045,526
		.9	I	591,927	284,011	67,500	49,709	33,103	28,053	22,544	15,537	11,123	9,901	1,113,409
		inflação		[526,480; 657,375]	[241,832; 326,191]	[58,768; 76,233]	[43,599; 55,820]	[29,122; 37,083]	[24,701; 31,405]	[19,833; 25,255]	[13,627; 17,447]	[9,684; 12,562]	[8,431; 11,372]	[1,037,901; 1,188,918]
	2	E.	II	725,304	251,646	46,110	30,428	19,662	18,295	16,663	12,316	8,372	6,236	1,135,031
91	iva/	Sem		[554,165; 896,444]	[171,710; 331,583]	[33,366; 58,854]	[22,440; 38,415]	[14,614; 24,710]	[13,627; 22,962]	[12,385; 20,941]	[9,083; 15,549]	[6,058; 10,686]	[4,255; 8,217]	[951,650; 1,318,413]
20	Estimativa/	ão	I	636,083	310,442	73,773	54,953	38,319	33,919	21,383	14,491	10,449	9,561	1,203,374
	Est	inflação		[567,261; 704,905]	[265,327; 355,558]	[64,433; 83,113]	[48,343; 61,563]	[33,810; 42,828]	[29,953; 37,885]	[18,867; 23,899]	[12,748; 16,234]	[9,126; 11,771]	[8,172; 10,951]	[1,123,539; 1,283,209]
		Com ir	II	836,779	250,048	45,396	29,736	19,227	17,910	16,294	12,036	8,242	6,196	1,241,865
		3		[643,958; 1,029,600]	[172,489; 327,607]	[33,144; 57,648]	[22,113; 37,359]	[14,406; 24,048]	[13,448; 22,372]	[12,209; 20,380]	[8,951; 15,121]	[6,018; 10,467]	[4,274; 8,118]	[1,039,613; 1,444,117]
	Val	lor real		658,862	243,478	48,142	31,584	26,567	20,853	20,386	12,467	9,104	7,209	1,078,653
		ão	I	571,457	263,776	67,227	45,530	32,363	27,522	23,668	16,949	12,535	9,838	1,070,865
		inflação		[514,698; 628,217]	[228,589; 298,963]	[59,415; 75,040]	[40,502; 50,557]	[28,867; 35,858]	[24,568; 30,476]	[21,111; 26,225]	[15,078; 18,821]	[11,079; 13,992]	[8,525; 11,150]	[1,006,063; 1,135,667]
	2	Sem ir	II	719,266	288,677	47,851	30,633	19,308	17,364	14,906	12,732	8,662	5,722	1,165,122
2017	Estimativa/	Š		[583,758; 854,775]	[215,536; 361,818]	[37,297; 58,406]	[24,215; 37,052]	[15,351; 23,265]	[13,828; 20,901]	[11,851; 17,961]	[10,064; 15,399]	[6,752; 10,573]	[4,272; 7,171]	[1,015,958; 1,314,285]
70	tima	ĝ	I	604,939	285,335	73,455	49,718	35,643	31,634	28,439	16,065	11,792	9,497	1,146,518
011	Es	Com inflação		[545,048; 664,830]	[247,395; 323,275]	[64,946; 81,964]	[44,246; 55,191]	[31,805; 39,481]	[28,250; 35,019]	[25,377; 31,502]	[14,297; 17,833]	[10,426; 13,158]	[8,234; 10,759]	[1,077,690; 1,215,345]
calendario		imo	II	818,675	332,818	47,433	30,090	18,769	16,814	14,495	12,476	8,539	5,676	1,305,785
cal		ŭ		[664,156; 973,194]	[248,339; 417,297]	[36,952; 57,914]	[23,774; 36,407]	[14,915; 22,623]	[13,383; 20,244]	[11,519; 17,471]	[9,857; 15,095]	[6,652; 10,426]	[4,235; 7,116]	[1,135,245; 1,476,324]
Ano de	Val	lor real		622,202	254,191	45,088	31,407	19,034	16,391	14,973	12,700	8,107	4,272	1,028,365
₹		ĝ	I	564,101	253,381	62,836	46,670	29,516	26,549	23,222	17,828	13,719	11,099	1,048,922
		inflação		[516,294; 611,908]	[224,545; 282,217]	[56,606; 69,066]	[42,273; 51,067]	[26,796; 32,237]	[24,118; 28,981]	[21,082; 25,363]	[16,148; 19,508]	[12,358; 15,079]	[9,836; 12,363]	[994,760; 1,103,084]
	71.	Sem i	II	668,571	281,120	53,770	32,610	19,686	17,061	13,608	11,326	9,124	5,924	1,112,801
810	ıtiva	S		[558,895; 778,248]	[219,128; 343,112]	[43,446; 64,095]	[26,661; 38,559]	[16,173; 23,199]	[14,036; 20,086]	[11,180; 16,036]	[9,260; 13,393]	[7,372; 10,876]	[4,618; 7,231]	[990,785; 1,234,818]
7	Estimativa/	ção	I	582,703	269,892	67,609	50,423	31,722	28,766	26,383	21,318	13,047	10,722	1,102,585
	ш	Com inflação		[532,882; 632,525]	[238,903; 300,880]	[60,846; 74,372]	[45,630; 55,216]	[28,772; 34,671]	[26,108; 31,424]	[23,929; 28,836]	[19,292; 23,344]	[11,742; 14,352]	[9,491; 11,953]	[1,045,645; 1,159,524]
		omo	II	741,786	319,461	61,517	31,908	18,996	16,326	13,066	10,970	8,906	5,858	1,228,794
		1 -		[616,344; 867,228]	[246,834; 392,088]	[49,340; 73,695]	[25,907; 37,908]	[15,502; 22,491]	[13,342; 19,311]	[10,663; 15,470]	[8,907; 13,033]	[7,143; 10,668]	[4,526; 7,189]	[1,088,426; 1,369,162]
	Va	lor real		654,068	264,025	51,182	33,215	20,746	17,832	12,426	13,811	12,612	7,465	1,087,382
		ção	I	627,348	250,867	62,357	43,301	29,682	24,278	22,472	17,501	14,410	12,139	1,104,354
		Sem inflação		[609,871; 644,825]	[241,485; 260,249]	[60,325; 64,389]	[41,960; 44,642]	[28,783; 30,581]	[23,547; 25,009]	[21,791; 23,152]	[16,959; 18,043]	[13,940; 14,879]	[11,685; 12,593]	[1,085,111; 1,123,597]
	VIC	em	II	689,831	258,556	52,333	37,428	21,525	17,570	12,823	10,615	8,213	6,257	1,115,152
2019	ativa	- 01		[597,511; 782,151]	[212,064; 305,048]	[44,137; 60,529]	[31,859; 42,998]	[18,392; 24,658]	[15,029; 20,112]	[10,956; 14,689]	[9,036; 12,195]	[6,926; 9,499]	[5,132; 7,383]	[1,014,873; 1,215,430]
7	Estimativa/	ıção	I	634,512	259,859	65,645	45,745	31,424	25,653	24,064	19,723	17,267	11,849	1,135,740
	H	inflação	L.,	[615,479; 653,546]	[249,395; 270,324]	[63,341; 67,948]	[44,219; 47,270]	[30,399; 32,449]	[24,821; 26,484]	[23,279; 24,849]	[19,065; 20,380]	[16,661; 17,873]	[11,372; 12,326]	[1,114,667; 1,156,813]
		Com	II	746,185	285,184	58,608	41,966	20,668	16,724	12,139	10,064	7,914	6,136	1,205,588
				[641,857; 850,514]	[231,605; 338,763]	[49,018; 68,198]	[35,442; 48,490]	[17,525; 23,811]	[14,197; 19,251]	[10,293; 13,985]	[8,500; 11,629]	[6,619; 9,209]	[4,984; 7,289]	[1,091,805; 1,319,371]

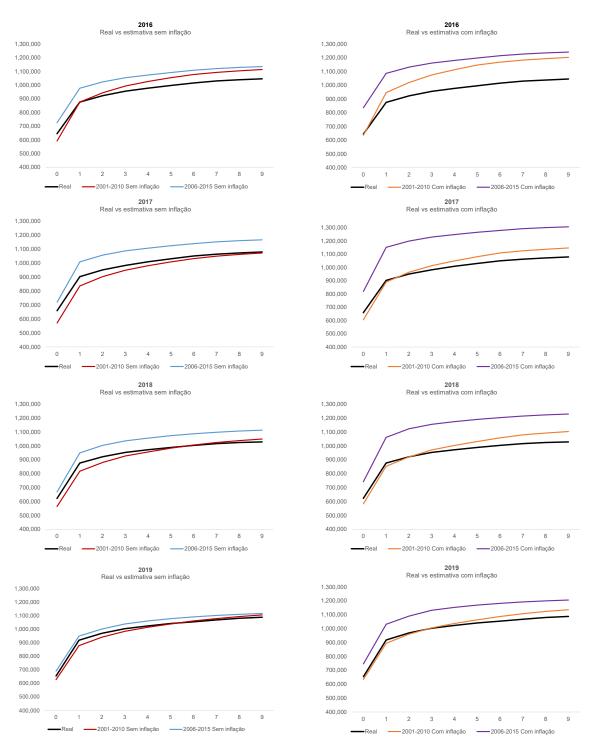


Figura 3.14: Gráficos comparativos entre os valores de pagamento reais e os valores estimados pelo modelo Loglinear no período de 2016 a 2019 por ano de desenvolvimento do sinistro (u.m.: EUR).

3.2 Modelo de Poisson sobre-dispersão

A aplicação do modelo linear generalizado de Poisson sobre-dispersão (adiante designado por ODP) teve por base a mesma lógica do modelo Loglinear, utilizando os mesmos dados e considerando ou desconsiderando a inflação. Dado que o processo é iterativo e para que exista uma base de comparação com a secção anterior, vejamos o exemplo da primeira estimativa obtida por este modelo com os dados históricos do período de 2006 a 2015 considerando a inflação. Os resultados da aplicação do modelo encontram-se na tabela 3.11.

Tabela 3.11: Resultado do ajustamento do modelo ODP aplicado aos dados incrementais de 2006 a 2015 com inclusão da inflação.

Resíduos:

Miı	nimo	1º Quartil	Média	Mediana	3º Quartil	Máximo
-1.8	5853	-0.51819	-0.00431	0	0.40324	2.3143

Coeficientes:

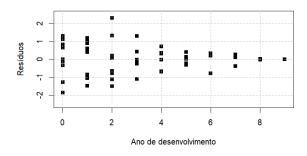
Cocheien	29•				
	Estimativa	Erro padrão	Valor da estatística <i>t</i>	Pr(> t)	
$\overline{\gamma}$	13.5005	0.0337	400.69	< 2e-16	***
$lpha_1$	-0.0246	0.0459	-0.54	0.59573	
$lpha_2$	-0.0344	0.0462	-0.74	0.46138	
$lpha_3$	-0.0296	0.0463	-0.64	0.52611	
$lpha_4$	0.0356	0.0458	0.78	0.44229	
$lpha_5$	-0.0340	0.0468	-0.73	0.47305	
$lpha_6$	-0.1563	0.0487	-3.21	0.0028	**
$lpha_7$	-0.2063	0.0500	-4.13	0.00021	***
$lpha_8$	-0.2087	0.0509	-4.10	0.00022	***
$lpha_9$	-0.1614	0.0561	-2.88	0.00674	**
eta_1	-0.9194	0.0268	-34.34	< 2e-16	***
eta_2	-2.5385	0.0555	-45.71	< 2e-16	***
eta_3	-2.9511	0.0713	-41.39	< 2e-16	***
eta_4	-3.4017	0.0947	-35.92	< 2e-16	***
eta_5	-3.5654	0.1120	-31.82	< 2e-16	***
eta_6	-3.7621	0.1387	-27.13	< 2e-16	***
eta_7	-4.0154	0.1810	-22.18	< 2e-16	***
eta_8	-4.3519	0.2608	-16.69	< 2e-16	***
eta_9	-4.741	0.4446	-10.66	1.1e-12	***

Grau de significância: '***'0.1% '**'1% '*'5% '.'10% ' '1

Deviance residual: 44,878 com 36 graus de liberdade

O modelo indica que as variáveis relacionadas com o ano de desenvolvimento são todas explicativas, sendo que os primeiros anos de origem poderiam ser agrupados, pois não apresentam significância estatística (o *p-value* da estatística *t* é superior aos níveis de significância 0.1%,

1%, 5% e 10%). No entanto, dado o valor muito residual das estimativas, optou-se por manter as variáveis.² Para avaliar a adequabilidade do modelo, realizaram-se os testes de normalidade dos resíduos, construiu-se o gráfico quantil-quantil destes e procedeu-se à representação gráfica dos resíduos relativamente aos valores modelizados e às variáveis explicativas. Nas próximas figuras encontram-se os referidos gráficos.

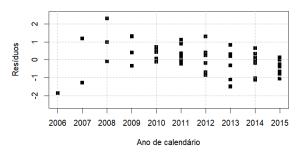


2006 2007 2008 2009 2010 2011 2012 2013 2014 2015

Ano de origem

Figura 3.15: Gráfico dos resíduos do modelo por ano de origem.

Figura 3.16: Gráfico dos resíduos do modelo por ano de desenvolvimento.



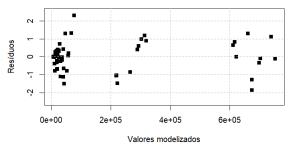


Figura 3.17: Gráfico dos resíduos do modelo por ano de calendário.

Figura 3.18: Gráfico dos resíduos *versus* valores modelizados.

Analisando os gráficos das figuras 3.15 a 3.18, verifica-se que os resíduos têm um comportamento aleatório com exceção do gráfico da figura 3.17 onde parece existir uma tendência decrescente entre os resíduos e as varáveis explicativas relativas ao ano de calendário, no entanto não se exclui a adequação deste modelo. Relativamente à normalidade dos resíduos, o gráfico da figura 3.19 mostra que o comportamento dos resíduos ajusta-se a uma distribuição normal e efetuando o teste de *Kolmogorov–Smirnov* este também indica que os resíduos seguem uma distribuição normal (o valor do *p-value* é superior a 0.05).

Neste sentido, pode-se dizer que o modelo ODP é um modelo que se ajusta aos dados e por isso proceder-se-á à determinação das estimativas para os valores das provisões para sinistros

²Foi explorada a possibilidade de se excluírem as variáveis não significativas. No entanto, quando se utilizam os dados sem incorporar a inflação, os valores estimados para a provisão para sinistros repetiam-se ao longo do triângulo inferior da matriz de desenvolvimento, o que não é realista. Tendo em conta estes resultados e o valor reduzido das estimativas para as variáveis não significativas, optou-se por manter todas as variáveis nos quatro cenários em estudo.

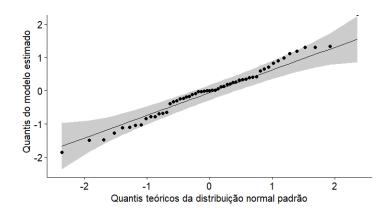


Figura 3.19: Gráfico quantil-quantil dos residuos do modelo ODP aplicado aos dados das indemnizações relativas ao período de 2006 a 2015 considerando a inflação.

para o primeiro ano após o último conhecido. Uma vez que se está a usar os dados de 2006 a 2015, as estimativas da primeira iteração são os valores da diagonal correspondente ao ano de pagamento em 2016, que se encontram na tabela 3.12. Nesta tabela também está o erro padrão da estimativa.

Comparando a tabela 3.12 com a tabela 3.4, verificam-se que as estimativas são semelhantes com exceção do ano de desenvolvimento zero, no entanto o erro padrão destas são muito diferentes. No ODP os valores do erro são menores que os erros do modelo Loglinear nos primeiros anos de desenvolvimento. Quando se analisam os últimos anos de desenvolvimento o erro padrão é mais reduzido quando se usa o modelo Loglinear.

Tabela 3.12: Resultados da primeira iteração da metodologia para o ano de calendário de 2016 com a aplicação do modelo ODP aos dados de 2006 a 2015 considerando a inflação (u.m.: EUR).

Ano de origem	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007
Ano de desenvolvimento	0	1	2	3	4	5	6	7	8	9
Estimativa	734,163	249,112	47,062	31,228	20,919	20,073	17,676	12,854	9,138	6,253
Erro padrão	38,965	21,854	8,276	6,714	5,518	5,514	5,313	4,638	4,131	3,932

Incorporando os resultados da primeira iteração da aplicação do modelo ODP aos dados históricos de 2006 a 2015, no triângulo de desenvolvimento, obtém-se um novo triângulo, com uma nova diagonal adicionada, tal como mostra a tabela 3.13.

O processo de aplicação do modelo ODP repete-se até atingir o ano de pagamento de 2019, sendo que na última iteração para além do cálculos dos valores relativos à diagonal do ano de calendário de 2019, calculam-se também os valores até ao ano de desenvolvimento 9, de modo a completar a matriz de desenvolvimento. Em cada iteração existe a verificação dos pressupostos do modelo, ou seja, se existe independência dos resíduos e se seguem uma distribuição normal. O detalhe da verificação dos pressupostos encontra-se no anexo A.2. Como resultado da aplicação

Tabela 3.13: Triângulo (ou matriz) de desenvolvimento com a inclusão dos resultados da primeira iteração da aplicação do modelo ODP aos dados relativos ao período 2006-2015, considerando a inflação (u.m.: EUR).

Ano de ocorrência	Ano de desenvolvimento	0	1	2	3	4	5	6	7	8	9
	2006	673,606	313,599	77,299	47,190	28,238	19,456	18,084	14,252	9,329	6,371
	2007	673,886	302,661	67,400	40,150	25,645	22,134	17,467	13,286	9,246	6,253 3,932
	2008	702,507	288,843	56,192	36,645	19,693	18,409	17,889	11,184	9,138 4,131	
	2009	698,641	293,695	57,550	35,856	25,308	20,106	12,905	12,854 4,638		
	2010	752,689	318,933	52,997	31,792	25,116	22,051	17,676 5,313			
	2011	738,662	265,436	43,262	35,605	19,828	20,073 5,514				
	2012	660,573	222,729	40,560	31,107	20,919 5,518					
	2013	616,410	219,041	41,896	31,228 6,714						
	2014	610,236	218,253	47,062 8,276							
	2015	621,016	249,112 21,854								
	2016	734,163 38,965									

Zona sombreada com os valores estimados pelo modelo ODP. A azul estão os valores do erro padrão das estimativas.

do modelo ODP aos dados históricos de 2006 a 2015, considerando a inflação, obteve-se a matriz de desenvolvimento da tabela 3.14.

As tabelas 3.15, 3.16 e 3.17 apresentam os resultados da aplicação do modelo ODP para o triângulo de desenvolvimento considerando o período 2006-2015 sem incorporação da inflação e o período 2001-2010, com incorporação e sem incorporação da inflação, respetivamente.

Ao efetuar as comparações entre as tabelas com os mesmos dados históricos, isto é, comparando a tabela 3.14 com a tabela 3.15 e comparando a tabela 3.16 com a tabela 3.17, é evidente o valor superior nas estimativas das provisões para sinistros nas tabelas que consideram a inflação histórica (até 2022) e futura (de 2023 a 2028).

Quando se efetua a comparação entre os resultados da tabela 3.14 com a tabela 3.16 e os resultados da tabela 3.15 com a tabela 3.17, ou seja, efetuando a comparação entre os resultados baseados em triângulos de desenvolvimento diferentes, mas com o mesmo pressuposto face à inflação, existe uma predominância de estimativas superiores quando se usa o triângulo de desenvolvimento com os dados relativos ao período de 2001 a 2010, sendo exceção para os anos de desenvolvimento 0 e 1. Quanto ao erro padrão, existe uma tendência decrescente com

Tabela 3.14: Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do modelo ODP aos dados relativos ao período 2006-2015, considerando a inflação (u.m.: EUR).

Ano de desenvolvimento Ano de ocorrênde cia	0	1	2	3	4	5	6	7	8	9
2006	673,606	313,599	77,299	47,190	28,238	19,456	18,084	14,252	9,329	6,371
2007	673,886	302,661	67,400	40,150	25,645	22,134	17,467	13,286	9,246	6,253 3,932
2008	702,507	288,843	56,192	36,645	19,693	18,409	17,889	11,184	9,138 4,131	6,279 3,086
2009	698,641	293,695	57,550	35,856	25,308	20,106	12,905	12,854 4,638	9,320 3,269	6,373 2,518
2010	752,689	318,933	52,997	31,792	25,116	22,051	17,676 5,313	13,461 3,740	9,972 2,760	6,823 2,476
2011	738,662	265,436	43,262	35,605	19,828	20,073 5,514	16,360 3,957	12,308 2,852	9,317 2,452	6,363 2,349
2012	660,573	222,729	40,560	31,107	20,919 5,518	18,131 4,054	14,212 2,942	11,145 2,504	8,247 2,264	5,706 2,146
2013	616,410	219,041	41,896	31,228 6,714	19,503 4,151	16,910 3,152	13,020 2,630	10,590 2,428	7,938 2,190	5,844 2,067
2014	610,236	218,253	47,062 8,276	30,246 5,161	19,013 3,311	17,026 2,967	12,976 2,625	10,691 2,423	8,528 2,186	6,121 2,062
2015	621,016	249,112 21,854	47,345 6,493	31,013 4,232	20,244 3,214	17,833 3,048	13,768 2,698	12,071 2,493	9,388 2,252	6,598 2,133
2016	734,163 38,965	291,509 18,567	53,601 5,625	36,027 4,317	23,802 3,527	21,240 3,350	17,451 2,972	14,917 2,755	11,358 2,506	7,920 2,406
2017	729,205 30,355	284,225 14,749	52,003 5,178	35,560 4,289	23,799 3,502	22,600 3,326	18,103 2,950	15,150 2,734	11,447 2,484	7,974 2,383
2018	724,671 24,442	278,740 13,649	51,447 5,194	35,636 4,289	25,380 3,495	23,498 3,317	18,429 2,939	15,303 2,721	11,550 2,470	8,046 2,365
2019	724,442 22,879	277,819 12,637	51,943 5,095	38,289 4,228	26,586 3,458	24,100 3,284	18,753 2,914	15,557 2,701	11,743 2,454	8,180 2,351

Zona sombreada com os valores estimados pelo modelo ODP. A azul estão os valores do erro padrão das estimativas.

o número de iterações feitas, visto que à medida que se avança no cálculo de mais iterações incorporam-se mais valores estimados pelo modelo. Relativamente ao ano de desenvolvimento zero, a estimativa é superior com os dados de 2005 a 2016. Na generalidade, o erro padrão é superior quando se utiliza o triângulo de desenvolvimento com os dados mais recentes.

Efetuando a comparação entre os resultados dos modelos Loglinear e ODP, é possível verificar que o erro padrão das estimativas obtidas pelo modelo ODP para os anos de desenvolvimento mais avançados é maior que o erro padrão das estimativas obtidas pelo modelo Loglinear.

Analisando os dados do quadro resumo 3.18, onde apresenta os valores reais ocorridos no período de 2016 a 2019 e as várias estimativas por ano de desenvolvimento e os respetivos intervalos de confiança a 95%, verifica-se que as estimativas produzidas (na maioria dos casos) pelo modelo ODP usando os dados mais antigos não são as mais adequadas quando se analisa

cada ano de desenvolvimento por ano de calendário, uma vez que os valores reais não estão incluídos nos intervalos de confiança para a maioria dos anos de desenvolvimento. Os intervalos de confiança que não contêm o valor real encontram-se marcados a *bordeaux*.

Tabela 3.15: Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do modelo ODP aos dados relativos ao período 2006-2015, não incorporando a inflação (u.m.: EUR).

Ano de desenvolvimento Ano de ocorrênde cia	0	1	2	3	4	5	6	7	8	9
2006	576,491	276,707	69,911	43,789	25,993	18,160	17,504	14,181	9,310	6,339
2007	594,609	273,732	62,543	36,958	23,937	21,424	17,380	13,260	9,200	6,345 4,199
2008	635,360	268,027	51,725	34,204	19,061	18,318	17,854	11,128	9,365 4,199	6,417 3,067
2009	648,292	270,349	53,717	34,706	25,182	20,066	12,841	13,236 4,722	9,536 3,247	6,555 2,402
2010	692,857	297,691	51,297	31,634	25,066	21,941	17,908 5,357	13,639 3,691	10,203 2,619	7,003 2,289
2011	689,464	256,924	43,047	35,534	19,729	20,370 5,566	16,682 3,921	12,657 2,718	9,685 2,288	6,653 2,202
2012	639,390	221,623	40,480	30,952	21,491 5,608	18,822 4,057	14,892 2,835	11,885 2,372	8,846 2,156	6,076 2,058
2013	613,349	218,607	41,688	32,240 6,841	20,318 4,162	17,940 3,057	13,974 2,498	11,529 2,328	8,581 2,113	5,894 2,012
2014	609,027	217,167	47,729 8,352	30,976 5,126	19,956 3,190	18,136 2,804	13,957 2,497	11,515 2,327	8,571 2,112	5,887 2,011
2015	617,926	249,479 21,930	47,607 6,386	31,883 4,032	21,329 3,018	18,987 2,880	14,612 2,566	12,055 2,393	8,973 2,176	6,163 2,080
2016	631,324 35,950	250,054 16,752	46,537 4,881	32,202 3,698	21,627 3,044	19,252 2,905	14,816 2,588	12,223 2,414	9,098 2,197	6,249 2,102
2017	638,764 27,696	248,775 12,890	46,361 4,441	32,410 3,714	21,766 3,057	19,376 2,917	14,912 2,599	12,302 2,425	9,157 2,206	6,290 2,112
2018	651,649 21,661	251,473 11,808	47,175 4,527	32,979 3,776	22,148 3,101	19,716 2,958	15,173 2,634	12,518 2,456	9,317 2,235	6,400 2,141
2019	667,463 19,996	257,576 11,090	48,320 4,507	33,779 3,778	22,686 3,115	20,195 2,975	15,542 2,654	12,822 2,478	9,543 2,260	6,555 2,171

Zona sombreada com os valores estimados pelo modelo ODP. A azul estão os valores do erro padrão das estimativas.

Não é possível avaliar qual o melhor modelo quando se analisa cada ano de desenvolvimento e ano de calendário, dado que para alguns anos a estimativa considerando os dados mais recentes sem incluir a inflação é mais próxima do valor real e para outros anos a melhor estimativa é a produzida pelos dados mais atuais e considerando a inflação. Para o ano de desenvolvimento zero, as melhores estimativas para os anos de calendário de 2016, 2017 e 2019 são as produzidas com os dados mais recentes e não considerando a inflação.

Quando se analisa a coluna relativa à estimativa total do ano de calendário, os modelos considerando a inflação não contemplam o valor real nos intervalos de confiança e excetuando o

Tabela 3.16: Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do modelo ODP aos dados relativos ao período 2001-2010, considerando a inflação (u.m.: EUR).

Ano de desenvolvimento Ano de ocorrênde cia	0	1	2	3	4	5	6	7	8	9
2001	819,739	367,543	95,822	70,033	38,386	30,288	28,342	18,695	13,881	12,664
2002	845,246	333,104	92,194	59,764	38,932	31,086	24,539	16,067	14,442	12,890 3,333
2003	745,065	334,219	85,226	58,242	36,224	25,775	23,570	18,007	13,388 2,913	12,198 3,028
2004	707,760	294,165	77,963	51,987	30,148	25,365	19,207	15,783 2,980	12,638 2,672	11,278 2,636
2005	651,769	290,478	78,117	49,718	30,909	24,229	21,083 3,350	14,912 2,750	11,893 2,357	10,681 2,501
2006	614,478	288,079	71,043	43,023	25,677	23,525 3,476	19,380 3,047	14,479 2,458	11,284 2,226	10,166 2,289
2007	615,439	278,335	61,966	37,045	28,940 3,823	23,009 3,270	18,705 2,721	13,931 2,345	11,099 2,084	10,020 2,152
2008	638,422	264,980	50,720	47,227 4,895	29,040 3,647	22,553 2,945	18,601 2,641	14,159 2,234	11,229 1,983	10,239 1,991
2009	635,461	270,012	71,165 6,059	46,960 4,646	28,865 3,306	23,245 2,908	19,604 2,560	14,566 2,139	11,617 1,843	10,573 1,803
2010	685,449	308,296 14,540	76,993 6,018	50,375 4,390	31,111 3,350	25,720 2,901	21,326 2,533	16,058 2,065	12,773 1,737	11,613 717
2011	870,121 24,751	387,813 15,693	94,879 6,145	62,600 4,821	39,881 3,634	32,615 3,135	27,127 2,661	20,358 2,122	16,154 622	14,671 690
2012	847,670 23,184	374,678 13,967	91,052 5,803	60,892 4,441	38,612 3,332	31,889 2,788	26,375 2,294	19,747 648	15,685 618	14,430 684
2013	756,998 19,782	332,025 12,627	81,311 5,102	53,950 3,881	35,199 2,853	28,351 2,294	23,519 732	17,601 651	14,162 620	13,865 685
2014	694,782 18,317	308,973 11,394	74,066 4,544	50,349 3,385	32,259 2,404	26,067 787	21,667 729	16,426 650	14,065 619	13,425 685
2015	664,088 16,796	295,560 10,435	70,856 4,027	48,959 2,953	30,597 894	24,933 814	20,994 755	16,937 674	14,139 644	13,213 717
2016	630,359 15,351	281,337 9,223	67,502 3,479	45,970 1,093	29,054 869	23,984 791	21,490 732	16,903 653	13,816 622	12,811 690
2017	623,391 13,874	276,780 8,112	67,937 1,338	45,440 1,088	29,093 864	25,557 787	22,326 728	17,193 648	13,943 618	12,916 684
2018	624,973 12,325	277,360 3,117	68,038 1,360	46,099 1,100	31,408 871	26,900 791	23,007 732	17,580 651	14,243 620	13,194 685
2019	632,055 4,749	279,664 2,865	69,495 1,329	50,107 1,085	33,284 864	27,910 787	23,686 729	18,080 650	14,649 619	13,570 685

Zona sombreada com os valores estimados pelo modelo ODP. A azul estão os valores do erro padrão das estimativas.

ano de 2017, o modelo que mais se aproxima dos valores reais é o modelo com os dados mais recentes, não incorporando a inflação.

Também é possível verificar a existência de dois intervalos de confiança com valores negativos no limite inferior (para o ano de calendário de 2016, ano de desenvolvimento 9 e usando os dados mais recentes). No contexto da provisão para sinistros o valor negativo não faz sentido, sendo

isto provocado pelo erro padrão, que neste caso é superior a metade da estimativa produzida.

Tabela 3.17: Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do modelo ODP aos dados relativos ao período 2001-2010, não incorporando a inflação (u.m.: EUR).

Ano de desenvolvimento Ano de ocorrênde cia	0	1	2	3	4	5	6	7	8	9
2001	642,807	300,894	81,270	61,298	34,405	27,771	26,792	18,115	13,800	12,489
2002	691,972	282,517	80,695	53,565	35,697	29,386	23,777	15,973	14,243	12,703 3,650
2003	631,916	292,534	76,387	53,402	34,243	24,975	23,432	17,758	13,453 3,219	12,085 3,165
2004	619,485	263,655	71,484	49,144	29,212	25,217	18,942	15,912 3,295	12,783 2,819	11,441 2,954
2005	584,169	266,338	73,846	48,174	30,728	23,894	21,243 3,702	15,255 2,917	12,415 2,719	11,133 2,685
2006	563,413	272,327	68,837	42,771	25,322	23,685 3,839	19,914 3,238	15,240 2,892	12,052 2,420	10,797 2,531
2007	581,787	269,694	61,603	36,534	29,168 4,225	23,743 3,483	19,962 3,203	15,103 2,569	12,117 2,337	10,860 2,345
2008	618,601	263,429	50,020	47,429 5,401	29,917 3,881	24,163 3,405	20,260 2,900	15,565 2,513	12,491 2,188	11,193 2,079
2009	631,741	266,284	70,899 6,656	47,980 4,923	30,745 3,724	25,305 3,191	21,509 2,875	16,247 2,361	12,972 1,942	11,625 1,824
2010	675,985	300,917 15,840	76,196 6,268	51,938 4,619	32,963 3,620	27,427 3,205	23,177 2,753	17,587 2,149	14,088 1,754	12,625 704
2011	665,908 23,523	294,677 14,141	74,289 5,565	50,411 4,454	32,507 3,434	27,002 2,900	22,740 2,353	17,284 1,813	13,838 630	12,401 689
2012	680,790 21,527	304,157 11,816	75,498 5,481	51,240 4,305	33,006 3,169	27,612 2,545	23,274 2,042	17,712 672	14,171 634	12,699 693
2013	646,464 19,417	287,647 12,268	71,113 5,075	48,058 3,797	31,745 2,682	26,146 2,111	22,123 769	16,789 688	13,433 649	12,038 711
2014	613,945 17,941	274,480 11,456	66,531 4,476	45,720 3,202	30,073 2,232	24,868 830	21,017 777	15,950 697	12,761 660	11,436 725
2015	596,758 16,932	266,253 10,310	63,897 3,799	45,327 2,734	29,144 884	24,159 813	20,418 759	15,495 681	12,397 643	11,110 704
2016	580,212 15,281	256,743 8,764	62,130 3,212	43,729 1,081	28,263 869	23,428 799	19,801 746	15,026 668	12,022 630	10,774 689
2017	585,347 13,329	257,759 7,540	64,362 1,321	44,137 1,088	28,526 875	23,647 804	19,985 751	15,166 672	12,134 634	10,874 693
2018	603,548 11,631	266,727 3,100	66,436 1,365	45,559 1,120	29,445 897	24,409 823	20,629 769	15,655 688	12,525 649	11,225 711
2019	625,104 4,774	276,254 2,884	68,809 1,353	47,186 1,119	30,497 903	25,280 830	21,366 777	16,214 697	12,973 660	11,625 725

Zona sombreada com os valores estimados pelo modelo ODP. A azul estão os valores do erro padrão das estimativas.

Efetuando uma análise aos gráficos da figura 3.20 que apresentam as estimativas de forma acumulada, as estimativas produzidas pelos modelos sem inflação estão mais próximas da realidade ocorrida, sendo que ao usar os dados menos recentes, as estimativas encontram-se

abaixo dos valores reais nos primeiros anos de desenvolvimento. No entanto quando se compara o crescimento ao longo dos anos de desenvolvimento, os dados mais recentes incorporando a inflação têm um comportamento mais ajustado ao crescimento ocorrido com os dados reais (isto é, a tendência de crescimento é semelhante). Contudo, as provisões determinadas com os dados mais recentes e tendo em conta a incorporação da inflação encontram-se sempre acima dos valores reais ocorridos.

Comparando os resultados produzidos pelos dois modelos para cada ano de calendário ao longo dos anos de desenvolvimento (tabelas 3.10 e 3.18), o modelo ODP apresenta estimativas mais próximas dos valores reais com os dados mais recentes, no entanto, como foi referido anteriormente, não é possível afirmar que a incorporação da inflação melhora o modelo, uma vez que para uns anos a estimativa que está mais próxima do valor real incorpora a inflação e para outros anos tal não acontece. Quando se analisa a totalidade da provisão para sinistros por ano de calendário, verifica-se que: para os anos de 2016, 2018 e 2019 o modelo ODP com os dados mais recentes e sem a incorporação da inflação produz valores mais próximos da realidade; e para o ano de 2017 o melhor modelo é o modelo ODP com os dados mais antigos e não tendo em conta a inflação. De uma forma geral, para a estimativa total por ano de calendário, a inclusão da inflação não produz uma aproximação tão boa à realidade como os modelos sem incorporação da inflação, tanto no modelo Loglinear, como no ODP.

Comparando os resultados acumulados obtidos com a aplicação do modelo ODP e do modelo Loglinear, através da análise dos gráficos da figura 3.21, conclui-se que o modelo Loglinear com os dados de 2006 a 2015 produz, na maioria dos casos, estimativas superiores do que os modelos ODP (com qualquer dos dois triângulos de desenvolvimento de base). Já o modelo Loglinear com os dados de 2001 a 2010 produz, na maioria dos casos, estimativas inferiores aos restantes modelos. Novamente constata-se que ao considerar a inflação, as estimativas revelam-se superiores aos valores reais, principalmente a partir do ano 1 de desenvolvimento. O modelo que aparentemente mais se ajusta ao comportamento da curva dos pagamentos acumulados é o modelo ODP com os dados mais recentes e não incorporando a inflação. Isto não coincide com os modelos anteriormente referidos, visto que a análise das tabelas 3.10 e 3.18 é relativa aos valores incrementais e ao valor acumulado apenas no ano de desenvolvimento 9 (valor total das provisões para o ano de calendário).

Tabela 3.18: Comparação dos resultados obtidos através do modelo ODP e os valores reais das indemnizações relativas ao período de 2016 a 2019 por ano de calendário.

									Ano de desenv	olvimento					u.m.: EUR
					0	1	2	3	4	5	6	7	8	9	Total
		Valor	r real		645,259	230,506	46,957	33,058	21,340	19,834	18,389	14,800	8,418	6,965	1,045,526
			.9	I		266,253	66,531	48,058	33,006	27,002	23,177	16,247	12,491	10,860	1,083,838
			inflação		[550,262; 610,162]	[246,046; 286,460]	[57,758; 75,304]	[40,616; 55,500]	[26,795; 39,217]	[21,318; 32,686]	[17,781; 28,573]	[11,620; 20,874]	[8,203; 16,779]	[6,264; 15,456]	[1,045,288; 1,122,388]
			E.	II	631,324	249,479	47,729	32,240	21,491	20,370	17,908	13,236	9,365	6,345	1,049,487
	91	C	Sem		[560,863; 701,785]	[206,497; 292,461]	[31,359; 64,099]	[18,832; 45,648]	[10,500; 32,482]	[9,461; 31,279]	[7,408; 28,408]	[3,981; 22,491]	[1,135; 17,595]	[-1,885; 14,575]	[963,861; 1,135,113]
	2016	iva/													
		Estimativa/	go	I	630,359	295,560	74,066	53,950	38,612	32,615	21,326	14,566	11,229	10,020	1,182,304
		Est	inflação		[600,273; 660,445]	[275,107; 316,013]	[65,161; 82,971]	[46,344; 61,556]	[32,082; 45,142]	[26,470; 38,760]	[16,361; 26,291]	[10,374; 18,758]	[7,342; 15,116]	[5,803; 14,237]	[1,143,506; 1,221,101]
			Com in	II	734,163	249,112	47,062	31,228	20,919	20,073	17,676	12,854	9,138	6,253	1,148,478
			ರೆ		[657,793; 810,533]	[206,279; 291,945]	[30,841; 63,283]	[18,069; 44,387]	[10,104; 31,734]	[9,266; 30,880]	[7,263; 28,089]	[3,764; 21,944]	[1,041; 17,235]	[-1,454; 13,960]	[1,058,325; 1,238,632]
		Valor	r real		658,862	243,478	48,142	31,584	26,567	20,853	20,386	12,467	9,104	7,209	1,078,653
			ão	I	585,347	256,743	63,897	45,720	31,745	27,612	22,740	17,587	12,972	11,193	1,075,555
			inflação		[559,223; 611,471]	[239,566; 273,920]	[56,451; 71,343]	[39,444; 51,996]	[26,488; 37,002]	[22,624; 32,600]	[18,128; 27,352]	[13,375; 21,799]	[9,166; 16,778]	[7,118; 15,268]	[1,042,210; 1,108,899]
			Sem in	II	638,764	250,054	47,607	30,976	20,318	18,822	16,682	13,639	9,536	6,417	1,052,815
	17	IC	Š		[584,481; 693,047]	[217,221; 282,887]	[35,091; 60,123]	[20,929; 41,023]	[12,161; 28,475]	[10,870; 26,774]	[8,997; 24,367]	[6,405; 20,873]	[3,172; 15,900]	[406; 12,428]	[987,143; 1,118,487]
	20	Estimativa/	ão	I	623,391	281,337	70,856	50,349	35,199	31,889	27,127	16,058	11,617	10,239	1,158,063
		tima	nflaç		[596,198; 650,584]	[263,260; 299,414]	[62,963; 78,749]	[43,715; 56,983]	[29,607; 40,791]	[26,425; 37,353]	[21,912; 32,342]	[12,011; 20,105]	[8,005; 15,229]	[6,337; 14,141]	[1,123,181; 1,192,945]
		Est	Com inflação	II	729,205	291,509	47,345	30,246	19,503	18,131	16,360	13,461	9,320	6,279	1,181,360
			٥		[669,710; 788,700]	[255,118; 327,900]	[34,619; 60,071]	[20,131; 40,361]	[11,367; 27,639]	[10,185; 26,077]	[8,604; 24,116]	[6,131; 20,791]	[2,913; 15,727]	[231; 12,327]	[1,109,963; 1,252,757]
		Valor	r real		622,202	254,191	45,088	31,407	19,034	16,391	14,973	12,700	8,107	4,272	1,028,365
iTi			ão	I	603,548	257,759	62,130	45,327	30,073	26,146	23,274	17,284	14,088	11,625	1,091,254
calendário			inflação		[580,752; 626,344]	[242,981; 272,537]	[55,835; 68,425]	[39,968; 50,686]	[25,698; 34,448]	[22,009; 30,283]	[19,272; 27,276]	[13,731; 20,837]	[10,650; 17,526]	[8,050; 15,200]	[1,062,388; 1,120,121]
e cal			Sem ii	II	651,649	248,775	46,537	31,883	19,956	17,940	14,892	12,657	10,203	6,555	1,061,048
Ano de	2018	JC IC	Ň		[609,194; 694,104]	[223,511; 274,039]	[36,970; 56,104]	[23,980; 39,786]	[13,704; 26,208]	[11,948; 23,932]	[9,336; 20,448]	[7,330; 17,984]	[5,070; 15,336]	[1,847; 11,263]	[1,010,009; 1,112,086]
₹	20	Estimativa/	ção	I	624,973	276,780	67,502	48,959	32,259	28,351	26,375	20,358	12,773	10,573	1,148,902
		tima	Com inflação		[600,816; 649,130]	[260,881; 292,679]	[60,683; 74,321]	[43,171; 54,747]	[27,547; 36,971]	[23,855; 32,847]	[21,879; 30,871]	[16,199; 24,517]	[9,369; 16,177]	[7,039; 14,107]	[1,118,047; 1,179,757]
		Es	om i	II	724,671	284,225	53,601	31,013	19,013	16,910	14,212	12,308	9,972	6,373	1,172,299
			O		[676,766; 772,576]	[255,317; 313,133]	[42,576; 64,626]	[22,718; 39,308]	[12,524; 25,502]	[10,732; 23,088]	[8,445; 19,979]	[6,718; 17,898]	[4,563; 15,381]	[1,438; 11,308]	[1,114,889; 1,229,709]
		Valor	r real		654,068	264,025	51,182	33,215	20,746	17,832	12,426	13,811	12,612	7,465	1,087,382
			žão	I	625,104	266,727	64,362	43,729	29,144	24,868	22,123	17,712	13,838	12,625	1,120,232
			Sem inflação		[615,747; 634,461]	[260,651; 272,803]	[61,773; 66,951]	[41,610; 45,848]	[27,411; 30,877]	[23,241; 26,495]	[20,616; 23,630]	[16,395; 19,029]	[12,603; 15,073]	[11,245; 14,005]	[1,108,431; 1,132,034]
			em i	II	667,463	251,473	46,361	32,202	21,329	18,136	13,974	11,885	9,685	7,003	1,079,511
	2019) IC	S		[628,272; 706,654]	[228,330; 274,616]	[37,657; 55,065]	[24,954; 39,450]	[15,414; 27,244]	[12,640; 23,632]	[9,078; 18,870]	[7,236; 16,534]	[5,201; 14,169]	[2,517; 11,489]	[1,032,568; 1,126,455]
	2	Estimativa/	ção	I	632,055	277,360	67,937	45,970	30,597	26,067	23,519	19,747	16,154	11,613	1,151,018
		stim	infla		[622,747; 641,363]	[271,251; 283,469]	[65,315; 70,559]	[43,828; 48,112]	[28,845; 32,349]	[24,525; 27,609]	[22,084; 24,954]	[18,477; 21,017]	[14,935; 17,373]	[10,208; 13,018]	[1,139,164; 1,162,872]
		ш	Com inflação	II	724,442	278,740	52,003	36,027	20,244	17,026	13,020	11,145	9,317	6,823	1,168,788
			٥		[679,600; 769,284]	[251,988; 305,492]	[41,854; 62,152]	[27,566; 44,488]	[13,945; 26,543]	[11,211; 22,841]	[7,865; 18,175]	[6,237; 16,053]	[4,511; 14,123]	[1,970; 11,676]	[1,115,130; 1,222,446]

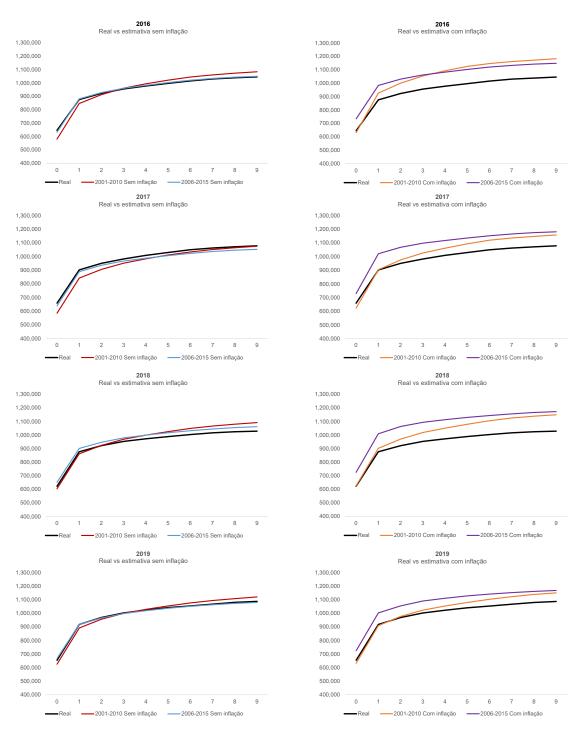


Figura 3.20: Gráficos comparativos entre os valores de pagamento reais e os valores estimados pelo modelo ODP no período de 2016 a 2019 por ano de desenvolvimento do sinistro (u.m.: EUR).

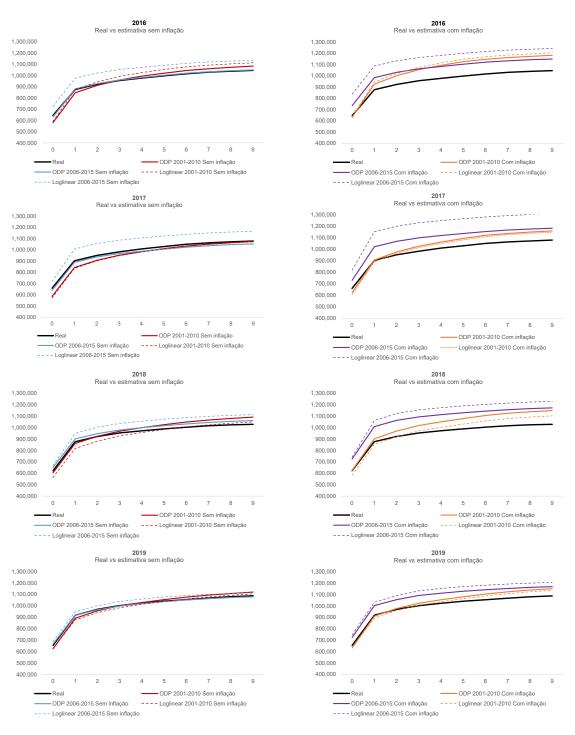


Figura 3.21: Gráficos comparativos entre os valores de pagamento reais e os valores estimados pelos modelos Loglinear e ODP no período de 2016 a 2019 por ano de desenvolvimento do sinistro (u.m.:EUR).

3.3 Bootstrap

A aplicação do modelo *Bootstrap* foi apenas aplicada para completar os triângulos de desenvolvimento a partir da construção das diagonais obtidas pelo método Loglinear até ao ano de calendário de 2019. Neste sentido, e de modo a obter comparações na mesma base, procederse-á à comparação das metodologias relativas aos dados do triângulo inferior da matriz de desenvolvimento a partir do ano de calendário de 2019.

Antes da aplicação da metodologia de *Bootstrap* é necessário verificar os dados, de modo a perceber se existem *outliers* ou valores que não façam sentido, pois esta metodologia ao efetuar simulações a partir dos dados, irá repetir os mesmos comportamentos. Da análise gráfica das figuras 3.1, 3.2, 3.3 e 3.4 verifica-se que os dados são bem comportados para a aplicação da metodologia.

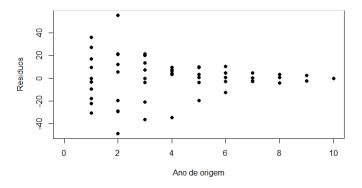


Figura 3.22: Análise gráfica dos resíduos de um triângulo obtido por simulação de *Bootstrap* em função do ano de origem do sinistro.

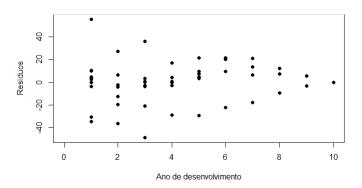


Figura 3.23: Análise gráfica dos resíduos de um triângulo obtido por simulação de *Bootstrap* em função do ano de desenvolvimento do sinistro.

Para analisar a adequabilidade do modelo, procedeu-se a análise gráfica dos resíduos. Para exemplificar, escolheu-se uma simulação feita pelo modelo de Bootstrap que se obteve dos dados

do período de 2005 a 2016, considerando a inflação e após já ter sido aplicado o modelo Loglinear até perfazer a diagonal do triângulo de desenvolvimento relativa ao ano de pagamento de 2019. Estes resíduos estão presentes nas figuras 3.22 e 3.23. Nestas duas figuras é possível verificar que não existe um padrão que sugira dependência entre os resíduos e os anos de origem ou desenvolvimento. Sendo assim, procedeu-se com a aplicação da metodologia efetuando 50,000 simulações para cada um dos quatro conjuntos de dados. Note-se que ao usar o número elevado de simulações o *software* utilizado demorou mais tempo a fazer o cálculo das estimativas do que nos modelos anteriores, não sendo porém uma espera muito demorada que inviabilize a possível escolha desta metodologia.

Tabela 3.19: Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do modelo Loglinear e *Bootstrap* aos dados relativos ao período 2006-2015, considerando a inflação (u.m.: EUR).

Ano de desenvolvimento Ano de ocorrência	0	1	2	3	4	5	6	7	8	9
2006	673,606	313,599	77,299	47,190	28,238	19,456	18,084	14,252	9,328	6,371
2007	673,886	302,661	67,400	40,150	25,645	22,134	17,467	13,286	9,246	6,196 981
2008	702,507	288,843	56,192	36,645	19,693	18,409	17,889	11,184	8,242 1,135	5,676 735
2009	698,641	293,695	57,550	35,856	25,308	20,106	12,905	12,036 1,574	8,539 963	5,858 679
2010	752,689	318,933	52,997	31,792	25,116	22,051	16,294 2,085	12,476 1,336	8,906 899	6,136 588
2011	738,662	265,436	43,262	35,605	19,828	17,910 2,277	14,495 1,519	10,970 1,053	7,914 661	5,717 1,919
2012	660,573	222,729	40,560	31,107	19,227 2,460	16,814 1,750	13,066 1,226	10,064 798	7,197 1,864	5,120 1,782
2013	616,410	219,041	41,896	29,736 3,889	18,769 1,966	16,326 1,523	12,139 942	9,563 2,029	6,935 1,829	5,257 1,853
2014	610,236	218,253	45,396 6,251	30,090 3,223	18,996 1,783	16,724 1,289	12,380 2,261	9,656 2,045	7,461 1,968	5,508 1,933
2015	621,016	250,048 39,572	47,433 5,347	31,908 3,062	20,668 1,604	17,127 2,635	13,178 2,355	10,948 2,291	8,236 2,135	5,958 2,065
2016	836,779 98,380	332,818 43,102	61,517 6,213	41,966 3,329	26,300 3,288	23,224 3,165	19,039 3,046	15,386 2,883	11,350 2,659	8,136 2,630
2017	818,675 78,838	319,461 37,055	58,608 4,893	39,835 3,992	25,870 3,280	24,324 3,352	19,432 3,139	15,366 2,910	11,250 2,675	8,057 2,611
2018	741,786 64,002	285,184 27,337	51,687 4,533	36,361 3,830	25,132 3,317	23,023 3,327	18,032 3,029	14,159 2,800	10,331 2,532	7,412 2,467
2019	746,185 53,230	284,554 12,154	52,512 4,633	39,318 4,185	26,467 3,530	23,763 3,444	18,443 3,099	14,466 2,855	10,576 2,601	7,574 2,526

Zona sombreada a azul com os valores estimados pelo modelo Loglinear. A azul estão os valores do erro padrão das estimativas obtidas pelo modelo Loglinear (tabela 3.6). Zona sombreada a verde com os valores estimados pela metodologia *Bootstrap*. A verde estão os valores do erro padrão das estimativas.

As tabelas 3.19, 3.20, 3.21 e 3.22 apresentam as estimativas resultantes da combinação dos dois métodos, sendo que a região a azul é relativa ao modelo Loglinear e a região a verde é relativa à metodologia *Bootstrap*, aplicadas aos dados de 2006 a 2015 considerando ou desconsiderando a inflação e aos dados de 2001 a 2010 incorporando ou não incorporando a inflação, respetivamente.

Tabela 3.20: Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do modelo Loglinear e *Bootstrap* aos dados relativos ao período 2006-2015, não incorporando a inflação (u.m.: EUR).

Ano de desenvolvimento Ano de ocorrência	0	1	2	3	4	5	6	7	8	9
2006	576,491	276,707	69,911	43,789	25,993	18,160	17,504	14,181	9,310	6,339
2007	594,609	273,732	62,543	36,958	23,937	21,424	17,380	13,260	9,200	6,236 1,011
2008	635,360	268,027	51,725	34,204	19,061	18,318	17,854	11,128	8,372 1,181	5,722 740
2009	648,292	270,349	53,717	34,706	25,182	20,066	12,841	12,316 1,650	8,662 975	5,924 667
2010	692,857	297,691	51,297	31,634	25,066	21,941	16,663 2,183	12,732 1,361	9,124 894	6,257 574
2011	689,464	256,924	43,047	35,534	19,729	18,295 2,381	14,906 1,559	11,326 1,054	8,213 656	5,936 1,816
2012	639,390	221,623	40,480	30,952	19,662 2,576	17,364 1,804	13,608 1,239	10,615 806	7,701 1,786	5,397 1,683
2013	613,349	218,607	41,688	30,428 4,075	19,308 2,019	17,061 1,543	12,823 952	10,331 1,955	7,461 1,751	5,234 1,647
2014	609,027	217,167	46,110 6,502	30,633 3,275	19,686 1,792	17,570 1,297	13,314 2,170	10,345 1,958	7,473 1,748	5,231 1,647
2015	617,926	251,646 40,785	47,852 5,385	32,610 3,035	21,525 1,599	18,137 2,499	14,011 2,229	10,888 2,008	7,856 1,808	5,512 1,720
2016	725,304 87,318	288,677 37,317	53,770 5,268	37,428 2,842	23,860 2,873	21,080 2,726	16,290 2,442	12,647 2,213	9,127 2,005	6,412 1,914
2017	719,266 69,138	281,120 31,629	52,333 4,182	36,219 3,492	23,499 2,842	20,761 2,704	16,054 2,422	12,461 2,190	9,003 1,979	6,317 1,890
2018	668,571 55,958	258,556 23,721	47,557 3,968	33,502 3,359	21,789 2,721	19,226 2,583	14,857 2,312	11,542 2,101	8,328 1,878	5,848 1,789
2019	689,831 47,103	264,610 10,700	48,943 4,091	34,513 3,409	22,402 2,760	19,816 2,631	15,286 2,349	11,873 2,118	8,574 1,907	6,018 1,831

Zona sombreada a azul com os valores estimados pelo modelo Loglinear. A azul estão os valores do erro padrão das estimativas obtidas pelo modelo Loglinear (tabela 3.7). Zona sombreada a verde com os valores estimados pela metodologia *Bootstrap*. A verde estão os valores do erro padrão das estimativas.

Tal como acontecia nos outros modelos, quando se usam os dados incorporando a inflação histórica (até 2022) e futura (de 2023 a 2028), as estimativas produzidas são superiores. Os resultados produzidos a partir do triângulo de desenvolvimento relativo às indemnizações de 2001 a 2010 são superiores aos produzidos pelo mesmo método a partir do triângulo de desenvolvimento com os dados mais recentes.

Tabela 3.21: Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do modelo Loglinear e *Bootstrap* aos dados relativos ao período 2001-2010, considerando a inflação (u.m.: EUR).

Ano de desenvolvimento Ano de ocorrência	0	1	2	3	4	5	6	7	8	9
2001	819,739	367,543	95,822	70,033	38,386	30,288	28,342	18,695	13,881	12,664
2002	845,246	333,104	92,194	59,764	38,932	31,086	24,539	16,067	14,442	12,519 1,331
2003	745,065	334,219	85,226	58,242	36,224	25,775	23,570	18,007	13,556 1,255	12,267 1,217
2004	707,760	294,165	77,963	51,987	30,148	25,365	19,207	15,263 1,342	12,492 1,079	11,015 1,004
2005	651,769	290,478	78,117	49,718	30,909	24,229	21,242 1,827	15,186 1,247	12,241 971	10,920 932
2006	614,478	288,079	71,043	43,023	25,677	22,765 1,946	18,968 1,523	14,102 1,064	11,224 834	10,041 799
2007	615,439	278,335	61,966	37,045	26,915 2,315	21,614 1,725	17,698 1,306	13,123 927	10,636 737	9,561 709
2008	638,422	264,980	50,720	42,514 3,738	26,411 2,121	20,624 1,512	17,106 1,182	12,935 852	10,449 675	9,497 644
2009	635,461	270,012	70,265 6,503	46,547 3,821	28,560 2,107	23,075 1,585	19,507 1,257	14,491 889	11,792 697	10,722 925
2010	685,449	310,402 32,991	77,245 6,674	50,371 3,799	31,067 2,147	25,549 1,636	21,383 1,284	16,065 902	13,047 888	11,849 243
2011	917,994 72,594	412,259 40,898	99,995 7,938	65,530 4,629	41,426 2,670	33,919 2,024	28,439 1,562	21,318 1,034	17,267 309	15,819 609
2012	858,220 63,356	381,250 34,748	91,817 6,827	60,703 3,999	38,319 2,301	31,634 1,727	26,383 1,252	19,723 336	16,152 498	14,907 583
2013	783,456 53,146	345,407 29,484	83,486 5,788	54,953 3,373	35,643 1,958	28,766 1,356	24,064 400	18,110 491	14,908 478	14,644 581
2014	696,686 44,266	311,505 24,794	73,773 4,765	49,718 2,792	31,722 1,505	25,653 424	21,568 519	16,378 464	14,350 477	13,741 562
2015	693,658 41,101	310,443 23,019	73,455 4,341	50,423 2,445	31,424 523	25,685 558	21,777 524	17,596 503	15,031 500	14,088 576
2016	636,083 35,114	285,335 19,357	67,609 3,451	45,745 778	28,861 580	23,868 535	21,532 534	16,965 500	14,188 489	13,199 554
2017	604,939 30,557	269,892 15,810	65,645 1,175	43,446 710	27,797 576	24,463 566	21,518 547	16,596 498	13,770 483	12,796 544
2018	582,704 25,420	259,859 5,339	62,921 863	42,367 710	28,844 608	24,748 581	21,311 552	16,308 497	13,520 479	12,568 540
2019	634,512 9,711	284,396 2,135	69,447 929	49,760 808	33,031 671	27,748 628	23,713 594	18,123 537	15,028 519	13,965 588

Zona sombreada a azul com os valores estimados pelo modelo Loglinear. A azul estão os valores do erro padrão das estimativas obtidas pelo modelo Loglinear (tabela 3.8). Zona sombreada a verde com os valores estimados pela metodologia *Bootstrap*. A verde estão os valores do erro padrão das estimativas.

Ao comparar as tabelas que usam dados diferentes mas o mesmo critério quanto à inflação, ou seja, a tabela 3.19 com a tabela 3.21 e a tabela 3.20 com a tabela 3.22, verifica-se que o erro padrão é superior quando o triângulo de desenvolvimento que serviu de base para os cálculos foi

o que continha os dados mais recentes. Isto poderá estar relacionado com o facto de ao usar os dados mais antigos, ser necessário fazer mais iterações em que assumimos como dados reais as estimativas das iterações anteriores, e por isso reduzindo a variabilidade dos valores entre si.

Tabela 3.22: Triângulo (ou matriz) de desenvolvimento com os resultados da aplicação do modelo Loglinear e *Bootstrap* aos dados relativos ao período 2001-2010, não incorporando a inflação (u.m.: EUR).

Ano de desenvol vimento Ano le ocorrên ia	0	1	2	3	4	5	6	7	8	9
2001	642,807	300,894	812,70	612,98	34,405	27,771	26,792	18,115	13,800	12,48
2002	691,972	28,2517	80,695	53,565	35,697	29,386	23,777	15,973	14,243	12,22 1,428
2003	631,916	292,534	763,87	53,402	34,243	24,975	23,432	1,7758	13,458 1,369	11,91- 1,258
2004	619,485	263,655	71,484	49,144	29,212	25,217	18,942	15,224 1,471	12,420 1,142	10,86 1,050
2005	584,169	266,338	73,846	48,174	30,728	23,894	21,223 2,006	15,307 1,337	12,474 1,056	11,01 991
2006	563,413	272,327	68,837	42771	25322	22,738 2,136	19,215 1,642	14,558 1,171	11,647 913	10,26 862
2007	581,787	269,694	61,603	36,534	26,891 2,542	21,946 1,863	18,468 1,453	13,829 1,030	11,198 819	9,90 750
2008	618,601	263,429	50,020	42,407 4,098	26,747 2,285	21,553 1,685	18,096 1,318	13,735 954	11,123 734	9,83 670
2009	631,741	266,284	69,758 7,096	46,832 4,091	29,693 2,336	24,418 1,767	20,713 1,408	15,537 974	12,535 743	11,09 644
2010	675,985	303,161 35,415	76,128 6,999	51,204 4,119	32,246 2,349	26,601 1,797	22,544 1,383	16,949 954	13,719 694	12,13 232
2011	716,116 62,227	319,240 33,704	79,538 6,734	53,285 3,968	33,920 2,306	28,053 1,710	23,668 1,304	17,828 857	14,410 240	12,80 477
2012	704,623 55,350	316,933 30,809	77,927 6,107	51,995 3,613	33,103 2,031	27,522 1,507	23,222 1,092	17,501 277	14,224 432	12,61 471
2013	683,359 49,436	306,182 27,550	74,742 5,467	49,709 3,118	32,363 1,784	26,549 1,241	22,472 347	16,993 445	13,764 422	12,2 459
2014	626,417 41,952	282,130 23,689	67,500 4,456	45,530 2,565	29,516 1,388	24,278 373	20,667 469	15,580 422	12,619 398	11,19
2015	631,863 39,493	284,011 21,521	67,227 3,986	46,670 2,243	29,682 459	24,621 510	20,836 474	15,705 421	12,722 401	11,28
2016	591,927 33,392	263,776 17,953	62,836 3,179	43,301 684	27,844 530	22,997 487	19,466 455	14,670 408	11,885 384	10,54
2017	571,457 28,959	253,381 14,713	62,357 1,037	42,032 650	26,898 523	22,217 478	18,805 448	14,177 396	11,483 377	10,18
2018	564,101 24,392	250,867 4,787	61,491 800	41,525 649	26,575 520	21,953 476	18,578 446	14,004 396	11,344 375	10,00
2019	627,348 8,917	280,694 1,964	68,513 850	46,265 696	29,612 558	24,457 509	20,705 475	15,602 424	12,636 400	11,21 433

Zona sombreada a azul com os valores estimados pelo modelo Loglinear. A azul estão os valores do erro padrão das estimativas obtidas pelo modelo Loglinear (tabela 3.9). Zona sombreada a verde com os valores estimados pela metodologia *Bootstrap*. A verde estão os valores do erro padrão das estimativas.

Ao efetuar a comparação dos três modelos, verifica-se que as estimativas produzidas pelo *Bootstrap* são semelhantes às estimativas produzidas pelo modelo Loglinear, mas com um erropadrão associado superior (nomeadamente para os anos de desenvolvimento mais avançados). Ao comparar com o modelo ODP, as estimativas também são semelhantes, com um erro padrão associado normalmente inferior.

Tabela 3.23: Resultados da provisão para sinistros por ano de calendário para os anos de 2020 a 2024, aplicando os três modelos (u.m.: EUR).

Ano de	n.	ados	Modelo				
calendário	D	auos	Loglinear	ODP	ODP Bootstrap		
		:	499,184	503,626	501,370		
	2001 2010	sem inflação	[491,008; 507,360]	[496,172; 511,080]	[496,454; 506,286		
	2001-2010	com inflação	514,353	516,753	516,956		
2020		com innação	[505,430; 523,227]	[509,253; 524,253]	[511,563; 522,348		
2020		aam inflaassa	430,905	418,759	427,665		
	2005 2016	sem inflação	[394,133; 467,676]	[391,693; 445,825]	[401,944; 453,386		
	2005-2016	com inflação	457,537	444,637	454,360		
		com mnação	[416,312; 498,761]	[414,034; 475,240]	[425,378; 483,343		
		aam inflaasa	222,233	228,821	222,731		
	2001 2010	sem inflação	[219,349; 225,117]	[223,804; 233,839]	[219,649; 225,812		
	2001-2010	com inflação	230,971	234,682	231,448		
2021		com innação	[227,780; 234,163]	[229,656; 239,709]	[228,020; 234,876		
2021		sam inflacão	166,640	163,101	164,238		
	2005-2016	sem inflação	[155,870; 177,410]	[146,391; 179,811]	[149,194; 179,282		
		com inflação	175,830	170,721	172,856		
			[163,742; 187,918]	[152,128; 189,315]	[156,027; 189,685		
2022	2001-2010	sem inflação	154,535	160,372	155,057		
			[152,462; 156,608]	[156,124; 164,620]	[152,453; 157,66]		
		com inflação	170,657	173,429	171,190		
			[168,211; 173,102]	[169,231; 177,627]	[168,117; 174,263		
	2005-2016	sem inflação	118,517	116,639	116,948		
			[110,574; 126,459]	[102,332; 130,947]	[104,208; 129,688		
		:g2-	133,573	130,164	131,480		
		com inflação	[124,092; 143,054]	[114,334; 145,994]	[115,999; 146,962		
	2001 2010	sem inflação	108,439	113,750	108,958		
		sem mnação	[106,861; 110,017]	[110,117; 117,383]	[106,757; 111,159		
	2001-2010	com inflação	124,407	126,977	125,033		
2023		com mnação	[122,466; 126,347]	[123,420; 130,534]	[122,322; 127,745		
2023		sem inflação	84,756	84,396	83,417		
	2005-2016	sem mnação	[78,599; 90,913]	[72,019; 96,774]	[72,436; 94,398]		
	2003-2016	com inflação	99,969	98,614	98,054		
		com mnação	[92,311; 107,628]	[84,937; 112,292]	[84,149; 111,958		
		sem inflação	79,942	84,208	80,381		
	2001-2010	sem mnação	[78,658; 81,226]	[81,032; 87,384]	[78,459; 82,303]		
	2001-2010	com inflação	93,371	95,139	93,931		
2024			[91,761; 94,982]	[92,056; 98,222]	[91,480; 96,383]		
2027		sem inflação	62,445	62,931	61,772		
	2005 2017	sem mnação	[57,557; 67,333]	[52,050; 73,812]	[52,112; 71,432]		
	2005-2016	com inflação	75,650	75,635	74,469		
		com mnação	[69,441; 81,859]	[63,593; 87,677]	[61,911; 87,027]		

Para uma melhor perceção das estimativas e erros de cada modelo, optou-se por construir a tabela com os resultados por ano de calendário, assumindo que a partir de final de 2019 não existiriam mais apólices em vigor deste ramo. Os resultados encontram-se nas tabelas 3.23 e 3.24.

Tabela 3.24: Resultados da provisão para sinistros por ano de calendário para os anos de 2025 a 2028, aplicando os três modelos (u.m.: EUR).

Ano de	Ano de Dados		Modelo					
calendário	υ	ados	Loglinear	ODP	Bootstrap			
		aam inflaasa	56,386	59,929	56,735			
	2001 2010	sem inflação	[55,377; 57,396]	[57,192; 62,666]	[55,099; 58,371]			
	2001-2010	· a ~	66,533	68,019	66,990			
2025		com inflação	[65,251; 67,815]	[65,385; 70,653]	[64,885; 69,095]			
2023		sem inflação	42,412	43,466	42,243			
	2005 2016	sem mnação	[38,702; 46,123]	[34,201; 52,730]	[34,005; 50,481]			
	2005-2016	com inflação	52,551	53,423	51,988			
		com mnação	[47,753; 57,349]	[43,082; 63,763]	[40,965; 63,011]			
		sem inflação	36,900	39,614	37,132			
	2001 2010	sem innação	[36,127; 37,673]	[37,307; 41,920]	[35,765; 38,499]			
	2001-2010	aam inflaasa	44,085	45,239	44,439			
2026		com inflação	[43,087; 45,084]	[43,032; 47,447]	[42,663; 46,214]			
2020	2005-2016	sem inflação	26,902	28,429	26,517			
			[24,197; 29,607]	[20,695; 36,162]	[19,835; 33,200]			
		com inflação	33,544	35,081	32,855			
			[30,017; 37,071]	[26,533; 43,630]	[23,757; 41,952]			
	2001-2010	sem inflação	22,527	24,197	22,697			
			[21,937; 23,117]	[22,299; 26,095]	[21,588; 23,806]			
		com inflação	27,304	27,843	27,596			
2027			[26,529; 28,079]	[26,037; 29,648]	[26,124; 29,068]			
2027		sem inflação	14,484	15,943	14,422			
	2005 2016	sem mnação	[12,693; 16,275]	[9,849; 22,038]	[9,285; 19,559]			
	2005-2016	com inflação	18,213	19,788	17,988			
		com mnação	[15,853; 20,572]	[13,120; 26,457]	[10,955; 25,021]			
		sem inflação	11,101	11,625	11,215			
	2001 2010	sem mnação	[10,707; 11,495]	[10,205; 13,046]	[10,366; 12,063]			
	2001-2010	com inflação	13,763	13,570	13,965			
2028		com mnação	[13,238; 14,289]	[12,227; 14,912]	[12,814; 15,117]			
2028		sem inflação	5,947	6,555	6,018			
	2005 2017	sem mnação	[4,933; 6,961]	2,300; 10,810]	[2,429; 9,606]			
	2005-2016	com inflação	7,549	8,180	7,574			
		Com mnação	[6,204; 8,894]	[3,572; 12,787]	[2,623; 12,526]			

Observando as tabelas que comparam as estimativas de 2020 a 2028 produzidas pelos três modelos (tabelas 3.23 e 3.24), verifica-se que ao usar os dados das indemnizações do período de 2006 a 2015 sem incorporação da inflação, obtém-se os valores mais baixos e as estimativas de provisão mais altas são obtidas quando o triângulo de desenvolvimento que serviu de projeção tem os dados relativos ao período de 2001 a 2010 incorporando a inflação (para qualquer um dos modelos). Quando se compara as estimativas dos três modelos obtidas com o triângulo de desenvolvimento com os dados mais recentes sem considerar a inflação, verifica-se que o modelo

ODP produz valores inferiores nos anos de calendário de 2020 a 2022 e o modelo *Bootstrap* nos anos de 2023 a 2027. Já para as estimativas mais altas (utilizando os dados de 2001 a 2010 e considerando a inflação), o modelo ODP produz valores superiores. As estimativas produzidas pelo *Bootstrap* são semelhantes às produzidas pelo ODP quando o triângulo de desenvolvimento que serviu de base para a projeção é o que contém os dados mais antigos e ajustados com a inflação.

Relativamente aos intervalos de confiança a 95%, estes apresentam maior amplitude quando se usam os dados mais recentes e considerando a inflação, sendo que o modelo Loglinear produz intervalos de maior amplitude no ano de calendário de 2020, o modelo ODP para o período de 2021 a 2023 e a metodologia de *Bootstrap* para o período de 2024 a 2028.

Para uma melhor perceção das estimativas considerando os vários modelos e os triângulos de desenvolvimento que serviram de base de projeção, optou-se por representar graficamente para cada ano de calendário, as estimativas produzidas, que se encontram na figura 3.24. Nestes gráficos é possível verificar que as estimativas calculadas pelo modelo de *Bootstrap* são semelhantes às estimativas obtidas pelo método Loglinear, sendo que ao usar o triângulo de desenvolvimento com as indemnizações do período de 2001 a 2010 e incorporando a inflação, as estimativas são muito próximas, mas também são as mais elevadas, sendo que anteriormente se verificou não ter os melhores resultados quando se comparava os dados estimados com a realidade ocorrida.

Na mesma figura é claramente percetível que ao usar os dados mais recentes as estimativas são inferiores, mesmo considerando a inflação histórica e a futura. Também se observa que o modelo ODP apresenta, na maioria dos gráficos, estimativas superiores.

Note-se que apesar de à data da realização deste trabalho já se ter a informação relativa às indemnizações reais dos anos de 2020 e 2021, optou-se por não efetuar essa comparação devido à diminuição significativa de sinistralidade provocada pelas medidas de confinamento devido à pandemia da COVID-19.

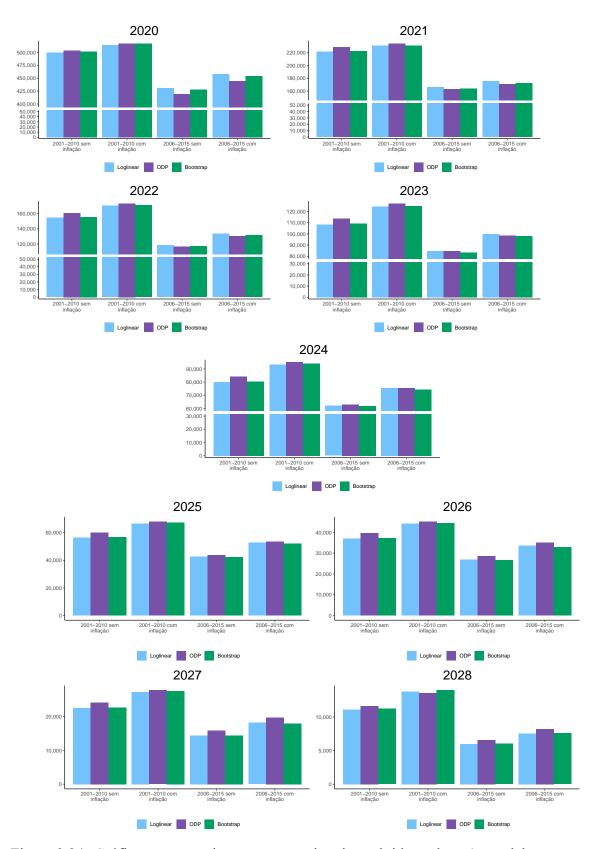


Figura 3.24: Gráficos comparativos entre as estimativas obtidas pelos três modelos por ano de calendário (u.m.: EUR).

4. Conclusão

Neste trabalho pretendeu-se efetuar estimativas para a provisão para sinistros que ocorram em anos futuros usando metodologias conhecidas para a estimação de provisões para sinistros já ocorridos mas com pagamentos ainda no futuro. Escolheram-se os modelos Loglinear, Poisson sobre dispersão (ODP) e o *Bootstrap*. Utilizaram-se os dados disponibilizados pela Autoridade de Supervisão de Seguros e Fundos de Pensões (ASF) para efetuar as projeções, sendo que foram escolhidos dois triângulos de desenvolvimento relativos às indemnizações de 2001 a 2010 e de 2006 a 2015 no ramo automóvel em Portugal. Em cada conjunto de dados, efetuaram-se as estimativas sem considerar a inflação e incorporando a inflação passada (até 2022) e futura (de 2023 a 2028). Para fazer uma análise das estimativas e ter um ponto de referência, optou-se por efetuar os cálculos até completar o ano de origem de 2019, uma vez que não se pretendia efetuar comparações com os dados de 2020 e 2021 dada a diminuição de sinistralidade provocada pela pandemia devido ao COVID-19.

Uma vez que os dados da ASF são resultados agregados de toda a industria seguradora em Portugal, os triângulos de desenvolvimento que serviram como base tinham um comportamento desejável, sem valores negativos, ou *outliers*, o que não gerou nenhum problema com a aplicação dos modelos escolhidos.

Foi elaborada uma comparação dos resultados das estimativas obtidas pelos modelos Loglinear e ODP e a realidade observada para os anos de 2016 a 2019, verificando-se que quando se compara o valor total de provisão por ano de calendário o modelo que mais se aproximou das indemnizações efetivamente pagas foi o modelo ODP considerando os dados sem inflação do período de 2006 a 2015. No entanto, quando se compara os valores individuais por ano de desenvolvimento, não é possível definir um conjunto de dados que proporcione estimativas mais próximas da realidade, uma vez que para alguns anos de calendário e anos de desenvolvimento os dados mais recentes incorporando a inflação produzem melhores aproximações, para outros anos a aproximação à realidade é verificada se forem utilizados os dados de 2001 a 2010. No entanto, a utilização dos dados menos recentes produz intervalos de confiança que na sua maioria não contêm o valor real, tanto no modelo Loglinear como no modelo ODP.

Foi também possível analisar que ao usar modelo Loglinear utilizando os dados menos recentes, as estimativas para os anos de desenvolvimento mais avançados não eram adequadas, nem o intervalo de confiança calculado. Isto pode ser explicado pelo facto de o modelo Logli-

near ter apresentado os menores erros de estimativa, e por isso os intervalos de confiança não incorporarem o valor real.

Há também a salientar que ao estar a usar um processo iterativo, em que se incorpora uma diagonal estimada pelo modelo ao triângulo anterior, o modelo diminui o cálculo do erro padrão, uma vez que no triângulo de partida já se encontram valores que o modelo esperaria, e desse modo, os intervalos de confiança diminuem a sua amplitude com o número de iterações, em vez de aumentarem como era esperado. Este comportamento é mais evidente no modelo Loglinear, mas também acontece no modelo ODP.

Quando se efetuou as comparações entre os três modelos para os anos de calendário (ou pagamento) de 2020 a 2028, foi possível verificar que ao usar o triângulo de desenvolvimento com os dados das indemnizações de 2001 a 2010 as estimativas são superiores do que quando se usam os dados mais recentes e que a inflação também provocou um aumento das provisões para sinistros.

Se forem consideradas as comparações entre os valores reais no período de 2016 a 2019, a escolha dos dados a usar para a provisão para sinistros seria do triângulo de desenvolvimento com as indemnizações de 2006 a 2015 não incorporando a inflação. No entanto, como estes dados provocam estimativas mais baixas do que os restantes conjuntos de dados e de modo a que exista algum nível de prudência, o modelo mais adequado será o modelo ODP, pois apresenta um erro padrão superior, o que consequentemente gera um intervalo de confiança com maior amplitude na generalidade dos anos de calendário a estimar. No caso das seguradoras, caberá ao atuário responsável, decidir qual o modelo a adotar e o triângulo de desenvolvimento que servirá de base à projeção, podendo, por exemplo, ainda adicionar uma margem de prudência às estimativas produzidas: considerando o valor superior do intervalo de confiança ou adicionando um valor ou percentual à estimativa central obtida.

Dentro do tema abordado neste trabalho, é ainda possível efetuar outros trabalhos futuramente, tais como considerando como base um triângulo de desenvolvimento em outro ramo não-vida, nomeadamente de acidentes de trabalho, uma vez que tem uma cauda mais pesada e que poderá influenciar os resultados, ou considerando outros modelos de estimação, nomeadamente o Modelo de Thomas Mack (em que será necessário produzir estimativas e os respetivos intervalos de confiança para o valor total as provisões para sinistros por ano de calendário), ou até mesmo, recorrendo a técnicas de *Machine Learning* quando se tem mais detalhe sobre os sinistros.

BIBLIOGRAFIA

- [1] G. Taylor, *Basic Concepts*. Boston, MA: Springer US, 2000, pp. 3–16, https://doi.org/10 .1007/978-1-4615-4583-5 1.
- [2] H. M. M. Borginho, "Metodologias estocásticas de estimação de provisões para sinistros," Master's thesis, Universidade Nova de Lisboa, 2001.
- [3] J. Lowe, "Practical guide to measuring reserve variability using: Bootstrap-ping, operational time and a distribution free approach," 1994, the 1994 General Insurance Convention.
- [4] X. G. Constitucional. (2023) Programa de estabilidade 2023-2027. [Online]. Available: https://app.parlamento.pt/webutils/docs/doc.pdf?path=6148523063484d364c793968636d3 56c6443397a6158526c63793959566b786c5a79394562324e31625756756447397a51574 e3061585a705a47466b5a564268636d786862575675644746794c7a45774d3259324f44497 94c546c695a6a41744e4441344f533169596a55324c574d335a6d4e6c4f444d79595464694 f5335775a47593d&fich=103f6822-9bf0-4089-bb56-c7fce832a7b9.pdf&Inline=true
- [5] https://www.r-project.org/.
- [6] L. B. Afonso, "Reserves, slides de apoio à disciplina de provisões para sinistros, universidade nova de lisboa," 2021.
- [7] B. Avanzi, G. Taylor, P. A. Vu, and B. Wong, "A multivariate evolutionary generalised linear model framework with adaptive estimation for claims reserving," *Insurance: Mathematics and Economics*, vol. 93, pp. 50–71, 2020.
- [8] T. Mack, "Distribution-free calculation of the standard error of chain ladder reserve estimates," *ASTIN Bulletin: The Journal of the IAA*, vol. 23, no. 2, p. 213–225, 1993.
- [9] B. Zehnwirth, "Interactive claims reserving forecasting system," Turramurra, NSW, Australia, 1985.
- [10] A. E. Renshaw, "Chain ladder and interactive modelling (claims reserving and glim)," *Journal of the Institute of Actuaries (1886-1994)*, vol. 116, no. 3, pp. 559–587, 1989. [Online]. Available: http://www.jstor.org/stable/41140947

- [11] S. Christofides, *Regression models based on log-incremental payments*. London: Institute of Actuaries, 1990.
- [12] R. J. Verrall, "Bayes and empirical bayes estimation for the chain ladder model," *ASTIN Bulletin: The Journal of the IAA*, vol. 20, no. 2, p. 217–243, 1990.
- [13] R. J. Verrall, "On the estimation of reserves from loglinear models," *Insurance: Mathematics and Economics*, vol. 10, no. 1, pp. 75–80, 1991. [Online]. Available: https://www.sciencedirect.com/science/article/pii/016766879190026T
- [14] T. S. Wright, "A stochastic method for claims reserving in general insurance," *Journal of the Institute of Actuaries (1886-1994)*, vol. 117, no. 3, pp. 677–731, 1990. [Online]. Available: http://www.jstor.org/stable/41140991
- [15] T. Mack, "A simple parametric model for rating automobile insurance or estimating ibnr claims reserves," *ASTIN Bulletin: The Journal of the IAA*, vol. 21, no. 1, p. 93–109, 1991.
- [16] P. England and R. Verrall, "Analytic and bootstrap estimates of prediction errors in claims reserving," *Insurance: Mathematics and Economics*, vol. 25, no. 3, pp. 281–293, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016766879900016
- [17] P. D. England and R. J. Verrall, "Stochastic claims reserving in general insurance," *British Actuarial Journal*, vol. 8, no. 3, pp. 443–544, 2002. [Online]. Available: http://www.jstor.org/stable/41141552
- [18] P. Cai, "Claim reserving: Classical versus machine learning methods," Master's thesis, McMaster University, 2021.
- [19] A. Despeyroux, C. Levi, C. Partrat, and J. Vignancour, "Techniques for valuation a general insurance company within the framework of ias standards: Some proposals." XXXIV International ASTIN Colloquium, Berlim, Alemanha, 2003.
- [20] P. England and R. Verrall, "A flexible framework for stochastic claims reserving," *Proceedings of the Casualty Actuarial Society Forum*, vol. 88, 01 2001.
- [21] G. Taylor and G. McGuire, "Loss reserving with glms: A case study. research," 2004, the Institute of Actuaries of Australia Accident Compensation Seminar 28 November to 1 December 2004.

- [22] M. Wüthrich, "Machine learning in individual claims reserving," *Swiss Finance Institute Research Paper Series*, no. 16-67, 2016, https://ssrn.com/abstract=2867897.
- [23] L. B. Afonso, "Provisões para sinistros documento de apoio à disciplina de gestão do risco em actuariado não vida, universidade nova de lisboa," 2015.
- [24] T. Mack, "The standard error of chain ladder reserve estimates: Recursive calculation and inclusion of a tail factor," *ASTIN Bulletin: The Journal of the IAA*, vol. 29, no. 2, p. 361–366, 1999.
- [25] J. A. Boor, "Estimating tail development factors: What to do when the triangle runs out," *CAS Forum*, vol. issue 2, pp. 345–390, 2006.
- [26] M. Merz and M. V. Wüthrich, "Estimation of tail development factors in the paid-incurred chain reserving method," *Variance, Advancing the Science of Risk*, vol. 7, no. 1, pp. 61–73, 2013.
- [27] J. Evans, "Tail factor convergence in sherman's inverse power curve loss development factor model," *Variance*, vol. 9, pp. 227–233, 2014.
- [28] PORDATA, "Taxa de inflação (taxa de variação do ipc)," acedido em 2023-04-20. [Online]. Available: https://www.pordata.pt/db/portugal/ambiente+de+consulta/tabela
- [29] J. Aitchison and J. Brown, *Lognormal Distribution*, ser. Department of Applied Economics Monographs. Cambridge University Press, 1957.
- [30] L. Tee, "Stochastic reserving methods in non-life insurance," Master's thesis, University of Tartu, 2013.
- [31] D. Hindley, *Claims Reserving in General Insurance*, ser. International Series on Actuarial Science. Cambridge University Press, 2017.
- [32] D. W. Hosmer, S. Lemeshow, and S. May, *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. John Wiley & Sons, Inc., 2008.
- [33] S. C. Strascia and A. Tripodi, "Overdispersed-poisson model in claims reserving: Closed tool for one-year volatility in glm framework," *Risks*, vol. 6, no. 4, p. 139, 2018.
- [34] M. Goldburd, A. Khare, D. Tevet, and D. Guller, *Generalized Linear Models for Insurance Rating*. Casualty Actuarial Society, 2016, https://books.google.pt/books?id=HHAayAE ACAAJ.

- [35] B. Efron, "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1 26, 1979. [Online]. Available: https://doi.org/10.1214/aos/1176344552
- [36] D. da Costa, "Metodologias de estimação de provisões para sinistros do ramo não vida," Master's thesis, Universidade de Lisboa, 2016.
- [37] P. J. R. Pinheiro, J. M. Andrade e Silva, and M. De Lourdes Centeno, "Bootstrap methodology in claim reserving," *Journal of Risk and Insurance*, vol. 70, no. 4, pp. 701–714, 2003. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.00 22-4367.2003.00071.x
- [38] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and their Application*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.
- [39] ANSR, "Relatório de sinistralidade e fiscalização de 2021," 2022. [Online]. Available: http://www.ansr.pt/Estatisticas/RelatoriosDeSinistralidade/Documents/2021/RelatorioSinistralidadeFiscalizacaoDezembro2021-ISBN.pdf
- [40] ASF. https://www.asf.com.pt/NR/exeres/2C86501F-8840-4968-9C3F-FCB10D0AD37C. htm.

A. Anexos

A.1 Verificação dos pressupostos do modelo Loglinear

Na tabela A.1 encontram-se os pressupostos para aplicação do modelo Loglinear e medidas de ajustamento ao modelo. Apesar de em algumas iterações não ser possível garantir a normalidade dos resíduos, optou-se pela aplicação da metodologia.

Tabela A.1: Verificação dos pressupostos para aplicação do modelo Loglinear.

Dados	Inflação	Ano de calendário	Nº coeficientes não significativos¹ mas considerados no modelo	Erro padrão dos resíduos (%)	R ajustado	p-value do teste estatístico de Kolmogorov Smirnov
		2016	2	10.55%	99.49%	24.0%
	Sim	2017	2	8.65%	99.67%	37.3%
	Siiii	2018	4	7.76%	99.74%	33.2%
2006-2015		2019	2	6.42%	99.83%	12.0%
2000-2013		2016	4	10.8%	99.45%	18.1%
	Não	2017	4	8.64%	99.66%	28.5%
	INao	2018	5	7.52%	99.75%	21.2%
		2019	6	6.14%	99.83%	12.8%
	Sim	2011	1	7.11%	99.72%	58.1%
		2012	2	6.64%	99.76%	18.8%
		2013	1	6.10%	99.87%	21.1%
		2014	3	5.72%	99.84%	14.6%
		2015	2	5.33%	99.86%	1.3%
		2016	2	4.97%	99.88%	0.7%
		2017	2	4.55%	99.90%	0.1%
		2018	1	3.93%	99.92%	0.1%
2001 2010		2019	1	1.38%	99.99%	12.3%
2001-2010		2011	3	7.81%	99.64%	18.7%
		2012	3	7.06%	99.72%	26.9%
		2013	3	6.51%	99.77%	12.8%
		2014	3	6.02%	99.81%	0.5%
	Não	2015	3	5.62%	99.83%	0.3%
		2016	4	5.08%	99.87%	0.4%
		2017	2	4.56%	99.89%	0.2%
		2018	2	3.89%	99.92%	0.1%
		2019	2	1.28%	99.99%	2.2%

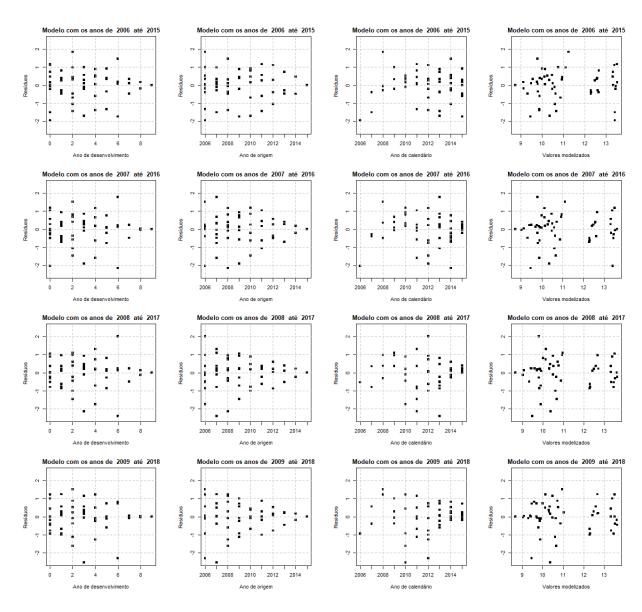


Figura A.1: Análise de tendência por parte dos resíduos das várias iterações aplicando o modelo Loglinear aos dados do triângulo de desenvolvimento com as indemnizações de 2006 a 2015 e considerando a inflação.

¹Abaixo de 0.05.

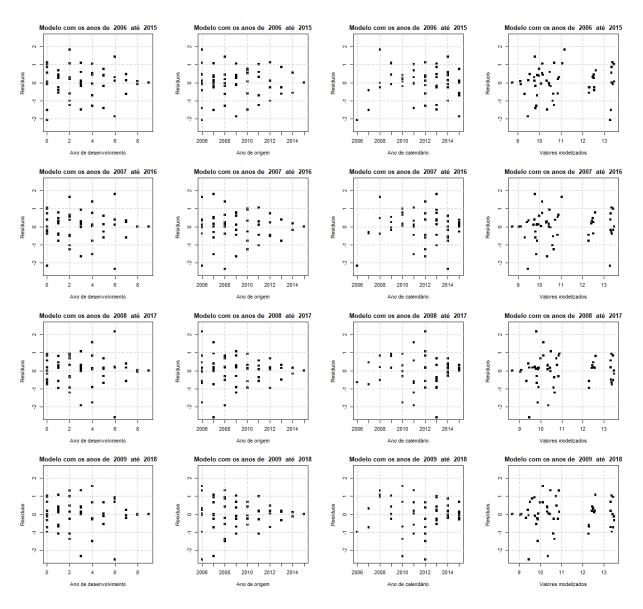


Figura A.2: Análise de tendência por parte dos resíduos das várias iterações aplicando o modelo Loglinear aos dados do triângulo de desenvolvimento com as indemnizações de 2006 a 2015 e sem incorporar a inflação.

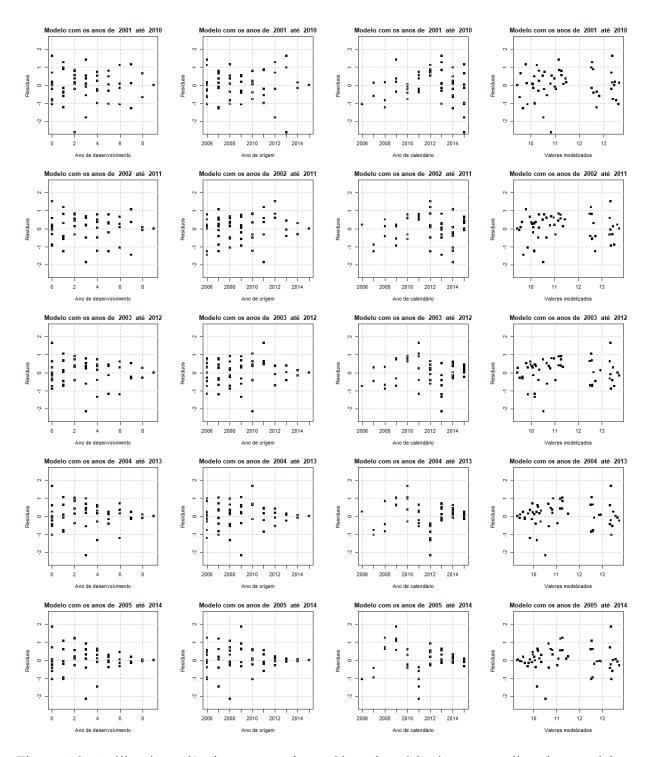


Figura A.3: Análise de tendência por parte dos resíduos das várias iterações aplicando o modelo Loglinear aos dados do triângulo de desenvolvimento com as indemnizações de 2001 a 2010 e considerando a inflação. Iterações de 1 a 5.

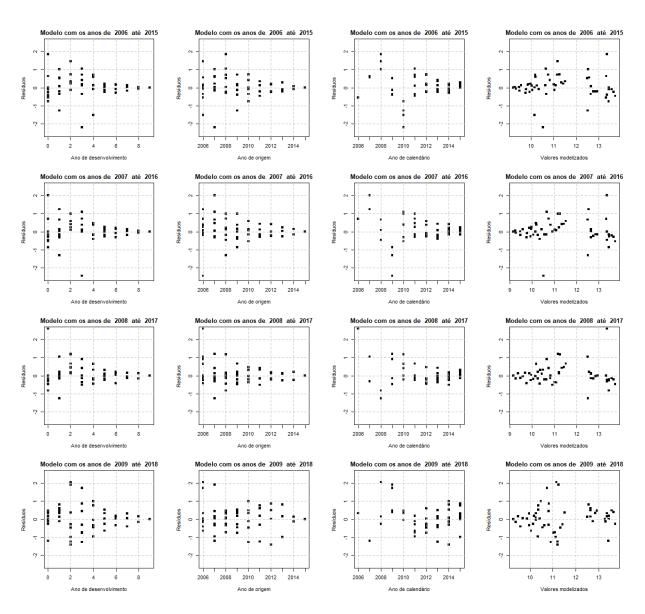


Figura A.4: Análise de tendência por parte dos resíduos das várias iterações aplicando o modelo Loglinear aos dados do triângulo de desenvolvimento com as indemnizações de 2001 a 2010 e considerando a inflação. Iterações de 6 a 9.

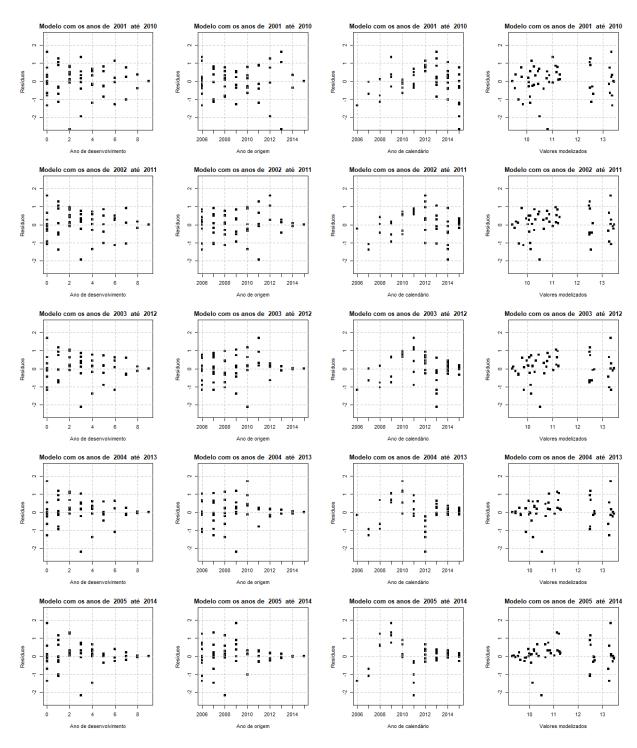


Figura A.5: Análise de tendência por parte dos resíduos das várias iterações aplicando o modelo Loglinear aos dados do triângulo de desenvolvimento com as indemnizações de 2001 a 2010 e sem incorporar a inflação. Iterações de 1 a 5.

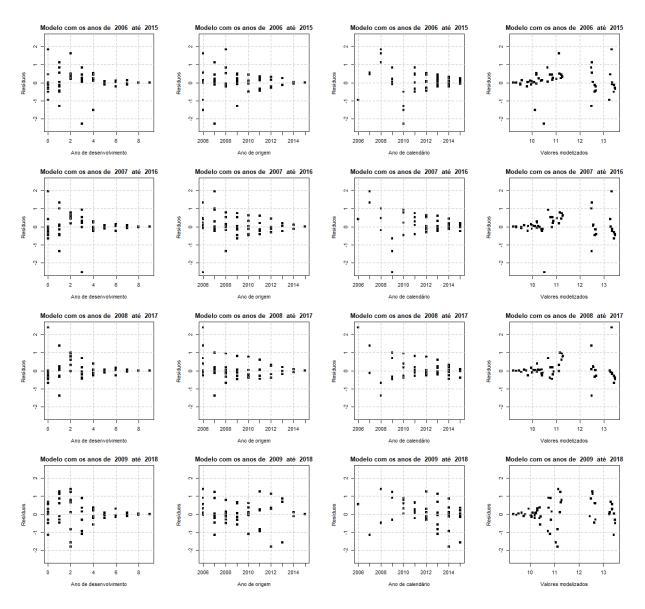


Figura A.6: Análise de tendência por parte dos resíduos das várias iterações aplicando o modelo Loglinear aos dados do triângulo de desenvolvimento com as indemnizações de 2001 a 2010 e sem incorporar a inflação. Iterações de 6 a 9.

A.2 Verificação dos pressupostos do modelo de Poisson sobredispersão

Na tabela A.2 encontram-se os pressupostos para aplicação do modelo ODP e medidas de ajustamento ao modelo. Apesar de em algumas iterações não ser possível garantir a normalidade dos resíduos, optou-se pela aplicação da metodologia.

Tabela A.2: Verificação dos pressupostos para aplicação do modelo de Poisson sobre-dispersão.

	Inflação		Nº coeficientes		<i>p-value</i> do
Dadaa		Ano de	não significativos ²	D: 41	teste estatístico
Dados		calendário	mas considerados	Deviance residual	de Kolmogorov
			no modelo		Smirnov
		2016	5	44,878	41.6%
	Sim	2017	5	27,777	33.9%
	Siiii	2018	4	18,122	19.6%
2006-2015		2019	4	15,727	8.1%
2006-2013		2016	8	44,210	35.4%
	Não	2017	8	26,072	46.4%
	INão	2018	6	15,704	28.2%
		2019	4	13,017	4.9%
	Sim	2011	1	16,541	16.8%
		2012	1	14,812	8.4%
		2013	0	11,900	12.7%
		2014	1	11,064	2.3%
		2015	1	9,837	0.4%
		2016	3	8,689	0.4%
		2017	3	7.270	0.1%
		2018	3	5,733	0.1%
2001-2010		2019	3	800	14.5%
2001-2010		2011	3	19,113	21.5%
		2012	3	15,679	9.1%
		2013	4	13,474	7.4%
		2014	4	12,117	0.5%
	Não	2015	5	11,110	0.2%
		2016	3	9,339	0.1%
		2017	3	7,094	0%
		2018	3	5,268	0%
		2019	0	817	0.1%

²Abaixo de 0.05.

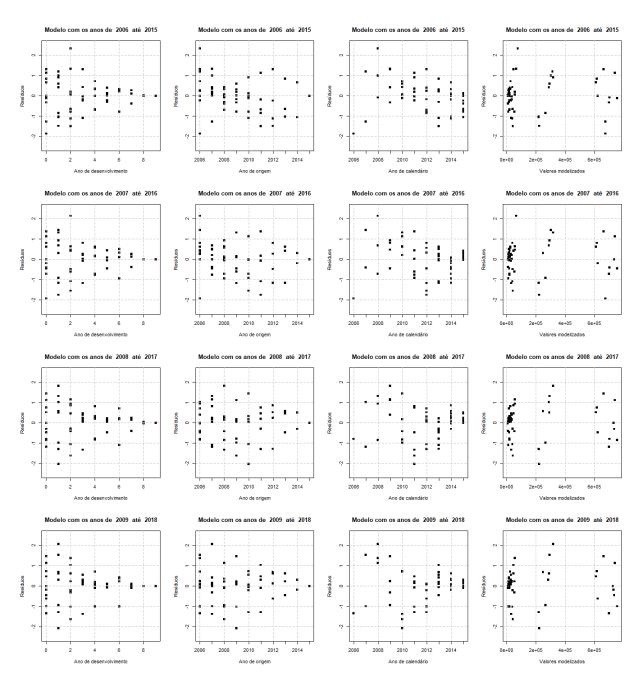


Figura A.7: Análise de tendência por parte dos resíduos das várias iterações aplicando o modelo ODP aos dados do triângulo de desenvolvimento com as indemnizações de 2006 a 2015 e considerando a inflação.

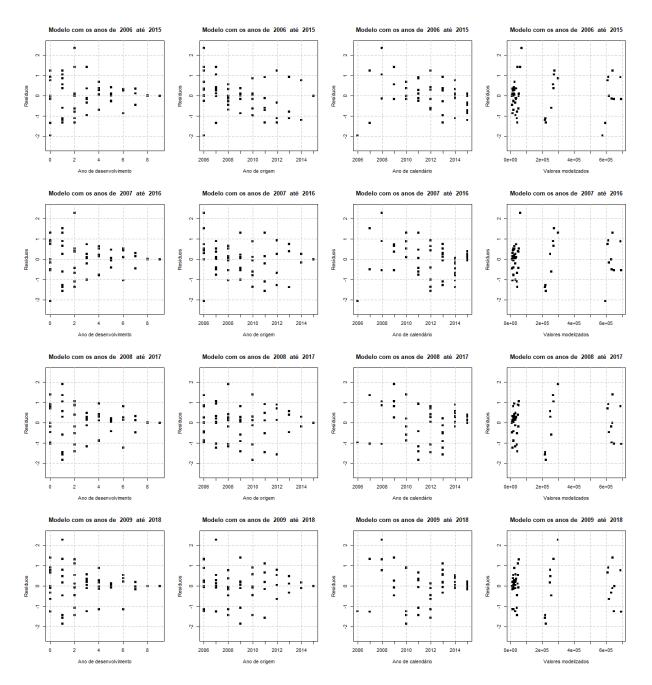


Figura A.8: Análise de tendência por parte dos resíduos das várias iterações aplicando o modelo ODP aos dados do triângulo de desenvolvimento com as indemnizações de 2006 a 2015 e sem incorporar a inflação.

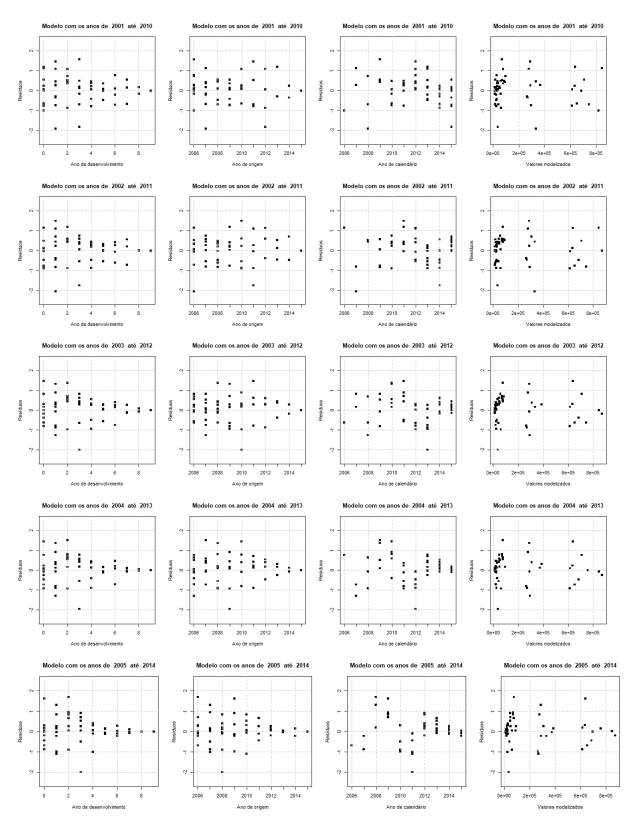


Figura A.9: Análise de tendência por parte dos resíduos das várias iterações aplicando o modelo ODP aos dados do triângulo de desenvolvimento com as indemnizações de 2001 a 2010 e considerando a inflação. Iterações de 1 a 5.

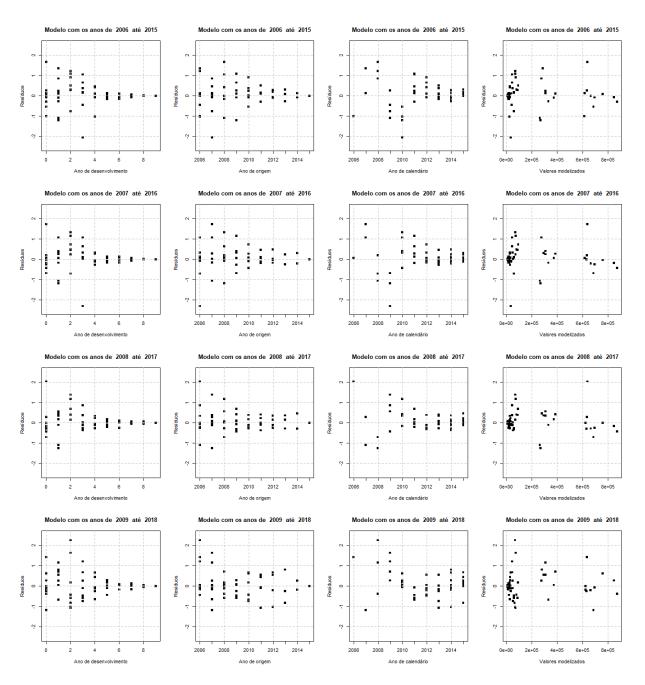


Figura A.10: Análise de tendência por parte dos resíduos das várias iterações aplicando o modelo ODP aos dados do triângulo de desenvolvimento com as indemnizações de 2001 a 2010 e considerando a inflação. Iterações de 6 a 9.

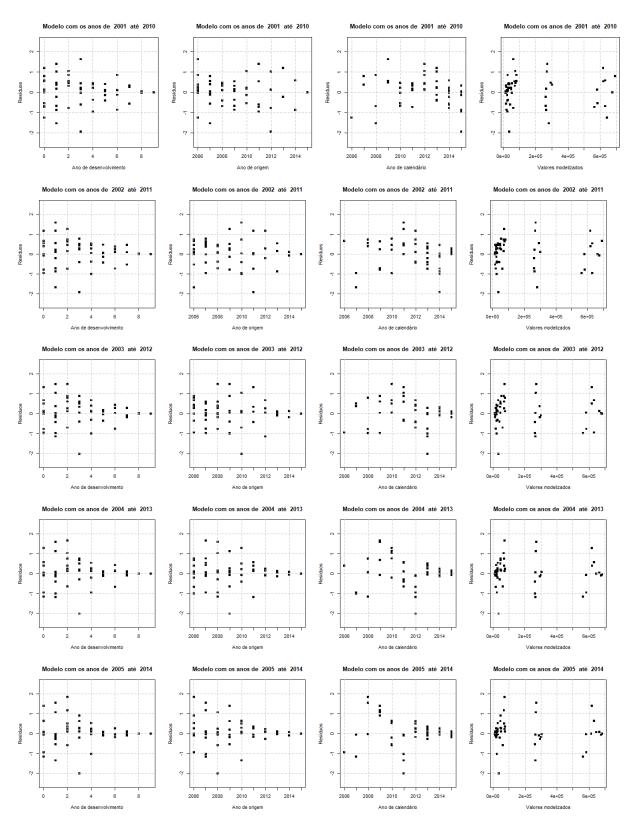


Figura A.11: Análise de tendência por parte dos resíduos das várias iterações aplicando o modelo ODP aos dados do triângulo de desenvolvimento com as indemnizações de 2001 a 2010 e sem incorporar a inflação. Iterações de 1 a 5.

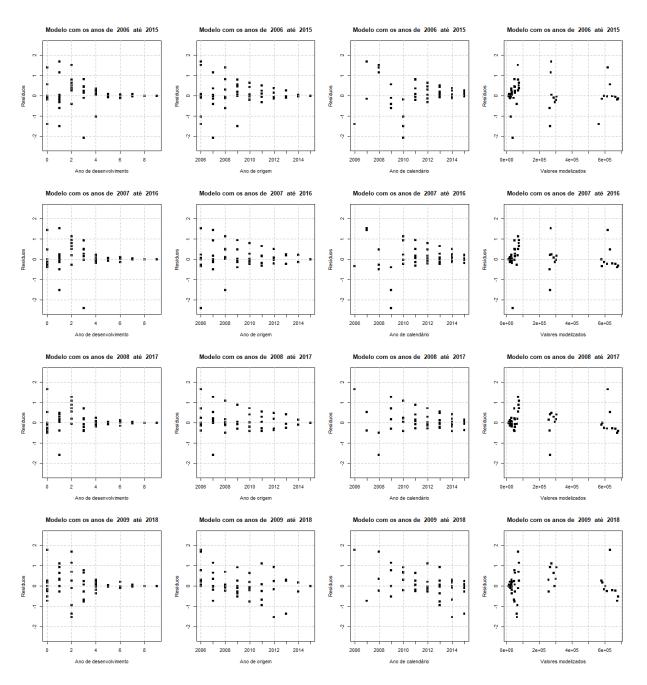


Figura A.12: Análise de tendência por parte dos resíduos das várias iterações aplicando o modelo ODP aos dados do triângulo de desenvolvimento com as indemnizações de 2001 a 2010 e sem incorporar a inflação. Iterações de 6 a 9.