



Instituto Superior de Economia e Gestão

UNIVERSIDADE TÉCNICA DE LISBOA

DESDE 1911

MESTRADO

DECISÃO ECONÓMICA E EMPRESARIAL

TRABALHO FINAL DE MESTRADO

DISSERTAÇÃO

**MODELOS LINEARES GENERALIZADOS TWEEDIE APLICADOS AO
CÁLCULO DE PROVISÕES PARA SINISTROS**

ANDRÉ PEREIRA POUSINHO

ORIENTAÇÃO:

PROFESSOR DOUTOR JOSÉ MANUEL DE MATOS PASSOS

Júri:

Presidente: Professor Doutor António Da Ascensão Costa

Arguente: Professora Doutora Nicoletta Rosati

Vogal: Professor Doutor José Manuel De Matos Passos

DEZEMBRO - 2013

“Education is the most powerful weapon which you can use to change the world.”

Nelson Mandela (1918 – 2013)

Agradecimentos

Ao meu orientador, Professor Doutor José Manuel de Matos Passos, pelo apoio e disponibilidade demonstrada em fazer parte deste trabalho.

À Professora Doutora Nicoletta Rosati pelo acompanhamento e apoio na fase preliminar desta dissertação.

Ao Departamento de Gestão de Risco e Actuariado Não Vida da Companhia de Seguros Fidelidade, na pessoa da Dra. Luísa Santos, pela disponibilidade e autorização do uso de informação.

Aos meus colegas e amigos que sempre me apoiaram e contribuíram para a elaboração deste trabalho, em especial à Ana Cristina Praxedes pela preciosa ajuda na exposição das ideias.

A toda a minha família, em especial à minha irmã e aos meus pais pela oportunidade e apoio incondicional durante a elaboração deste trabalho.

À Ana Margarida Luís, pela paciência e apoio demonstrado ao longo de todo este ano.

**MODELOS LINEARES GENERALIZADOS TWEEDIE APLICADOS AO CÁLCULO DE PROVISÕES PARA
SINISTROS**

Por André Pereira Pousinho

Mestrado em: Decisão Económica e Empresarial

Orientação: Professor Doutor José Manuel De Matos Passos

Resumo

Em 2013, ano internacional da estatística, é elaborada esta dissertação que relaciona o cálculo atuarial com a área da estatística. Partindo das vantagens e desvantagens de métodos como os Modelos Lineares Generalizados, é pretendido elucidar e estabelecer uma ligação entre as necessidades das empresas seguradoras e os estudos efetuados para o provisionamento da reserva para sinistros. Para a aplicação dos Modelos Lineares Generalizados no cálculo de provisões técnicas para sinistros é utilizada a distribuição Tweedie, pertencente à família de dispersão exponencial. Em estudos como este, esta distribuição apresenta uma boa flexibilidade de ajustamento aos dados, permitindo obter estimativas com melhor rigor.

Palavras-chave: companhia de seguros; provisão técnica; reserva para sinistros; Modelos Lineares Generalizados; distribuição Tweedie; modelos estocásticos; Atuário

TWEEDIE GENERALIZED LINEAR MODELS APPLIED TO THE CALCULATION OF CLAIMS OUTSTANDING

RESERVE

By André Pereira Pousinho

Master Degree: Economic and Corporate Decision-making

Supervisor: Professor José Manuel De Matos Passos

Abstract

In 2013, the international year of statistics, this paper is developed to connect the actuarial calculation with the field of statistics. Starting with the advantages and disadvantages of methods like the Generalized Linear Models, it is intended to clarify and establish a link between the needs of insurance companies and studies carried out for the provisioning of claims outstanding reserve. The Tweedie distribution belongs to the family from exponential dispersion and it is used for the implementation of Generalized Linear Models in the calculation of claims technical provisions. In researches like these, this distribution has a good flexibility in the adjustment to data, allowing us to get estimations with better accuracy.

Keywords: insurance companies; technical reserve; claims outstanding reserve; Generalized Linear Models; Tweedie distribution; stochastic models; Actuary

Glossário

Atuário – do latim *actuarius* “escrivão”. Técnico especializado na aplicação de cálculos estatísticos e matemáticos a operações financeiras, especialmente no estabelecimento e gestão de regimes de pensões e seguros.

Companhia de Seguros, Empresa de Seguros, Seguradora – Entidade legalmente constituída para assumir e gerir riscos, devidamente especificados no contrato de seguro.

Cosseguero – Contrato de seguro realizado por duas ou mais Companhias de Seguros referente ao mesmo risco. É uma forma de repartir uma responsabilidade de grandes proporções por várias seguradoras.

ISP – Instituto de Seguros de Portugal – entidade de supervisão da atividade seguradora em Portugal.

Prémio – Preço pago pelo Tomador de Seguro à empresa de seguros pela contratação do seguro.

Provisões Técnicas – O montante das provisões técnicas deve, em qualquer momento, ser suficiente para permitir à seguradora cumprir, na medida do razoavelmente previsível, os compromissos decorrentes dos contratos de seguro. Somas obrigatoriamente inscritas no passivo do balanço de uma empresa de seguros ou de resseguros.

Resseguro – Operação que uma empresa de seguros realiza para segurar parte dos riscos que assume.

Sinistro – Evento ou série de eventos, resultantes de uma mesma causa, suscetível de fazer funcionar as garantias de um ou mais contratos de seguro.

Solvência de uma Seguradora – É a capacidade da seguradora honrar os seus compromissos futuros. De forma mais objetiva, uma seguradora será considerada solvente se o fluxo de receitas futuras adicionado à sua Margem de Solvência ultrapassar o fluxo de pagamentos futuros.

Índice

Introdução	3
1 Mercado segurador	5
1.1 Seguros do ramo Não-Vida	5
2 Enquadramento	7
3 Modelos Lineares Generalizados (MLG)	11
3.1 Definição de um Modelo Linear Generalizado	11
3.2 Família de dispersão exponencial	13
4 Distribuição Tweedie	15
4.1 Características da distribuição Tweedie	15
4.2 Casos particulares da distribuição Tweedie.....	17
4.2.1 Distribuição Normal	17
4.2.2 Distribuição Gamma	18
4.2.3 Distribuição Poisson.....	18
4.3 Aplicação Prática 1.....	19
4.3.1 Análise preliminar dos dados.....	19
4.3.2 Testes de correlação	20
4.4 Estimar p através de Máxima Verosimilhança	21
4.5 Aplicação Prática 2.....	22
4.5.1 Estimação da constante p	22
5 Estimação dos parâmetros	24
5.1 Estimação de parâmetros dos MLG	24
5.2 Aplicação Prática 3.....	25
5.2.1 Estimar modelo com $p = p$	25
6 Adequabilidade do modelo.....	27
6.1 Resíduos.....	27
6.1.1 Casos particulares de resíduos	28
6.2 Análise dos resíduos	29
6.3 Aplicação prática 4.....	31
6.3.1 Adequabilidade do modelo com $p = p$	31
6.3.2 Adequabilidade do modelo com $p=1$, Poisson	32
6.3.3 Adequabilidade do modelo com $p=2$, Gamma	34
6.3.4 Resultados obtidos.....	36

7	Conclusões	37
	Anexos	A
	Bibliografia	42

Introdução

Tanto no início do negócio segurador, como nos dias de hoje, sempre existiu a necessidade de transferir determinados riscos para uma outra entidade, a troco de um prémio. A transferência desse risco garante aos indivíduos a eliminação de parte (ou a totalidade) dos encargos provenientes de acontecimentos ocasionais indesejáveis. As entidades que aceitam adquirir esses riscos são as atuais companhias de seguros. Esta realidade origina enormes quantidades de risco, provenientes da celebração de contratos de seguro em massa, que as companhias de seguros na sua atividade, têm que suportar. A celebração destes contratos origina, como responsabilidade das companhias, grandes quantidades de capital seguro. De forma a poder minimizar esse risco, as seguradoras pretendem homogeneizar quer quantitativamente, quer qualitativamente os riscos inerentes aos capitais seguros. Esta preocupação é positiva para estas entidades uma vez que as leva a tomar medidas no sentido de diminuir e controlar esse risco, tais como a realização de contratos de cosseguro e de resseguro, e a especialização em determinados ramos da atividade seguradora. No entanto, o risco tem que ser gerido essencialmente no seio das empresas seguradoras. Essa gestão passa essencialmente por garantir que, no futuro, as companhias estão preparadas para eventualidades provenientes das suas carteiras de contratos de seguro.

Ainda que exista a obrigatoriedade do seu cálculo, as provisões técnicas garantem à companhia uma boa resposta às suas obrigações no futuro, consequentes dos contratos de seguro celebrados até então. Atualmente, para a estimação das provisões técnicas são habitualmente utilizados nas seguradoras portuguesas métodos determinísticos e estocásticos como o tradicional *Chain Ladder*, o *Thomas Mack* e o *Bootstrap*. No caso do

método determinístico *Chain Ladder*, os seus resultados são obtidos a partir das evoluções verificadas no passado para a resolução dos sinistros.

No sentido de obter novos métodos para o cálculo de estimativas para as provisões dos custos com sinistros, este trabalho consiste no estudo da aplicação de Modelos Lineares Generalizados na obtenção dessas estimativas. Os Modelos Lineares Generalizados (MLG) apresentam a vantagem e flexibilidade de se poder construir e adequar um modelo que melhor se ajuste aos dados de que dispomos, que no caso do tema abordado representam o historial dos pagamentos efetuados com sinistros. Dado que a componente estocástica dos Modelos Lineares Generalizados pressupõe uma distribuição da família exponencial, é necessário proceder à escolha de uma distribuição dessa família que melhor se ajuste aos dados. Esta escolha pode passar pelas distribuições Gama, Poisson, Normal, Exponencial, etc... .

Os dados a trabalhar podem ter o valor zero, o que acarreta um maior cuidado na escolha de uma distribuição da família exponencial. Uma boa opção seria atribuir-se maior preferência à distribuição Poisson, ou à distribuição Gama, com o intuito de tentar modelar os “zeros” existentes nos dados e assim reduzir o seu impacto nas estimativas, ou uma outra distribuição que, dentro destes pressupostos, melhor se adapte aos dados. Essa opção incide sobre a distribuição Tweedie, uma distribuição que apresenta as características de uma dispersão exponencial. Essencialmente, a Tweedie permite a sua utilização como forma de uma distribuição entre a Poisson e a Gama, o que abrange um maior leque de distribuições e nos permite obter a distribuição que melhor se adapta aos dados disponíveis.

1 Mercado segurador

O negócio segurador é dos mais antigos de que há registo. A ideia deste negócio terá surgido em 4500 A.C., época em que os transportes fluviais eram feitos por intermédio de barcas que não possuíam segurança nem estabilidade. Como este meio de transporte era bastante suscetível a acidentes, tais como incêndios e naufrágios, houve a necessidade de minimizar estes riscos, acordando com determinados indivíduos ou entidades que assumiam parte ou a totalidade desse risco. Os acordos celebrados consistiam em garantir a cobertura dos sinistros, ou parte deles, habitualmente a troco de parte da mercadoria transportada. Esta era a forma de minimizar os impactos negativos provenientes dos sinistros. Em 1318, surgiu pela primeira vez uma legislação sobre seguros, sendo que emissão da primeira apólice foi em 1347, em Génova, de um contrato de seguro de transportes marítimos.

O início de contratos de seguro no nosso país teve percurso semelhante ao do restante mundo. Iniciou-se com os seguros marítimos e, em 1370, o rei D. Fernando I promulga as primeiras leis de seguros de Portugal. Foi em 1797 que surgiu a primeira companhia em território português, a companhia de seguros “Bom Conceito”, que posteriormente deu origem à conhecida Bonança (atualmente incorporada no Grupo Caixa Seguros).

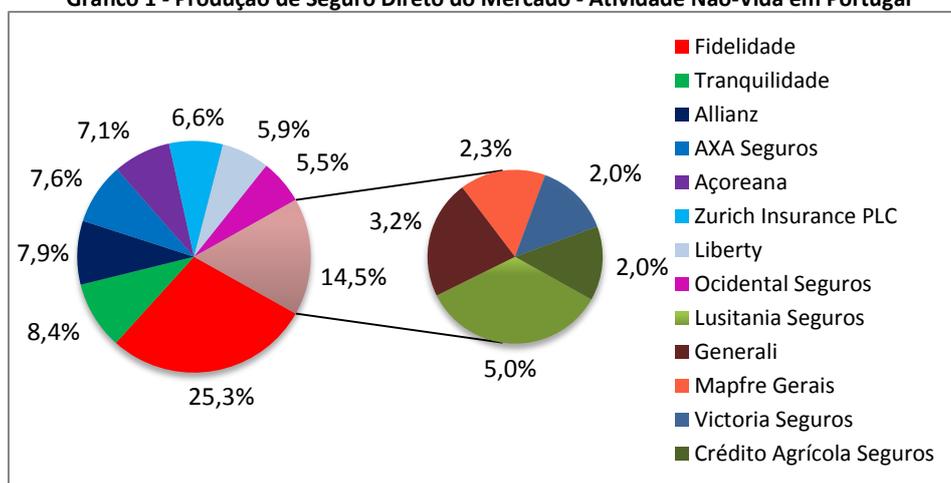
1.1 Seguros do ramo Não-Vida

Conforme Portugal L. (2007), os seguros Não-Vida, ou seguros reais, como por vezes também designados são todos os tipos de seguro que reembolsam, indemnizam ou prestam serviço ao segurado ou a um terceiro, no seguimento da resolução do sinistro que seja responsabilidade da seguradora em causa. O ramo Automóvel é o que possui

maior peso dos Não-Vida, sendo seguido pelos ramos de Acidentes de Trabalho, Doença e de Incêndio e Outros Danos, tal como pode ser verificado no Anexo 1.i. Nos ramos Não-Vida em 2012, relativamente ao ano transato, o volume de prémios teve uma variação de -3%, segundo o Instituto de Seguros de Portugal (2013). Conforme publicado pela Associação Portuguesa de Seguradores (2013), já no primeiro trimestre de 2013, a variação acentuou-se ainda mais, situando-se nos -4,3%, sendo um pouco o reflexo do resultado da conjuntura socioeconómica que tem provocado uma diminuição do poder de compra dos consumidores. Simultaneamente, os custos com sinistros aumentaram 4.8%, nesse trimestre, muito por culpa dos fortes temporais verificados durante o Inverno de 2012/2013, em que as indemnizações atingiram o nível de 100 milhões de euros.

Segundo as informações mais recentes de quotas de mercado, é possível reparar que tal como tem sucedido nos últimos três anos, o Grupo Caixa Seguros, agora denominado Fidelidade Companhia de Seguros, tem liderado este mercado, tanto nos ramos Vida, como nos Não-Vida. Direccionando o foco para os ramos Não-Vida, a Fidelidade Companhia de Seguros detém uma quota de 25,3%, seguida da Companhia de Seguros Tranquilidade e da Companhia de Seguros Allianz Portugal, S.A., com 8,4% e 7,9%, respetivamente.

Gráfico 1 - Produção de Seguro Direto do Mercado - Atividade Não-Vida em Portugal



2 Enquadramento

De seguida, procedemos a um enquadramento do processo de gestão de sinistros. A gestão do risco é a base e a razão da existência de estudos como este. Parte das necessidades de uma seguradora surgem da responsabilidade que esta tem para com a gestão de sinistros. Tal como é expectável, a atividade de uma companhia de seguros não é moldada apenas pelos prémios, mas também pelos pagamentos que serão efetuados pela ocorrência de sinistros. Por essa razão, a seguradora tem de estar preparada para a obrigatoriedade quer de pequenos, quer de grandes volumes de pagamentos a serem feitos. Ou seja, a companhia de seguros, após o conhecimento da ocorrência do sinistro, tem de possuir recursos suficientes para fazer face a todas as despesas inerentes a esse sinistro, até este ser considerado encerrado. No entanto, a participação dos sinistros, muitas vezes, não é efetuada num período de tempo relativamente pequeno após a sua ocorrência. Isto é algo que dificulta a análise, pois como se não bastasse a aleatoriedade da ocorrência dos sinistros, também se torna impossível saber os atrasos na sua participação e na sua regularização. Estas dificuldades levantam algumas dúvidas, tais como:

- ❖ Dado que por vezes a seguradora tem a impossibilidade de saber instantaneamente das ocorrências dos sinistros, quanto tempo passa até os sinistros serem participados?
- ❖ Mediante, muitas vezes, a gravidade dos sinistros, qual é o volume de pagamentos que terá de ser efetuado até à sua resolução?
- ❖ Após os pagamentos dos sinistros serem iniciados, qual o intervalo de tempo que passa até que os sinistros sejam regularizados?

Estas três questões, naturalmente, não possuem uma resposta certa. O que nos incute a necessidade de as tentar prever, com base no passado. Cabe aos atuários, a responsabilidade de efetuar cálculos, com base em métodos, estatísticos e determinísticos, para obter a reserva necessária para a seguradora fazer face às suas responsabilidades futuras. E assim tentar, obter a melhor aproximação das respostas às dúvidas acima referidas. A esta tarefa é habitualmente designada por provisionamento, pois as reservas da seguradora correspondem às suas provisões técnicas.

O cálculo de provisões técnicas provém dos métodos aplicados e desenvolvidos pelos atuários. Garantem à companhia de seguros estimativas adequadas, a fim de honrar os compromissos que aquela assume perante os clientes. Todavia, existe uma grande complexidade no cálculo destes valores que acarreta vários custos: no caso de as provisões serem excessivas, a companhia terá de sustentar enormes custos, correndo o risco de afetar a sua rentabilidade; se, por outro lado, forem insuficientes, corre-se o risco de enfrentar, no futuro, um cenário de insolvência. No ramo Não-Vida, as provisões técnicas são compostas por diferentes tipos de provisões: provisão para desvios de sinistralidade, provisão para prémios não adquiridos; provisão para riscos em curso; provisão para sinistros. Nesta dissertação, iremos focar-nos apenas nas provisões para sinistros e no seu cálculo. Esta provisão subdivide-se em três categorias: provisões para sinistros já declarados, provisões para sinistros ocorridos mas diminutamente provisionados (IBNER - *incurred but not enough reported*) e provisões para sinistros ocorridos mas não declarados (IBNR - *incurred but not reported*).

Este trabalho é elaborado com o objetivo de se poder estudar a aplicação dos Modelos Lineares Generalizados, com a distribuição Tweedie, no cálculo das provisões

técnicas para sinistros. Esta aplicação é feita no sentido de poder obter previsões para a variável aleatória objeto de estudo, que se considera ser os montantes pagos com sinistros. De forma a melhorar a interpretação e organização dos dados, iremos de seguida expor como é habitualmente apresentada essa informação.

Assumimos que os pagamentos de sinistros, Y , são variáveis aleatórias i.i.d. Intuitivamente, esta variável pertence ao conjunto \mathbb{R}_0^+ , assumindo o pressuposto de que possam existir reembolsos passados, mas que estes não excedem o total de pagamentos efetuados. A análise efetuada baseia-se numa visão global e anual dos sinistros, e não sinistro-a-sinistro. Por este motivo, na aplicação prática efetuada nesta dissertação, as principais variáveis explicativas dos montantes pagos, são os anos em que ocorreram os sinistros, i (com $1 \leq i \leq N$), e o período existente entre a ocorrência e o pagamento do sinistro, j (com $0 \leq j \leq (N - 1)$). Para o cálculo atuarial, é bastante comum dispor estas variáveis numa matriz incompleta, a que chamamos “Matriz *Run Off*”:

Tabela I - Matriz Run Off
Ano de desenvolvimento j

		0	1	2	...	N-2	N-1
Ano de ocorrência i	Ano 1	$y_0(\text{Ano 1})$	$y_1(\text{Ano 1})$	$y_2(\text{Ano 1})$...	$y_{N-2}(\text{Ano 1})$	$y_{N-1}(\text{Ano 1})$
	Ano 2	$y_0(\text{Ano 2})$	$y_1(\text{Ano 2})$	$y_2(\text{Ano 2})$...	$y_{N-2}(\text{Ano 2})$	
	Ano 3	$y_0(\text{Ano 3})$	$y_1(\text{Ano 3})$	$y_2(\text{Ano 3})$...		
	⋮	⋮	⋮	...			
	Ano N-1	$y_0(\text{Ano N-1})$	$y_1(\text{Ano N-1})$				
	Ano N	$y_0(\text{Ano N})$					

Como N é o número de anos de ocorrência existentes na matriz, sabe-se que o valor total de observações existentes de y_i irá ser sempre, $\Psi = \frac{N(N+1)}{2}$.

Nestas matrizes, é possível identificar facilmente o período que decorre até à regularização da generalidade dos sinistros, ou seja o intervalo de tempo médio que os sinistros estão em *Run Off*. A utilização dos dados sob a forma de matriz auxilia a

aplicação e interpretação de métodos estocásticos e determinísticos, tais como o *Bootstrap*, *Thomas Mack*, *Chain Ladder*, assim como o método objeto de estudo. A aplicação destes métodos servem para obter estimativas futuras dos montantes pagos, que correspondem à área sombreada da matriz. Estes métodos têm como base as evoluções verificadas no passado, para poder estimar as do futuro. No caso do *Chain Ladder*, é também avaliada a evolução verificada entre colunas de desenvolvimento e atribuída uma ponderação a cada uma delas. Estes fatores correspondem à proporção de um ano de desenvolvimento em relação ao anterior. Este método tem como grande vantagem a sua simplicidade aliada às suas boas estimativas, no entanto este processo de cálculo não considera a possibilidade de avaliar o erro de estimação ou o grau de incerteza relativamente às estimativas obtidas, o que não nos permite quantificar o nível de prudência que pretendemos refletir no processo de provisionamento. O resultado do cálculo das provisões técnicas é o volume de pagamentos estimado para a área sombreada presente na Tabela I. Esse valor total, será designado por reserva, e poderá ser apresentado pela seguinte expressão:

$$Reserva = \sum_{i=2}^N \sum_{j=N+1-i}^N \mu_{ij}, \quad (1.1.i)$$

sendo μ_{ij} o valor esperado de pagamentos de sinistros ocorridos no ano i e pago no ano $i+j$.

Após obter as provisões técnicas, interessa apresentar os custos totais com os sinistros como resultados. Este interesse provém da necessidade de se tomar conhecimento da despesa total, que os sinistros ocorridos num determinado ano gerem à companhia de seguros. O custo total com os sinistros é o resultado da soma do volume de pagamentos que a seguradora já fez e espera vir a fazer até os sinistros estarem regularizados.

3 Modelos Lineares Generalizados (MLG)

Os Modelos Lineares Generalizados, apresentados por Nelder & Wedderburn (1972) são casos específicos de modelos de regressão, que têm como principal objetivo o estudo da relação entre variáveis, mais especificamente, a análise do efeito que uma ou mais variáveis explicativas provocam sobre a variável de interesse, que designamos habitualmente por variável resposta.

Os MLG surgiram com o intuito de sintetizar numa só generalização uma vasta gama de modelos existentes à época. Foram uma simplificação dos estudos e aplicações feitas, até à altura, em diversas áreas da ciência como a genética e a física, assim como também nas áreas das telecomunicações móveis, logística e comércio, por exemplo. Esta dissertação, estuda a aplicação dos Modelos Lineares Generalizados no cálculo atuarial efetuado nas seguradoras, mais concretamente no cálculo de estimativas para o provisionamento de sinistros. Os MLG apresentam vantagens por serem uma generalização dos modelos lineares, entre as quais, a flexibilidade da função de regressão, ou seja da relação existente entre a variável resposta e a combinação linear das variáveis explicativas. Essa relação é habitualmente efetuada pela função de ligação. A sua maior vantagem consiste na possibilidade de construção de modelos que permitem obter intervalos de confiança das estimativas, com base em distribuições pertencentes à família exponencial, tal como pode ser verificado em Turkman & Silva (2000).

3.1 Definição de um Modelo Linear Generalizado

O estudo dos Modelos Lineares Generalizados iniciou-se há bastante tempo, no entanto a sua aplicação tem-se tornado maior nas últimas décadas. Tal como

apresentado por McCullagh & Nelder (1989), um Modelo Linear Generalizado terá que assentar sobre três componentes:

- **Componente aleatória:** as variáveis aleatórias y_i são i.i.d segundo uma distribuição família exponencial.
- **Componente sistemática / determinística:** combinação linear de p variáveis explicativas:

$$H = \mathbf{X} \cdot \mathbf{B} \quad (3.1.i)$$

$$\eta(i) = \beta_0 + \beta_1 x_1(i) + \beta_2 x_2(i) + \beta_3 x_3(i) + \dots + \beta_p x_p(i), \quad \forall i \in \{1, \dots, n\} \quad (3.1.ii)$$

- \mathbf{X} é a matriz das p variáveis explicativas
- \mathbf{B} representa o vetor de parâmetros, de dimensão p
- **Função de ligação:** A relação entre a componente aleatória e a componente sistemática é caracterizado por uma função monótona e diferenciável em todo o seu domínio, regularmente nomeada por função de ligação, g :

$$g(\mu) = g(E[Y]) = H, \quad (3.1.iii)$$

em que μ é o vetor dos valores esperados da componente aleatória, Y :

$$E[Y] = g^{-1}(H) = g^{-1}(\mathbf{X} \cdot \mathbf{B}) = g^{-1}(\sum_{j=1}^p \beta_j x_j(i)). \quad (3.1.iv)$$

Existem inúmeras funções que podem ser consideradas como função de ligação, e a decisão de qual função utilizar provém essencialmente do tipo de resposta que se pretende. De seguida, na Tabela II, é apresentada uma relação entre as distribuições mais comuns, da componente aleatória, com a função de ligação que frequentemente é utilizada.

Tabela II - Funções de ligação e distribuições mais comuns nos MLG

Distribuição	Função de Ligação (designação)	Função de Ligação	Função Variância
Normal/Gaussian	Identidade	μ	1
Binomial	Logit	$\text{Log}(\mu/(1-\mu))$	$\mu \cdot (1-\mu)/n$
Poisson	Logarítmica	$\text{Log}(\mu)$	μ
Gamma	Recíproca	$1/\mu$	μ^2
Normal inversa	Quadrática inversa	$1/\mu^2$	μ^3

Os Modelos Lineares Generalizados apresentam uma grande flexibilidade na sua aplicação a diferentes estudos estatísticos, sendo a sua única limitação a necessidade de se garantir que as variáveis explicativas entrem no modelo através de uma combinação linear, com a distribuição da variável dependente pertencente à família exponencial.

3.2 Família de dispersão exponencial

A família exponencial é uma importante classe de distribuições que partilham, entre si, certas características. Esta família inclui as distribuições mais utilizadas e comuns, entre as quais se encontram a Poisson, Gamma, Exponencial, Normal, etc... . A representação das distribuições pertencentes à família de dispersão exponencial é do tipo bi-paramétrico. A distribuição das variáveis aleatórias, y_i , é considerada pertencente à família de dispersão exponencial, se a função de densidade/probabilidade de Y , $f(Y|\theta, \phi)$, poder ser escrita sob a forma:

$$f(Y|\theta, \phi) = e^{\left[\frac{Y\theta - b(\theta)}{a(\phi)} + c(Y, \phi) \right]} \quad (3.2.i)$$

sendo θ o parâmetro usual da distribuição e ϕ o parâmetro de dispersão que normalmente é conhecido. As funções $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ são funções reais específicas de cada distribuição, onde se admite que $b(\cdot)$ é diferenciável. A função $a(\phi)$, depende apenas do parâmetro de dispersão e em muitos casos é escrita como o rácio $a(\phi) = \phi/w$, sendo w um valor constante conhecido dependendo da distribuição a considerar. O fator de dispersão da função de distribuição, ϕ , pode ser alterado e adequado consoante os dados que

temos disponíveis e a distribuição que desejamos utilizar. Naturalmente, se a este fator for atribuído um valor, estaremos perante uma família de distribuições exponencial simplificada. Sendo que, desta feita, a função $a(\phi)$, passa a ser uma constante.

De forma a poder encontrar uma expressão simplificada para o valor médio e para a variância, recorre-se à derivada do logaritmo da função densidade de probabilidade, $\ell(\theta; \phi, Y)$, habitualmente designada por função Score:

$$S(\theta) = \frac{\partial \ell(\theta; Y, \phi)}{\partial \theta} = \frac{Y - b'(\theta)}{a(\phi)}, \quad (3.2.ii)$$

da função score, torna-se intuitivo retirar a função de $E[Y]$,

$$\mu = E[Y] = a(\phi)E[S(\theta)] + b'(\theta). \quad (3.2.iii)$$

Sob certas condições de regularidade as seguintes igualdades se verificam:

$$E[S(\theta)] = 0 \quad \text{e} \quad E[S^2(\theta)] = -E\left[\frac{\partial^2 \ell(\theta; Y, \phi)}{\partial \theta^2}\right]$$

é então possível obter a função de μ e a variância de Y ,

$$\mu = b'(\theta), \quad (3.2.iv)$$

$$\text{var}(Y) = a^2(\phi)\text{var}(S(\theta)) = a^2(\phi)\frac{b''(\theta)}{a(\phi)} = a(\phi)b''(\theta). \quad (3.2.v)$$

Verificamos que a variância de Y consiste no produto das funções $a(\phi)$, que depende do parâmetro de dispersão, e a função $b''(\theta)$, que depende do parâmetro θ . Consequentemente, sabendo que a função de variância de μ é $b''(\theta)$, a variância dependerá do valor da média μ , o que pode ser facilmente verificado da seguinte forma:

$$\text{var}(Y) = a(\phi)\frac{\partial b'(\theta)}{\partial \theta} = a(\phi).v(\mu). \quad (3.2.vi)$$

Deste modo, confirma-se a relação entre a variância da variável aleatória tanto com o parâmetro de dispersão, ϕ , como com a função variância, $v(\mu)$, sendo que esta função de variância define a relação entre a variância de Y e a sua própria média.

4 Distribuição Tweedie

4.1 Características da distribuição Tweedie

A distribuição Tweedie pertence à família exponencial, e é um caso especial dentro dos modelos de dispersão exponencial. Esta distribuição é uma das grandes conquistas de Maurice Charles Kenneth Tweedie (1919 – 1996), que foi físico, médico, e estatístico da Universidade de *Liverpool*, e embora tenha feito uma carreira na física o seu grande interesse era a área da estatística e matemática.

Embora tenham sido estudados por Tweedie (1947), a aplicação dos Modelos Lineares Generalizados e as suas análises estatísticas estiveram décadas paradas, devido à falta de meios computacionais adequados. Jorgensen & Souza (1994) foram uns dos grandes responsáveis pela retoma do estudo e aplicação destes modelos de dispersão.

A opção por utilizar a distribuição Tweedie deve-se essencialmente ao facto de esta ser bastante flexível. Esta flexibilidade provém em grande parte da possibilidade de a Tweedie se aproximar de algumas das mais conhecidas distribuições da família exponencial. Permitindo assim, que possamos contar com uma melhor acomodação de “zeros”, na eventualidade de estes existirem nos dados.

A Tweedie tem como sua função de variância, $V(\mu) = \mu^p$, dependendo da constante $p \in (-\infty, 0] \cup [1, \infty)$, tal como verificado em Jorgensen (1997). A utilização deste resultado em **(3.2.vi)**, a variância da variável aleatória fica,

$$\text{Var}(Y) = a(\phi) \cdot \mu^p. \quad (4.1.i)$$

Conforme exposto em Jorgensen & Souza (1994), a cada valor atribuído à constante p , corresponde a uma distribuição específica. Para exemplos mais comuns e

conhecidos, quando p assume os valores: 0, 1, 2, 3 estamos perante as distribuições Guaussiana, Poisson, Gamma, e Guassiana inversa, respetivamente.

Quando a variável aleatória, y_i , respeita uma distribuição exponencial Tweedie, a função $b(\theta)$ é representada da seguinte forma, para p pertencente ao seu domínio e diferente de 1 e 2 ,

$$b(\theta_p) = \frac{1}{2-p} ((1-p)\theta_p)^{\frac{2-p}{1-p}}, \quad (4.1.ii)$$

para o caso de $p=1$, distribuição Poisson,

$$b(\theta) = e^\theta, \quad (4.1.iii)$$

e para $p=2$, distribuição Gamma,

$$b(\theta) = -\log(-\theta). \quad (4.1.iv)$$

Assim, pode-se dizer que o domínio do parâmetro θ fica condicionado pelo valor atribuído a p ,

$$D_\theta = \begin{cases} \mathbb{R}, & \text{para } p = 0, 1 \\ [0, \infty), & \text{para } p < 0 \\ (-\infty, 0[, & \text{para } 1 < p \leq 2 \\ (-\infty, 0], & \text{para } 2 < p < \infty \end{cases}$$

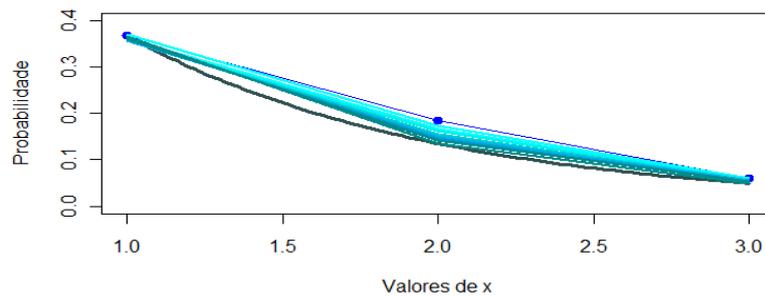
A função de distribuição Tweedie, tal como outra qualquer distribuição da família de dispersão exponencial, pode ser escrita na seguinte forma:

$$f(Y|\mu, \phi, p) = e^{c(Y, \phi, p)} \cdot e^{\frac{w}{\phi} [Y\theta_p - b(\theta_p)]}, \quad (4.1.v)$$

onde $c(Y, \phi, p)$ pode ser obtido na reparametrização feita em Wüthrich (2003).

Ao contrário de outras distribuições pertencentes à família exponencial, o comportamento da função densidade da distribuição Tweedie difere conforme o valor atribuído a p . No Gráfico 2, poderá ser verificado o comportamento da função densidade (4.1.v) perante a atribuição de diferentes valores a p entre 1 e 2 e o seu ajustamento às distribuições Poisson e Gamma.

Gráfico 2 - Ajustamento da Distribuição Tweedie à Poisson(1) e à Gama(1,1)



Tal como se verifica no Gráfico 2, a distribuição Tweedie começa por se ajustar bem à função de distribuição Poisson(1) para valores de p pequenos, e à medida que são escolhidos valores de p mais próximos do valor 2, acentua-se a sua curva ajustando-se a distribuição Gamma (1,1).

4.2 Casos particulares da distribuição Tweedie

A Tweedie é uma distribuição que possibilita o seu ajustamento a distribuições mais conhecidas da família exponencial, mediante o valor atribuído à constante p . Será, então, bastante curioso verificar alguns dos exemplos de valores a atribuir a p , que correspondam a algumas das distribuições mais famosas na área da estatística.

4.2.1 Distribuição Normal

Na distribuição Tweedie, estamos perante a distribuição normal quando o valor da constante p é igual a zero, possuindo as seguintes características,

$$c(y, \phi) = -\frac{1}{2} \left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right],$$

$$b(\theta) = \frac{\theta^2}{2},$$

de onde pode ser calculado μ ,

$$\mu(\theta) = \frac{\partial b(\theta)}{\partial \theta} = \theta. \quad (4.2.i)$$

Como $p=0$, temos então que

$$V_p(\mu) = \mu^p = \mu^0 = 1,$$

naturalmente que ao mesmo resultado chegaríamos, se aplicássemos a generalização da família exponencial:

$$V(\mu) = \frac{\partial b'(\theta)}{\partial \theta} = 1. \quad (4.2.ii)$$

4.2.2 Distribuição Gamma

A variável aleatória Y , tem distribuição Gama de parâmetros α e β , $Y \sim Ga(\alpha, \beta)$, quando é atribuído o valor dois à constante p . Verifica-se que,

$$c(y, \phi) = (\beta - 1) \ln(y) - \ln(\Gamma(\beta) + \beta \ln(\beta)), \quad a(\phi) = \frac{\phi}{w} = \frac{1/\beta}{1} = \frac{1}{\beta}$$

e com $b(\theta) = -\log(-\theta)$, temos que: $\frac{\partial b(\theta)}{\partial \theta} = -\frac{1}{\theta} \Rightarrow \mu(\theta) = -\frac{1}{\theta}$.

Sendo que a função de variância de μ pode ser obtida da seguinte forma,

$$V_p(\mu) = \frac{\partial b'(\theta)}{\partial \theta} = \mu^p = \mu^2. \quad (4.2.iii)$$

4.2.3 Distribuição Poisson

Na utilização dos MLG, com a distribuição Poisson, para o cálculo de estimativas para as provisões técnicas, é curioso verificar que quando utilizada uma função de ligação logarítmica, os resultados obtidos são iguais aos resultados que teriam sido obtidos com a estimação através do método determinístico, *Chain Ladder*. A variável aleatória com distribuição Poisson de média μ , $Po(\mu)$, é o equivalente à distribuição Tweedie com $p=1$.

As suas funções a , b e c são:

$$a(\phi) = 1, \quad b(\theta) = e^\theta, \quad c(y, \phi) = -\ln(y!).$$

Da forma habitual conseguimos obter a média e a função de variância,

$$\frac{\partial b(\theta)}{\partial \theta} = e^\theta = E(y) = \mu, \quad (4.2.iv)$$

$$\frac{\partial b'(\theta)}{\partial \theta} = e^\theta = \mu = \mu^1 = V_p(\mu). \quad (4.2.v)$$

4.3 Aplicação Prática 1

4.3.1 Análise preliminar dos dados

Para aplicação prática da investigação feita, são utilizados dados reais relativos a um ramo Não-Vida. No Anexo 2.i, é apresentada a matriz utilizada que agrega os montantes pagos a sinistros desse ramo. São considerados os sinistros ocorridos após o ano de 2004. Após a declaração da ocorrência dos sinistros à companhia de seguros, o comportamento dos pagamentos dos seus custos irá tendencialmente diminuir. Esta diminuição prolonga-se à medida que os anos passam após estes terem ocorrido, até à sua regularização. Consegue-se verificar esta situação de uma melhor forma, no Gráfico 3 e no Gráfico 4.

Gráfico 3 - Evolução dos pagamentos, em euros

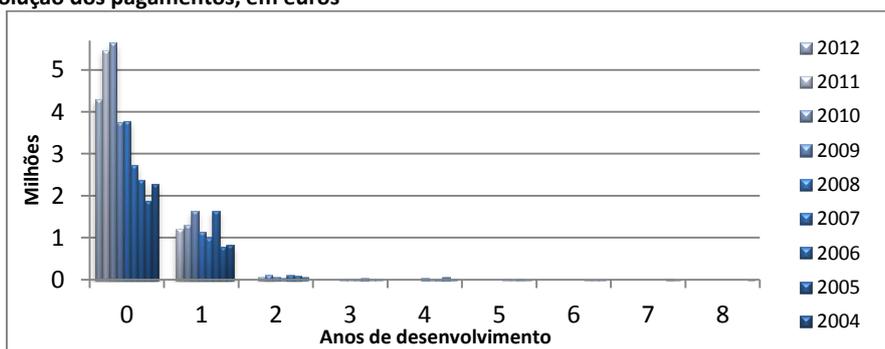
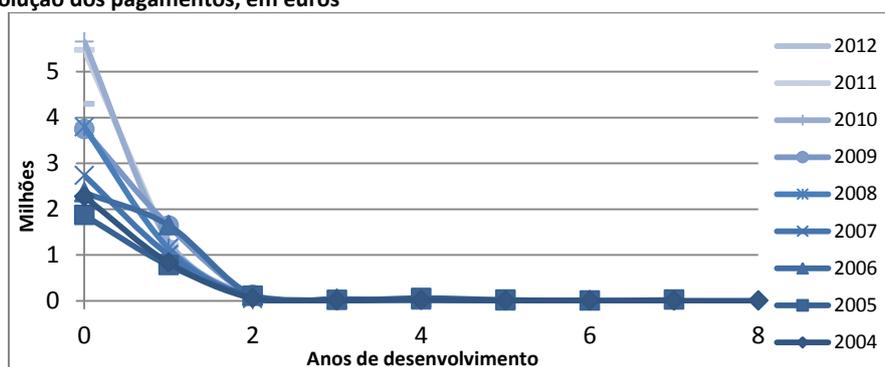


Gráfico 4 - Evolução dos pagamentos, em euros



Embora não seja o caso dos dados utilizados neste trabalho, existem ramos em que a regularização dos sinistros se torna um pouco mais demorada, devido ou à sua complexidade e gravidade ou a problemas “externos”, tais como a necessidade de

recolha de algumas informações contratuais ou legais e a demora na resolução de processos judiciais em tribunal. No entanto, verificamos que este ramo é de rápida resolução, isto porque se verifica que o grande volume de pagamentos é efetuado nos dois anos que se seguem à ocorrência do sinistro. Este fator é um aspeto importante para a análise, pois este acontecimento garante-nos uma maior estabilidade da matriz e assim reduzimos o risco de volatilidade dos resultados, o que nos permite uma melhor análise desta aplicação. Graficamente, este comportamento lembra-nos algumas distribuições à partida, tais como a Gamma e a Poisson, por exemplo. É por esse mesmo motivo que neste trabalho são essas as duas distribuições que temos como referência.

4.3.2 Testes de correlação

Tal como já referido, a variável aleatória são os montantes pagos que fazem face aos custos com sinistros. Sendo essa a nossa variável e assumindo como nossas variáveis explicativas os anos de ocorrência e os anos de desenvolvimento, faz sentido testarmos a existência de correlação entre as variáveis. Para efetuar esse teste, recorreu-se ao *software* R, e aplicou-se o teste de correlação de Pearson às variáveis, obtendo as seguintes conclusões.

- (Anexo 2.ii) Correlação entre montantes pagos e ano de ocorrência: com um $p\text{-value}$ = 0.0001579 e o coeficiente de Pearson igual a 0.5340426, estamos perante a evidência estatística de uma correlação positiva e significativa entre estas duas variáveis;
- (Anexo 2.iii) Correlação entre montantes pagos e o ano de desenvolvimento: com um $p\text{-value}$ = 3.092e-07 e o coeficiente de Pearson igual a -0.678099, estamos perante a evidência estatística de uma correlação negativa e significativa entre estas duas variáveis;

Tendo obtido estas conclusões, podemos estar um pouco mais certos nos bons resultados provenientes da utilização destas variáveis e na relação que é ser feita entre elas.

4.4 Estimar \hat{p} através de Máxima Verosimilhança

Habitualmente quando é utilizada a distribuição Tweedie é definido um p específico *a priori*. Naturalmente, tal como já referido, as características do tipo de dados utilizados indicam que a estimativa que iremos obter para este tipo de dados será um $p \in [1,2]$. Como esta distribuição pertence à família de dispersão exponencial, verifica-se também que esta constante, p , é a responsável na obtenção da função de variância de μ ,

$$V(\mu) = \mu^p, \quad p \in [1,2] \quad (4.4.i)$$

Foi então necessário tentar otimizar essa escolha. Smyth (1996) apresenta a utilização da estimativa por máxima verossimilhança, para obter uma otimização do valor de p . O valor estimado de p (estimativa de p , \hat{p}), que se considera mais adequado aos dados disponíveis consiste no valor de p que maximiza a função de Verosimilhança. Na aplicação prática presente neste trabalho foi obtido um valor de $\hat{p} = 1.62$, sendo que assim se poderá afirmar que a nossa estimativa se encontrará entre um modelo de Poisson, com variância de Y :

$$Var(Y) = a(\phi) \cdot \mu \quad (4.4.ii)$$

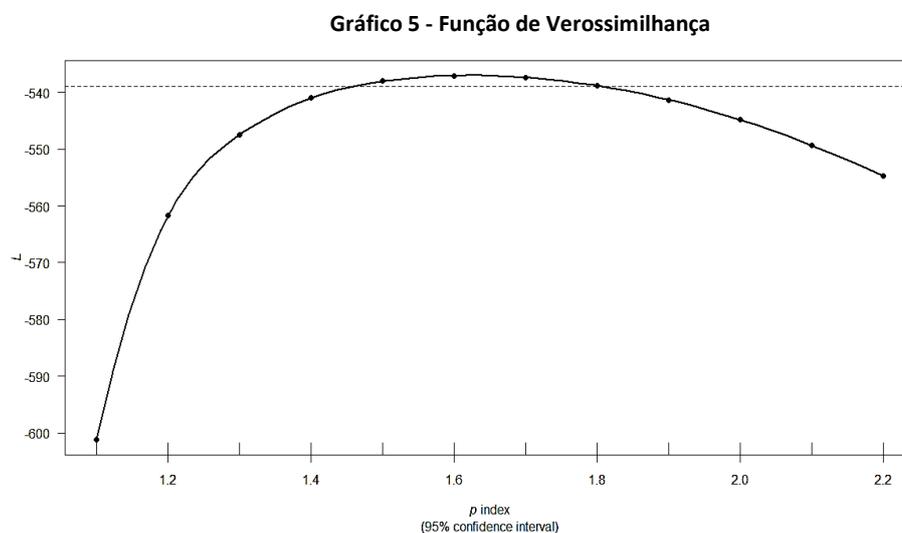
e um modelo de distribuição Gama:

$$Var(Y) = a(\phi) \cdot \mu^2. \quad (4.4.iii)$$

4.5 Aplicação Prática 2

4.5.1 Estimação da constante p

No exemplo apresentado nesta dissertação foi aplicada a estimação acima referida, obtendo-se assim uma estimativa para o valor de p . Utilizando o *package* “Tweedie” desenvolvido por Peter K Dunn no *software* R, é possível portanto obter o valor da constante $\hat{p}=1.62$, sendo o valor que maximiza a função de verossimilhança apresentada de seguida (Gráfico 5).

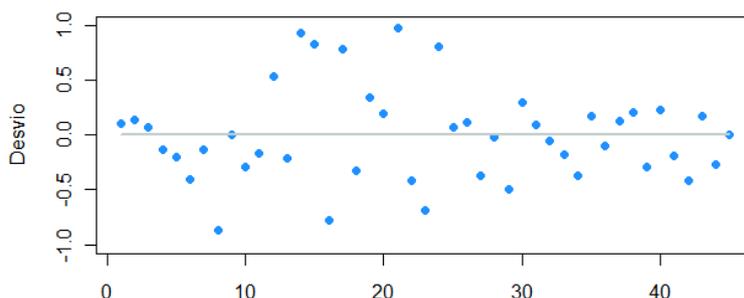
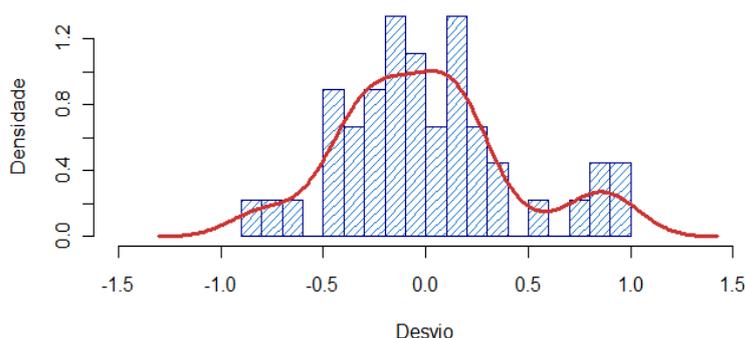


Para além desta estimativa, foi também obtido um intervalo de confiança a 95%,

$$IC_{\hat{p}} = [1.457969, 1.805540] ,$$

pelo que o valor da constante se encontra compreendido dentro deste intervalo, com um nível de confiança de 95%.

Relativamente aos resíduos resultantes desta estimação por Máxima Verossimilhança, podem ser analisados a partir do Gráfico 6, Gráfico 7, Anexo 3.i e Anexo 3.ii. Nos gráficos em anexo, é possível verificar que o desvio dos resíduos aumenta com o passar do desenvolvimento dos sinistros (embora esta tendência não seja muito acentuada) e diminui quando nos aproximamos do presente.

Gráfico 6 - Estimação de p - Dispersão dos resíduosGráfico 7 - Estimação de p - Histograma dos resíduos

Pelo que se verifica no Gráfico 6 e no Gráfico 7, há indícios dos resíduos se apresentarem sob a forma de uma distribuição Normal(0,1). No entanto, para se poder testar se é verificada a hipótese de normalidade dos resíduos, procedeu-se à realização do teste de normalidade Jarque-Bera. Este teste tem em consideração o nível de *kurtosis* e *skewness* (assimetria), testando como hipótese “nula” a existência de normalidade dos resíduos.

Tabela III - Estimação de p - Jarque Bera Test aos resíduos

Jarque Bera Test		
X-squared	df	p-value
1.8646	2	0.3936

Os resultados obtidos no teste Jarque-Bera vêm confirmar os indícios anteriores. Os resíduos resultantes da obtenção de um estimador para p , apresentam um comportamento similar a distribuição Normal. Com um $p\text{-value}=0.3936$, não foi rejeitada, estatisticamente, a hipótese de existência de normalidade.

5 Estimação dos parâmetros

5.1 Estimação de parâmetros dos MLG

A vantagem que mais se destaca nos Modelos Lineares Generalizados é a estimação dos seus parâmetros. Essa estimação é efetuada a partir do método de Máxima Verossimilhança. O seu estimador apresenta além da consistência e normalidade assintótica também a propriedade ótima de eficiência. Este método tem como objetivo a obtenção de estimativas para o parâmetro de interesse, θ . É também importante referir que o parâmetro de dispersão, ϕ , no caso de não assumir um valor conhecido, é estimado conjuntamente com os outros parâmetros.

Admitindo daqui em diante, o cumprimento das condições de regularidade de Sen & Singer (1993), a estimação da Máxima Verossimilhança consiste na obtenção dos máximos da função de verossimilhança, correspondente aos valores de y_i verificados.

A função de verossimilhança das n observações, independentes y_i com $i=1,2,\dots,n$ é:

$$L(\theta, \phi; y_1, y_2, \dots, y_n) = \prod_{i=1}^n f(y_i|\theta, \phi). \quad (5.1.i)$$

Recordando que θ é escrito em função de μ , podemos escrever o logaritmo da função de verossimilhança da seguinte forma:

$$l(\theta) = \ln(L(\theta, \phi; y_1, y_2, \dots, y_n)) = \sum_{i=1}^n \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i) \right]. \quad (5.1.ii)$$

Desta forma os estimadores para β presentes na equação (3.1.ii), podem ser obtidos mediante a resolução do sistema de equações seguinte,

$$\frac{\partial l(\beta)}{\partial \beta_j} = 0 \Leftrightarrow \sum_{i=1}^n \frac{\partial l(\beta_j, y_i)}{\partial \beta_j} = 0, \quad j = 1, 2, \dots, p. \quad (5.1.iii)$$

Para maior aprofundamento deste desenvolvimento, podemos consultar McCullagh & Nelder (1989).

5.2 Aplicação Prática 3

5.2.1 Estimar modelo com $p = \hat{p}$

Retomando os dados e o exemplo de que dispomos, é aplicado o método de estimação em estudo. Conhecendo agora o valor de p a utilizar, $\hat{p}=1.62$, é possível utilizar a distribuição Tweedie nos Modelos Lineares Generalizados. Para esta aplicação recorreu-se a uma função de ligação logarítmica (Capítulo 3.1), que entre todas é a que habitualmente é utilizada. Como variáveis explicativas, foram utilizadas variáveis *dummys* correspondentes a cada um dos anos de desenvolvimento, assim também como a cada um dos anos de ocorrência. Desta forma, é possível ter um modelo que apresenta resultados “passo a passo”, ou seja, este modelo apresenta estimativas de coeficientes para cada um dos anos desenvolvimento e não apenas um coeficiente que multiplica o número de anos, por exemplo. A variável explicada, que se assume comportar-se segundo a distribuição Tweedie com $\hat{p}=1.62$, é a que representa os montantes pagos com sinistros no passado (como tal já tinha sido referido, anteriormente). O modelo a estimar é o seguinte:

$$E(\text{montantes pagos}) = \exp\left\{\alpha + \sum_{i=2005}^{2012} \beta_i \cdot \theta_i + \sum_{j=1}^8 \gamma_j \cdot \delta_j\right\}, \quad (5.2.i)$$

com

$$\begin{cases} \theta_i = 1, & i = \text{Ano ocorrência considerado} \\ \theta_i = 0, & i \neq \text{Ano ocorrência considerado} \end{cases}$$

e

$$\begin{cases} \delta_j = 1, & j = \text{Ano desenvolvimento considerado} \\ \delta_j = 0, & j \neq \text{Ano desenvolvimento considerado} \end{cases}$$

Para este modelo poder ser utilizado, só é necessário estimar os coeficientes β_i e γ_j . As estimativas dos coeficientes são obtidas tal como foi evidenciado no ponto 5.1. No entanto, na aplicação prática desse processo recorreu-se a *softwares* que facilitam e

aceleram este processo. O *software* a ser utilizado é o *R-project*, com o *package* específico da Tweedie, desenvolvido por Gordon Smyth com base em: Tweedie (1984); Jorgensen (1987); Smyth (1996); Jorgensen (1997); Smyth & Verbyla (1999). O resultado da sua estimação encontra-se em anexo (Anexo 4.i). Após a estimação, pode ser feita uma rápida avaliação da estimativa. Um dos fatores que mais se destaca é a significância das variáveis às quais estimamos os seus coeficientes. Todas são significativas a pelo menos a 5%, à exceção dos fatores relativos aos anos de ocorrência de 2005 e 2007.

Como já são conhecidos os valores das estimativas dos coeficientes, utilizando a equação (5.2.i), podemos estimar os montantes que poderão vir a ser pagos, de sinistros já ocorridos. A estimativa obtida por este modelo para o “triângulo inferior” da matriz foi, cerca de 2.243.687€. Este valor, reunido aos 39.464.525€, que representavam os pagamentos já efetuados, perfaz uma estimativa de custo com sinistros, ocorridos entre 2005 e 2012, de 41.708.212€.

6 Adequabilidade do modelo

O facto de o modelo que estamos a considerar ser ou não adequado aos dados de que dispomos reflete grande parte da credibilidade que lhe é atribuída aos resultados obtidos, pelas suas estimativas. A avaliação de um modelo pode ser feita por diversas formas, dependendo do critério que se opta por utilizar. Esta análise é, por vezes, erradamente, efetuada apenas com base nos valores resultantes das estimativas. É um erro a análise ser feita, por vezes, dessa forma, pois é completamente ignorada a qualidade de ajustamento do modelo aos dados de que dispomos. Uma estimativa que aparenta ser boa pode vir a revelar-se um fracasso, assim como o contrário. Não sendo este o modo para a análise do modelo, é apresentado de seguida o meio mais eficaz e conveniente de se poder efetuar essa análise.

6.1 Resíduos

Uma vez estimado, é possível obter os resíduos do modelo. Os resíduos provêm das diferenças verificadas entre os dados de que dispomos e os dados obtidos pelo modelo, sendo ε o vetor dos resíduos, e que o modelo linear normal se pode escrever da seguinte forma,

$$Y = \hat{\mu} + \varepsilon, \quad (6.1.i)$$

verifica-se com facilidade que,

$$\varepsilon = Y - \hat{\mu}. \quad (6.1.ii)$$

No entanto, nos Modelos Lineares Generalizados não existe a componente ε , para o qual seja possível obter uma estimativa. Têm surgido e normalmente são utilizadas algumas propostas para o cálculo desses desvios de estimativa. Embora tenha como

grande desvantagem a sua assimetria e enviesamento (para modelos não normais), um dos métodos mais conhecidos e utilizados são os resíduos de Pearson:

$$R_i^p = \frac{Y_i - \hat{\mu}_i}{\sqrt{\text{var}(\hat{\mu}_i)}}, \quad \forall i \in \{1, \dots, n\} \quad (6.1.iii)$$

de forma a poder verificar o impacto, de cada observação do modelo, no cálculo da estimativa de Pearson generalizada. É utilizada a seguinte aproximação, $\text{Var}(Y_i - \hat{\mu}_i) \approx \text{Var}(Y_i) \cdot (1 - h_i)$, com h_i sendo a diagonal principal da matriz $H = X(X^T W X)^{-1} X^T W$, para se obter a equação estandardizada dos resíduos de Pearson, R_i^{*p} :

$$R_i^{*p} = \frac{(Y_i - \hat{\mu}_i) \cdot w_i}{\sqrt{\hat{\phi} \cdot \text{var}(\hat{\mu}_i)}} \approx \frac{(Y_i - \hat{\mu}_i) \cdot w_i}{\sqrt{\hat{\phi} \cdot \text{var}(\hat{\mu}_i) \cdot (1 - h_i)}}, \quad \forall i \in \{1, \dots, n\}. \quad (6.1.iv)$$

Uma outra abordagem ao cálculo dos resíduos é a obtenção de resíduos com base na função *deviance* (desvio). Este tipo de resíduos é definido na seguinte forma:

$$R_i^d = \zeta_i \cdot \sqrt{D_i}, \quad \forall i \in \{1, \dots, n\} \quad (6.1.v)$$

com $\zeta = \text{sinal}(y_i - \hat{\mu}_i)$, ou seja, $\zeta \in \{-1, 1\}$, conforme o sinal verificado na diferença entre o valor real e o valor estimado, e D_i a respetiva função de desvio para o modelo (Tabela IV). Um pouco à semelhança do que é feito nos resíduos de Pearson (6.1.iv), neste tipo de resíduos é também possível obter a função estandardizada através da

multiplicação de R_i^d por, $1/\sqrt{\hat{\phi} \cdot (1 - h_i)}$:

$$R_i^{*d} = \frac{\zeta \cdot \sqrt{D_i}}{\sqrt{\hat{\phi} \cdot (1 - h_i)}}, \quad \forall i \in \{1, \dots, n\}. \quad (6.1.vi)$$

6.1.1 Casos particulares de resíduos

Para diferentes modelos existem diferentes funções de cálculo dos seus desvios, ou resíduos. Estas diferenças advêm essencialmente das várias funções de distribuição

aplicadas. Por esse mesmo motivo, na Tabela IV são apresentados as diferentes funções de desvio correspondentes às distribuições mais conhecidas e comuns.

Tabela IV - Funções desvio para casos particulares

Distribuição	Função desvio
Poisson	$2 \cdot \left[\sum_i^n Y_i \cdot \ln \left(\frac{Y_i}{\hat{\mu}_i} \right) - \sum_i^n (Y_i - \hat{\mu}_i) \right]$
Gama	$2 \cdot \sum_i^n - \ln \left(\frac{Y_i}{\hat{\mu}_i} \right) + \frac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$
Normal	$\sum_i^n (Y_i - \hat{\mu}_i)^2$
Normal inversa	$\sum_i^n \frac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2 \cdot Y_i}$

6.1.1.1 Resíduos do modelo Normal

Este modelo torna-se relativamente simples na obtenção dos seus resíduos. O modelo normal apresenta a sua função de variância igual a um. Quer os desvios de Pearson, quer os desvios obtidos com base na função *deviance* (desvio) são:

$$R_i^d = R_i^p = (Y_i - \hat{\mu}_i).$$

6.1.1.2 Resíduos do modelo Poisson

Relativamente aos modelos com a distribuição Poisson, sabendo que a sua variância é, $Var(x)=x$, através de (6.1.iii) os resíduos de Pearson são obtidos pela função:

$$R_i^p = \frac{Y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}.$$

No que diz respeito aos resíduos R_i^d , com base no resultado apresentado na Tabela IV, podem ser obtidos através da seguinte função:

$$R_i^d = \zeta_i \cdot \sqrt{2 \cdot \left(Y_i \cdot \ln \left(\frac{Y_i}{\hat{\mu}_i} \right) - Y_i + \hat{\mu}_i \right)}.$$

6.2 Análise dos resíduos

A análise dos resíduos permite identificar a qualidade das escolhas da função de distribuição da variável aleatória e da função de ligação, assim como também pode permitir identificar observações que não sejam bem explicadas pelo modelo. Para

proceder ao estudo sobre a qualidade de ajustamento do modelo, é sempre aconselhável fazer essa análise com suporte nos resíduos baseados na função *deviance*, em detrimento dos resíduos de Pearson. É desejável que estes resíduos se distribuam de forma irregular e que estejam próximos de zero. Ou seja, é pretendido que os resíduos apresentem uma variância constante, unitária e que apresentem uma simetria na sua dispersão, o que revelaria que as estimativas do modelo se aproximam muito dos valores reais. Esta expectativa advém da teoria de que o valor esperado dos resíduos é igual a zero, e que se admite a existência da independência entre si. Para poder verificar e analisar o seu comportamento padrão, é necessária a construção de alguns gráficos dos resíduos e efetuar alguns testes que apoiem e enriqueçam a análise, tais como o teste de Jarque-Bera.

Visa-se sempre obter resíduos de um modelo que se aproximem de zero e que estejam distribuídos aleatoriamente, isto é, que respeitem uma distribuição Normal com média zero. No caso de ser utilizada a versão padronizada dos resíduos, essa expectativa assenta no pressuposto de estes apresentarem um comportamento semelhante à função de distribuição Normal estandardizado. Um possível teste para efetuar uma verificação, *a priori*, dessa semelhança, consiste em verificar se, no mínimo, 95% dos resíduos se encontram compreendidos dentro do intervalo $]-2, 2[$, e se existe simetria dos resíduos em torno de zero.

Para melhor verificar a repartição dos resíduos, é necessária a construção de um histograma com os valores dos resíduos obtidos. Desta forma, torna-se mais intuitiva a sua distribuição, o que permite verificar quer o desvio, quer a simetria em torno de zero.

6.3 Aplicação prática 4

Para poder averiguar o bom ajustamento do modelo aos dados, será necessário proceder à análise dos resíduos resultantes do modelo estimado no Anexo 4.i. . Para uma comparação viável deste modelo com outras possíveis propostas, são também apresentados dois modelos alternativos ao proposto neste trabalho: o modelo Poisson (Anexo 4.ii) e o modelo Gama (Anexo 4.iii). Para esta análise utilizaremos como suporte os resíduos baseados na função *deviance* e a sua padronização.

6.3.1 Adequabilidade do modelo com $p = \hat{p}$

Recorremos a alguns gráficos para efetuar a análise de adequabilidade do modelo estimado, em que a sua variável aleatória assume uma distribuição Tweedie, com o valor da constante p igual à sua estimativa, \hat{p} , calculado no ponto 4.5.1. . São os gráficos Gráfico 8, Gráfico 9, Anexo 5.i e Anexo 5.ii que nos podem dar a conhecer alguns dos moldes como os resíduos se encontram distribuídos.

Gráfico 8 - Modelo com p estimado - Dispersão dos resíduos padronizados

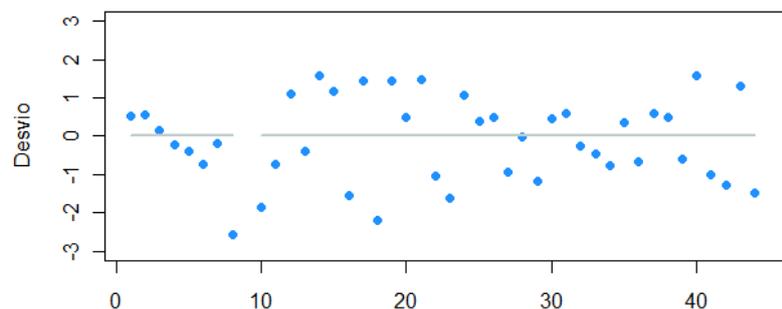
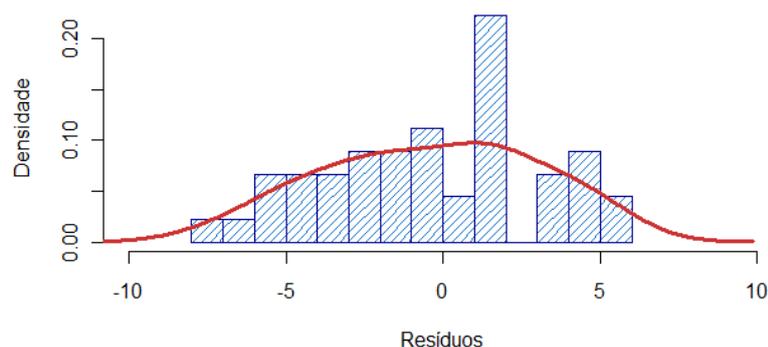


Gráfico 9 - Modelo com p estimado - Histograma dos resíduos



Com a análise do Gráfico 8 e do Gráfico 9 podemos verificar que apresentam indícios de o modelo, que representam, estar bem ajustado aos dados. Os resíduos padronizados encontram-se, na sua maioria, compreendidos entre o intervalo $]-2, 2[$, com uma dispersão aleatória e simétrica em relação a zero. Para testar a sua normalidade e eliminar algumas dúvidas existentes, procede-se à elaboração do teste Jarque-Bera, que tem como hipótese a testar: a existência de normalidade na distribuição dos resíduos.

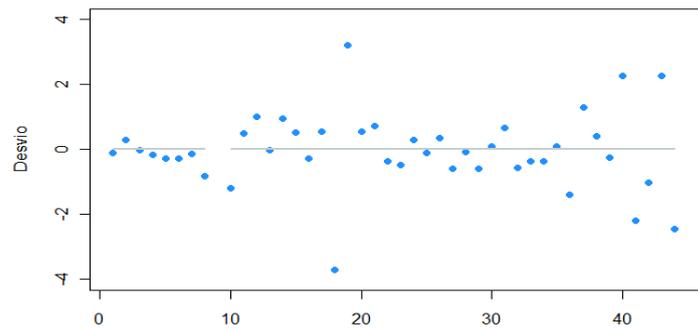
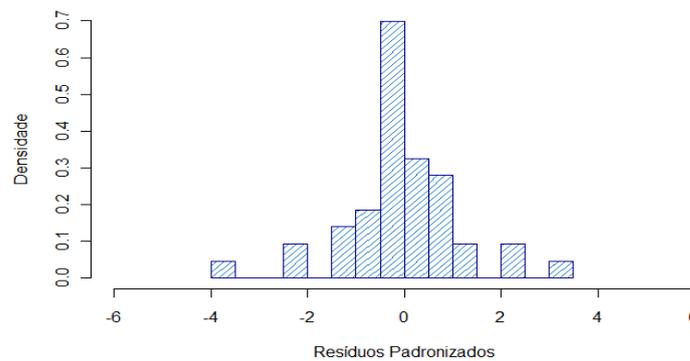
Tabela V - Modelo com p estimado - Jarque Bera Test aos resíduos

Jarque Bera Test		
X-squared	df	p-value
1.5663	2	0.457

Dados os resultados apresentados na Tabela V é possível admitir que os resíduos do modelo estimado com a constante $p = \hat{p}$, apresentam uma distribuição Normal, pois essa hipótese não é rejeitada com um $p\text{-value}=0,457$. Com a análise ao Anexo 5.i e ao Anexo 5.ii verificamos que quer em relação ao ano de desenvolvimento, quer ao ano de ocorrência dos sinistros, os resíduos Padronizados não aparentam ter um padrão constante. Relativamente ao comportamento dos resíduos em relação aos valores modelizados (Anexo 5.iii), também não apresentam uma tendência crescente ou decrescente, o que revela que a distribuição utilizada para a variável resposta é adequada.

6.3.2 Adequabilidade do modelo com $p=1$, Poisson

Para analisar a adequabilidade do modelo com $p=1$, procede-se também à observação dos gráficos dos resíduos resultantes da estimação do modelo. São os gráficos: Gráfico 10, o Gráfico 11, o Anexo 5.iv e o Anexo 5.v, que nos podem dar a conhecer alguns dos moldes como os resíduos se encontram distribuídos.

Gráfico 10 - Modelo com $p=1$ - Dispersão dos resíduos padronizadosGráfico 11 - Modelo com $p=1$ - Histograma dos resíduos padronizados

Com a análise do Gráfico 10 e do Gráfico 11 verifica-se que os resíduos apresentam uma dispersão simétrica em relação a zero, porém existem muitos *outliers* e consequentemente um grande número de resíduos padronizados fora do intervalo $]-2,2[$, o que apresenta indícios de estes resíduos não seguirem uma distribuição Normal. Para testar a sua normalidade e eliminar algumas dúvidas existentes procede-se à elaboração do teste Jarque-Bera, que tem como hipótese a testar: a existência de normalidade na distribuição dos resíduos.

Tabela VI - Modelo com $p=1$ - Jarque Bera Test aos resíduos

Jarque Bera Test		
X-squared	df	p-value
11.776	2	0.002773

Dados os resultados apresentados na Tabela VI verificamos que não existe evidência estatística de os resíduos do modelo, estimado com $p=1$, seguirem uma

distribuição Normal. É rejeitada a hipótese de existência de normalidade dos resíduos com um $p\text{-value} = 0,002773$. Com a análise ao Anexo 5.iv e ao Anexo 5.v verifica-se que quer em relação ao ano de desenvolvimento, quer ao ano de ocorrência dos sinistros, os resíduos Padronizados apresentam padrões sistemáticos. No caso dos resíduos em relação ao ano de desenvolvimento é verificável uma tendência decrescente, e uma tendência crescente na relação com os anos de ocorrência. Relativamente ao comportamento dos resíduos em relação aos valores modelizados (Anexo 5.vi), apresentam também uma tendência crescente. Todas estas conclusões trazem fortes indícios de que este modelo, com a distribuição Tweedie com $p=1$, não evidencia um bom ajustamento aos dados utilizados.

6.3.3 Adequabilidade do modelo com $p=2$, Gamma

A análise da adequabilidade do modelo estimado com a distribuição Tweedie com $p=2$ (Anexo 4.iii) é efetuada também com recurso aos gráficos e testes realizados aos resíduos. É com base no Gráfico 12, no Gráfico 13, no Anexo 5.vii e no Anexo 5.viii que podemos tomar conhecimento sobre alguns dos moldes, como os resíduos se encontram distribuídos.

Gráfico 12- Modelo com $p=2$ - Dispersão dos resíduos padronizados

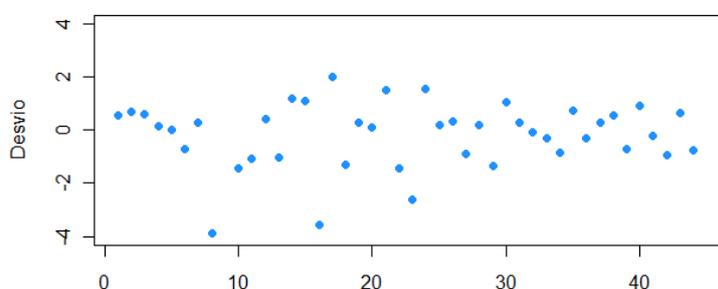
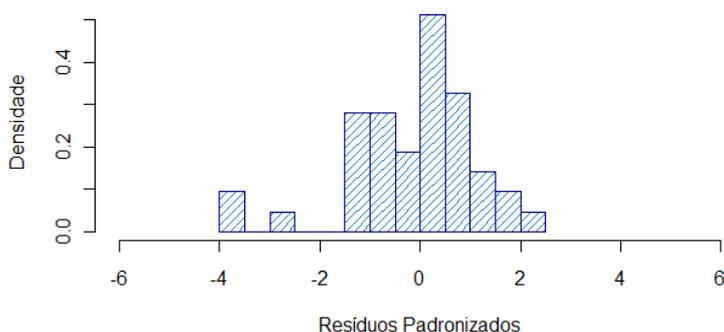


Gráfico 13 - Modelo com $p = 2$ - Histograma dos resíduos padronizados

O Gráfico 12 e o Gráfico 13 são suscetíveis de algumas dúvidas em relação à normalidade dos resíduos. Embora apresentem uma dispersão aleatória e simétrica em relação a zero, existem algumas dúvidas em relação à evidência de a generalidade dos resíduos estar compreendida entre o intervalo $[-2,2]$ ou não estar. Para poder apoiar e esclarecer esta situação é aplicado o teste Jarque-Bera, para testar a veracidade da hipótese de existência de normalidade dos resíduos.

Tabela VII - Modelo com $p = 2$ - Jarque Bera Test aos resíduos

Jarque Bera Test		
X-squared	df	p-value
8.0567	2	0.0178

Dados os resultados apresentados na Tabela VII verificamos que não existe evidência estatística de os resíduos, do modelo estimado com $p=2$, seguirem uma distribuição Normal. É rejeitada a 5% com um $p\text{-value} = 0,0178$ a hipótese que admite a existência de normalidade dos resíduos. A relação dos resíduos com os valores modelizados (Anexo 5.ix) apresenta-se um pouco constante. No entanto, no que diz respeito aos gráficos que relacionam os resíduos com os anos de desenvolvimento e de ocorrência (Anexo 5.vii e Anexo 5.viii) é possível verificar que apresentam tendências decrescentes e crescentes, respetivamente, evidenciando o fraco ajustamento do modelo aos dados.

6.3.4 Resultados obtidos

Com o apoio do *software R-project* é possível estimar os coeficientes dos modelos apresentados no Anexo 4.i, no Anexo 4.ii e no Anexo 4.iii. As análises de adequabilidade apuradas nos subcapítulos anteriores indicam que, estatisticamente, o modelo que melhor se adequa aos dados que dispomos para esta aplicação prática é o modelo que utiliza a distribuição Tweedie, com o valor da constante p estimada. Esta evidência contribui para a forma e maneira como observamos os valores estimados das provisões técnicas através destes três modelos, isto é, a credibilidade que devemos depositar em cada um dos modelos.

Tabela VIII - Provisão para sinistros, estimada por 3 modelos diferentes e por *Chain Ladder*

	Estimativa MLG ($p=1,62$)	Estimativa MLG ($p=2$)	Estimativa MLG ($p=1$)	Estimativa <i>Chain Ladder</i>
2005	184	285	128	128
2006	14.620	13.134	14.521	14.521
2007	13.929	10.598	17.321	17.320
2008	32.448	26.269	38.120	38.118
2009	85.475	75.781	88.439	88.441
2010	121.252	94.607	143.697	143.694
2011	244.131	214.335	267.044	267.042
2012	1.731.648	1.689.690	1.701.419	1.701.277
Total	2.243.687	2.124.698	2.270.689	2.270.541

Através do exposto na Tabela VIII é verificada, imediatamente, a aproximação muito forte do modelo Tweedie com $p=1$ ao método de estimação de *Chain Ladder*. Um acontecimento que é justificado pela igualdade existente entre os resultados obtidos pelo método *Chain Ladder* e os do modelo linear generalizado com a sua componente aleatória seguindo uma distribuição Poisson.

Relativamente a estes dados e comparativamente com os resultados obtidos através do método *Chain Ladder* podemos verificar que a estimativa da provisão, obtida através do modelo Tweedie com \hat{p} , é sensivelmente 1,18% inferior, e a do modelo Tweedie, com $p=2$, menor em cerca de 6,42%.

7 Conclusões

Os resultados obtidos com os modelos estimados mostram-se muito próximos, mas não podemos generalizar esta conclusão para todas as estimativas realizadas com estes modelos. Isto verifica-se apenas em situações em que os dados utilizados correspondem a um ramo de sinistros de rápida regularização e com montantes não muito elevados, como neste caso prático.

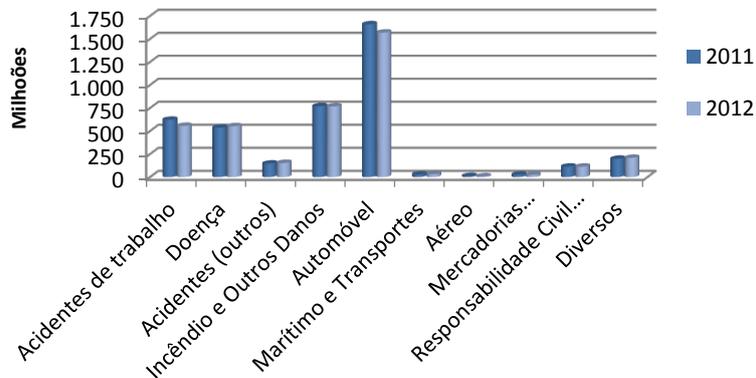
Na análise do comportamento dos resíduos é facilmente perceptível que o modelo Tweedie, com a constante p estimada, é tão bom ou melhor que outros modelos estimados e uma boa alternativa aos métodos determinísticos e estocásticos utilizados para o mesmo fim. Sendo ainda possível a obtenção do erro padrão da estimativa com os Modelos Lineares Generalizados, que podem permitir obter intervalos em que, com alguma confiança, o verdadeiro volume de pagamentos futuros se encontra.

Como futuros desenvolvimentos e aplicações deste modelo, pode ser estudado o cálculo da reserva para os sinistros ocorridos em anos anteriores aos da matriz *Run Off* considerada, normalmente designada por “cauda”.

Anexos

Anexo 1. Mercado Segurador

Anexo 1.i – Produção de seguro direto em Portugal, ramos Não-Vida



Anexo 2. Dados

Anexo 2.i - Matriz exemplo de pagamentos de sinistros, em euros

Ano Ocorrência	0	1	2	3	4	5	6	7	8
2004	2.275.970	830.626	57.965	10.428	18.679	4.842	2.237	1.115	144
2005	1.865.752	773.779	105.970	12.060	57.650	19.238	717	19.498	
2006	2.362.955	1.648.809	109.523	40.120	22.802	4.370	7.926		
2007	2.735.413	1.005.084	42.240	14.499	14.419	13.236			
2008	3.790.467	1.147.154	74.029	12.616	45.725				
2009	3.739.513	1.640.180	130.119	17.107					
2010	5.652.611	1.300.053	69.823						
2011	5.471.076	1.201.820							
2012	4.296.178								

Anexo 2.ii - (Output R) Teste de correlação de Pearson, entre montantes pagos e ano de ocorrência

Pearson's product-moment correlation

```
data: dat_alt[, 1] and dat_alt[, 2]
t = 4.1421, df = 43, p-value = 0.0001579
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2852196 0.7154265
sample estimates:
      cor
0.5340426
```

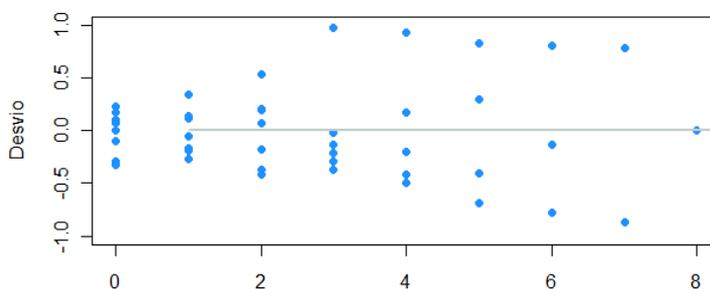
Anexo 2.iii - (Output R) Teste de correlação de Pearson, entre montantes pagos e o ano de desenvolvimento

Pearson's product-moment correlation

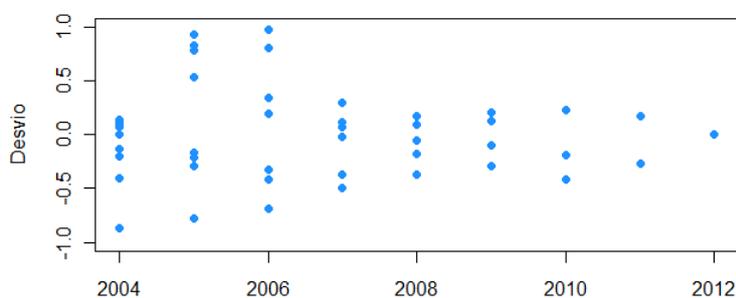
```
data: dat_alt[, 1] and dat_alt[, 3]
t = -6.05, df = 43, p-value = 3.092e-07
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.8103390 -0.4801333
sample estimates:
      cor
-0.678099
```

Anexo 3. Estimação de \hat{p}

Anexo 3.i - Estimação de p - Dispersão dos resíduos vs. Ano desenvolvimento



Anexo 3.ii - Estimação de p - Dispersão dos resíduos vs. Ano de ocorrência



Anexo 4. Estimação dos modelos

Anexo 4.i - (Output R) Modelo com o estimador de p

```
Call:
glm(formula = y ~ factor(year) + factor(dj), family = tweedie(var.power = p$p.max,
link.power = 0), data = dat_alt)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-7.5255	-2.7655	-0.1374	1.7991	5.5901

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	14.5417	0.1750	83.089	< 2e-16 ***
factor(year)2005	0.2473	0.2201	1.124	0.27072
factor(year)2006	0.5249	0.2185	2.402	0.02316 *
factor(year)2007	0.2150	0.2271	0.947	0.35176
factor(year)2008	0.5130	0.2244	2.286	0.03002 *
factor(year)2009	0.6953	0.2261	3.075	0.00467 **
factor(year)2010	0.7979	0.2285	3.492	0.00161 **
factor(year)2011	0.8112	0.2374	3.417	0.00195 **
factor(year)2012	0.7316	0.2820	2.595	0.01490 *
factor(dj)1	-1.0481	0.1293	-8.106	7.97e-09 ***
factor(dj)2	-3.6472	0.1975	-18.470	< 2e-16 ***
factor(dj)3	-5.1530	0.2749	-18.745	< 2e-16 ***
factor(dj)4	-4.4889	0.2713	-16.548	5.47e-16 ***
factor(dj)5	-5.5307	0.3649	-15.157	5.03e-15 ***
factor(dj)6	-6.6843	0.5155	-12.967	2.34e-13 ***
factor(dj)7	-5.4932	0.5205	-10.553	2.89e-11 ***
factor(dj)8	-9.5718	1.6052	-5.963	2.02e-06 ***

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

(Dispersion parameter for Tweedie family taken to be 17.13837)

Null deviance: 24836.83 on 44 degrees of freedom
Residual deviance: 515.53 on 28 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 6

Anexo 4.ii – (Output R) Modelo com $p=1$

```
glm(formula = y ~ factor(year) + factor(dj), family = tweedie(var.power = 1,
link.power = 0), data = dat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-404.63  -69.17  -10.19   79.61  557.59

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.6457    0.1205  121.577 < 2e-16 ***
factor(year)2005  -0.1148    0.1724  -0.666  0.510975
factor(year)2006   0.2739    0.1572   1.742  0.092433 .
factor(year)2007   0.1823    0.1605   1.136  0.265750
factor(year)2008   0.4671    0.1513   3.087  0.004526 **
factor(year)2009   0.5617    0.1489   3.772  0.000772 ***
factor(year)2010   0.8056    0.1430   5.632  4.95e-06 ***
factor(year)2011   0.7735    0.1443   5.360  1.04e-05 ***
factor(year)2012   0.6276    0.1580   3.973  0.000452 ***
factor(dj)1      -1.0721    0.0794 -13.502  8.78e-14 ***
factor(dj)2      -3.6603    0.2792 -13.112  1.79e-13 ***
factor(dj)3      -5.1151    0.6497  -7.873  1.41e-08 ***
factor(dj)4      -4.4594    0.5333  -8.362  4.26e-09 ***
factor(dj)5      -5.4909    1.0390  -5.285  1.27e-05 ***
factor(dj)6      -6.5162    2.0315  -3.208  0.003340 **
factor(dj)7      -5.3494    1.4777  -3.620  0.001152 **
factor(dj)8      -9.6759    17.6482  -0.548  0.587856
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Tweedie family taken to be 44847.89)

Null deviance: 89485760 on 44 degrees of freedom
Residual deviance: 1213257 on 28 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

Anexo 4.iii – (Output R) Modelo com $p=2$

```
glm(formula = y ~ factor(year) + factor(dj), family = tweedie(var.power = 2,
link.power = 0), data = dat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.33341  -0.33943   0.03948   0.22414   0.64916

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.4175    0.2495  57.795 < 2e-16 ***
factor(year)2005  0.6827    0.2444   2.793  0.009309 **
factor(year)2006  0.8476    0.2572   3.295  0.002671 **
factor(year)2007  0.3292    0.2715   1.212  0.235461
factor(year)2008  0.6212    0.2893   2.147  0.040583 *
factor(year)2009  0.8430    0.3131   2.693  0.011836 *
factor(year)2010  0.8042    0.3482   2.309  0.028514 *
factor(year)2011  0.8965    0.4084   2.195  0.036602 *
factor(year)2012  0.8557    0.5488   1.559  0.130200
factor(dj)1      -1.0630    0.2444  -4.349  0.000164 ***
factor(dj)2      -3.7005    0.2572 -14.388  1.84e-14 ***
factor(dj)3      -5.2362    0.2715 -19.284 < 2e-16 ***
factor(dj)4      -4.5918    0.2893 -15.872  1.57e-15 ***
factor(dj)5      -5.6396    0.3131 -18.012 < 2e-16 ***
factor(dj)6      -6.8175    0.3482 -19.578 < 2e-16 ***
factor(dj)7      -5.8081    0.4084 -14.222  2.45e-14 ***
factor(dj)8      -9.4477    0.5488 -17.214 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

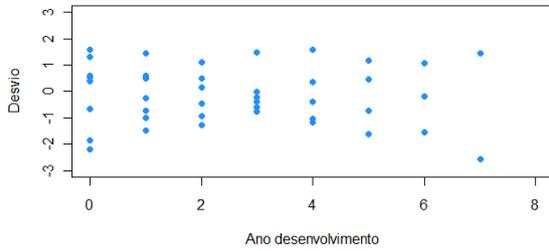
(Dispersion parameter for Tweedie family taken to be 0.2389806)

Null deviance: 205.9896 on 44 degrees of freedom
Residual deviance: 9.0675 on 28 degrees of freedom
AIC: NA

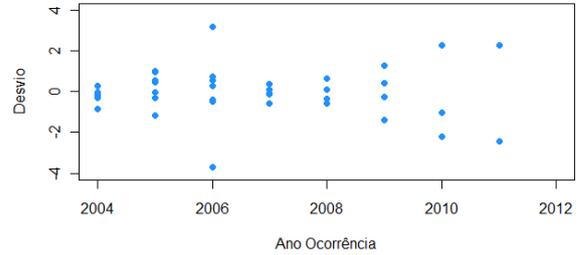
Number of Fisher Scoring iterations: 10
```

Anexo 5. Adequabilidade dos modelos

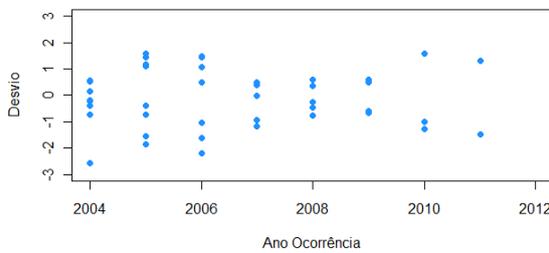
Anexo 5.i - Modelo com estimador de p - Dispersão dos resíduos padronizados vs. Ano desenvolvimento



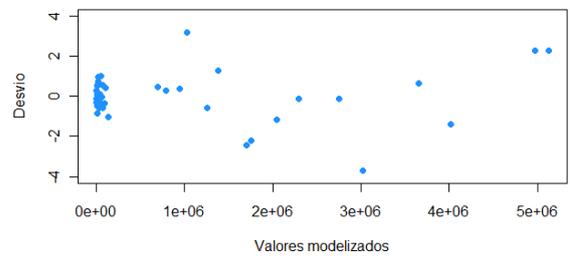
Anexo 5.v - Modelo com $p=1$ - Dispersão dos resíduos padronizados vs. Ano ocorrência



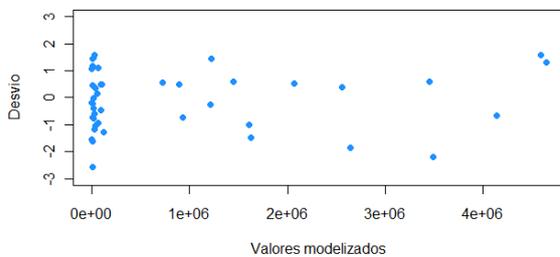
Anexo 5.ii - Modelo com estimador de p - Dispersão dos resíduos padronizados vs. Ano ocorrência



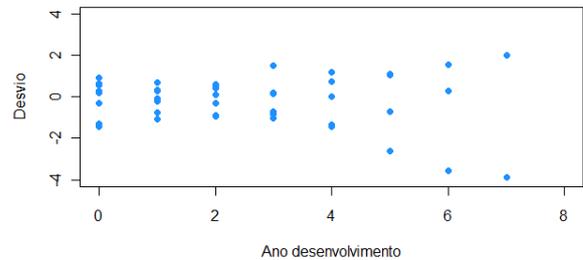
Anexo 5.vi - Modelo com $p=1$ - Dispersão dos resíduos padronizados vs. valores modelizados



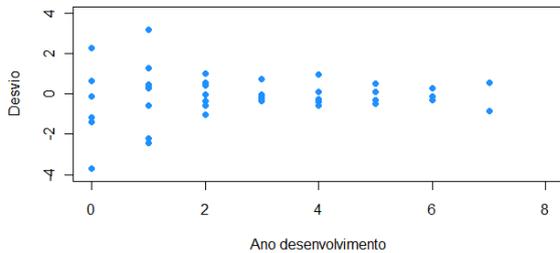
Anexo 5.iii - Modelo com estimador de p - Dispersão dos resíduos padronizados vs. valores modelizados



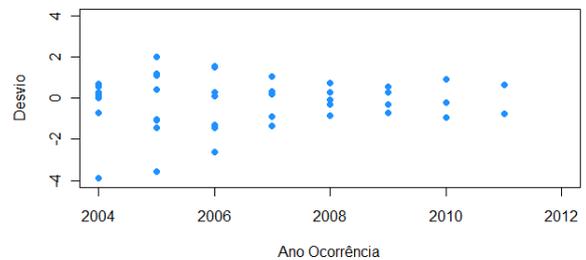
Anexo 5.vii - Modelo com $p=2$ - Dispersão dos resíduos padronizados vs. Ano desenvolvimento



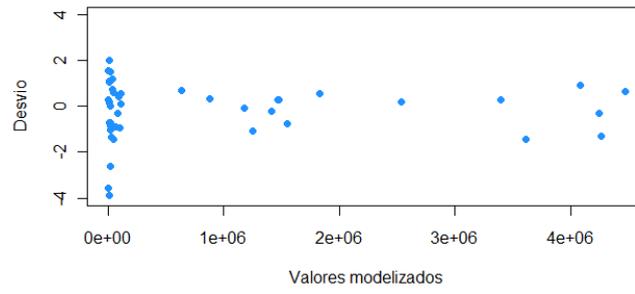
Anexo 5.iv - Modelo com $p=1$ - Dispersão dos resíduos padronizados vs. Ano desenvolvimento



Anexo 5.viii - Modelo com $p=2$ - Dispersão dos resíduos padronizados vs. Ano ocorrência



Anexo 5.ix - Modelo com $p=2$ - Dispersão dos resíduos padronizados vs. valores modelizados



Bibliografia

APS (Associação Portuguesa de Seguradores). (16 de Maio de 2013). Press Release.

Produção de seguro direto cresceu 6% no primeiro trimestre .

Borginho, H. M. (2001). *Metodologias Estocásticas de Estimação de Provisões para Sinistros*. Relatório de estágio - UNL FCT.

Dunn, P. K., & Smyth, G. K. (2008). Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Statistics and Computing, Volume 18*, 73–86.

Instituto de Seguros de Portugal – Dep. de Estatística e Controlo de Informação. (24 de Janeiro de 2013). *Atividade Seguradora - Prémios de Seguro Direto – 2012*.

Jørgensen, B., & Souza, M. (1994). Fitting Tweedie's compound Poisson model to insurance claims data. *Scandinavian Actuarial Journal (1)*, 69-93.

Jørgensen, B. (1987). Exponential Dispersion Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 49, No. 2, 127-162.

Jørgensen, B. (1997). *The Theory of Dispersion Models*. London: Chapman & Hall.

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. London: (2nd ed.) Chapman and Hall.

Nelder, J., & Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A (Royal Statistical Society)*, 135, 370–384.

Portugal L. (2007). *Gestão de seguros Não-Vida*. Lisboa: Instituto de Formação Actuarial.

Ramos, F. M. (Fevereiro de 2000). *Aplicação dos Modelos Lineares Generalizados à Previsão de Reservas para Sinistros*. ISEG-UTL.

Sen, P. K., & Singer, J. d. (1993). *Large sample methods in statistics: An introduction with applications*. New York: Chapman & Hall.

Smyth, G. K. (1996). Regression analysis of quantity data with exact zeros. *Proceedings of the Second Australia–Japan Workshop on Stochastic Models in Engineering, Technology and Management* (pp. 572-580). Australia: University of Queensland.

Smyth, G. K., & Verbyla, A. P. (1999). Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environmetrics* (10), 695-709.

Turkman, M. A., & Silva, G. L. (2000). *Modelos Lineares Generalizados - da teoria à prática*. Lisboa: FCT - PRAXIS XXI.

Tweedie, M. (1947). Functions of a statistical variate with given means with special reference to Laplacian distributions. *Proceedings of the Cambridge Philosophical Society* 43, 41-49.

Tweedie, M. C. (1984). An index which distinguishes between some important exponential families. In Indian Statistical Institute, *Statistics: Applications and New Directions* (pp. 579-604).

Wüthrich, M. V. (2003). Claims Reserving Using Tweedie's Compound Poisson Model.

Astin Bulletin 33(2), 331-346.