2
Reordenação do triângulo *runoff*

Para se prever o IBNR total, os dados de sinistros do tipo IBNR são dispostos em um formato particular, chamado triângulo *runoff*, representado graficamente pela Figura 2.1 (vide Hart *et al.*, 2001 e suas diversas referências). Cada linha do triângulo representa um "ano de acidente"[1] ou "ano de origem", isto é, o ano em que o sinistro ocorreu. Já as colunas referem-se aos "anos de desenvolvimento", que expressam o atraso entre o pagamento e o ano de origem. Os "anos de calendário" (ou "anos de pagamento") são as diagonais do triângulo.

Ano de Origem			Desenvolvimento d		
w	0	1	2	$\dots$	$J-1$
1	C_{10}	C_{11}	C_{12}	$\dots$	C_{1J-1}
2	C_{20}	C_{21}	$\dots$	C_{2J-2}	
3	C_{30}	$\vdots$			
$\vdots$	$\vdots$	$\vdots$			
$\vdots$	$\vdots$	$C_{J-1\,1}$			
J	C_{J0}				

Figura 2.1: Triângulo de *runoff*.

O triângulo possui duas representações básicas: *incremental* e *acumulada*. Na forma *incremental*, cada célula do triângulo é representada por C_{wd}, onde $0 < w \leq J$ e $0 \leq d < J$. O valor C_{wd} é o montante agregado pago pela seguradora referente a sinistros ocorridos no ano de origem w, e atrasados d anos (ou seja, pagos no ano de calendário $w + d$).

Na forma *acumulada* do triângulo, as células são calculadas da forma

$$D_{wd} = \sum_{k=0}^{d} C_{wk},$$

[1]Conforme Hart *et al.* (2001), qualquer unidade de tempo pode ser utilizada. Para algumas classes de negócios pode ser mais adequado utilizar meses, trimestres ou semestres.

isto é, há uma acumulação dos incrementos C_{wd} ao longo das linhas. Em geral, a forma *acumulada* é utilizada em métodos baseados nas razões ou taxas entre anos de desenvolvimento consecutivos, como os modelos em Hertig (1985) e em de Jong (2004), e o próprio método *Chain Ladder* (cf. Booth *et al.*, 2004; England & Verrall, 2002; Taylor, 2000 entre outros) só para citar alguns.

O ponto comum a todos os métodos de previsão de reservas – sobretudo aos aplicados em triângulos *runoff* – é a expectativa de que determinados "padrões" de comportamento dos sinistros ocorridos no passado se repitam no futuro. O padrão em questão é o de atraso entre o período de origem e o de pagamento do sinistro, ou seja, é o padrão apresentado ao longo das colunas. Os métodos estatísticos aplicados ao triângulo tentam modelar esta dependência entre colunas de diversas formas, dentre as quais está a *curva de Hoerl* (cf. Wright, 1990). Em de Jong & Zehnwirth (1983) e Renshaw (1989) a curva é utilizada no contexto de séries temporais, enquanto England & Verrall (2002) o faz no contexto de Modelos Lineares Generalizados, através de uma regressão de variáveis dependentes Poisson com sobredispersão. O trabalho de de Jong (2006), que é uma extensão do trabalho de Hertig (1985), verifica a existência de correlações entre as colunas do triângulo – sobretudo entre as primeiras (atrasos menores) – e as tratam explicitamente, levando a melhoras significativas no ajuste do modelo e, conseqüentemente, no seu poder preditivo. Ainda existem outros trabalhos que incorporam como informação adicional o número de sinistros ocorridos em cada "célula" do triângulo. Vide, por exemplo, Ntzoufras & Dellaportas (2002), no qual é realizada uma comparação entre estimações de reserva levando-se em conta ou não o número de sinistros. Os autores mostram que a inclusão dessa informação reduz o intervalo de confiança da previsão.

Na representação de "índice duplo" – tanto no contexto de séries temporais quanto no de regressão – é comum se parametrizar o triângulo através de fatores comuns às colunas e outros comuns às linhas. Por exemplo, na parametrização mais tradicional, onde $C_{wd} = \alpha_w \beta_d$, todas as observações da linha w compartilham do mesmo parâmetro α_w. A grande desvantagem desta abordagem é o grande número de parâmetros a serem estimados com poucas observações. Alguns trabalhos utilizam-se de reparametrizações no intuito de se reduzir o número de parâmetros a ser estimado. Um exemplo disso é a *curva de Hoerl*, empregada em de Jong & Zehnwirth (1983), que reduz o número de parâmetros que influenciam as colunas para apenas um.

2.1 O método Chain-Ladder

O método determinístico *Chain-Ladder* para previsão de reservas IBNR é, ainda hoje, um dos mais utilizados pelas empresas seguradoras. A principal razão disto, segundo Mack (1993), é sua simplicidade e por ser livre de distribuição[2]. Muitos dos trabalhos citados anteriormente tentaram construir um arcabouço estocástico para este método através da incorporação de premissas básicas, como, por exemplo, assumir uma determinada distribuição para as observações do triângulo. Assim, seria possível obter uma medida de dispersão para a reserva estimada. Porém, a maioria dos trabalhos, na realidade, estava apenas desenvolvendo um novo método – diferente do *Chain-Ladder* – pois os valores estimados das reservas não coincidiam. Coube ao trabalho de Mack (1993) desenvolver o arcabouço estocástico do método *Chain-Ladder* – mantendo idêntico o valor estimado da reserva –, e dessa forma obter a medida de dispersão para tal reserva sem assumir qualquer distribuição para as observações do triângulo, ou seja, preservou-se a característica principal do método *Chain-Ladder*: ser "livre de distribuição".

2.1.1 Expressões básicas

O método *Chain-Ladder* básico assume a existência de $J-1$ fatores ("Development Factors") $f_1, \ldots, f_{J-1} > 0$ tais que

$$\mathrm{E}(D_{w,k}|D_{w1}, \ldots, D_{w,k-1}) = D_{w,k-1} f_{wk}, \quad 1 \leq w \leq J,\ 1 \leq k \leq J-1. \tag{2-1}$$

Os estimadores dos fatores f_k são dados pela seguinte expressão

$$\hat{f}_k = \frac{\sum_{j=1}^{J-k} D_{j,k}}{\sum_{j=1}^{J-k} D_{j,k-1}}, \quad 1 \leq k \leq J-1. \tag{2-2}$$

Assim, a esperança do valor acumulado na última coluna da linha w é dada por

$$\hat{D}_{w,J-1} = D_{w,J-w} \cdot \hat{f}_{J-w} \cdot \ldots \cdot \hat{f}_{J-1} \tag{2-3}$$

As parcelas da reserva referentes a cada linha w (ano de acidente w) são dadas pelo valor obtido na coluna $J-1$, subtraído do valor da diagonal da linha correspondente.

[2]Segundo a classificação criado por Taylor (2003), o método *Chain-Ladder* é heurístico e determinístico.

$$\begin{aligned}\hat{R}_w &= \hat{D}_{w,J-1} - D_{w,J-i} \\ &= D_{w,J-w}\left(\hat{f}_{J-w} \cdot \ldots \cdot \hat{f}_{J-1}\right)\end{aligned}$$

A reserva total é dada pela soma de cada uma destas parcelas

$$\hat{R} = \sum_{i=2}^{J} R_i. \tag{2-4}$$

2.1.2
Expressões do erro médio quadrático obtidas por Mack

A grande contribuição do trabalho de Mack foi a obtenção de uma expressão para o erro médio quadrático da reserva, sem assumir qualquer distribuição subjacente. Para tal, alguns pressupostos tiveram que ser assumidos, como, por exemplo, a independência entre linhas do triângulo. Desta maneira, a expressão para o erro médio quadrático da parcela da reserva correspondente à linha w é dada por

$$\widehat{\mathrm{EQM}(\hat{R}_w)} = \hat{D}^2_{w,J-1} \sum_{k=J-w}^{J-2} \frac{\hat{\sigma}^2_k}{\hat{f}^2_k}\left(\frac{1}{\hat{D}_{w,k}} + \frac{1}{\sum_{j=1}^{J-k} D_{j,k}}\right)$$

com $\hat{\sigma}^2_k$ definido como o seguinte estimador não-viesado:

$$\hat{\sigma}^2_k = \frac{1}{J-k-1}\sum_{i=1}^{J-k} D_{ik}\left(\frac{D_{ik}}{D_{i,k-1}} - \hat{f}_k\right)^2, \quad 1 \leq k \leq J-2.$$

A expressão para o erro médio quadrático da reserva total encontra-se a seguir. Ela será utilizada no capítulo 4 para efeito de comparação com o modelo proposto no capítulo 3.

$$\widehat{\mathrm{EQM}(\hat{R})} = \sum_{i=2}^{J}\left\{(\mathrm{e.p.}(\hat{R}_i))^2 + \hat{D}_{i,J-1}\left(\sum_{j=i+1}^{J}\hat{D}_{j,J-1}\right)\sum_{k=J-i}^{J-2}\frac{2\hat{\sigma}^2_k/\hat{f}^2_k}{\sum_{n=1}^{J-k} D_{nk}}\right\} \tag{2-5}$$

2.2
Modelo em Espaço de Estado proposto

Este trabalho fará uso da forma *incremental* do triângulo, porém, como já citado, adotando-se uma nova ordenação: os índices w e d cedem lugar para um único índice t, o qual define uma série temporal univariada, cuja ordenação está exemplificada na Figura 2.2. Com esta ordenação, e utilizando-se a formulação em Espaço de Estado (a ser discutida na seção 3.1), as estruturas de dependência entre os valores do triângulo podem ser modeladas de forma natural. Cabe, aqui, ressaltar que este índice t não representa a ordem cronológica dos pagamentos, mas sim uma nova maneira de possibilitar a análise do triângulo através de modelos que admitam componentes periódicas, como será visto no próximo capítulo.

Ano de Origem	Desenvolvimento d				
w	0	1	2	...	$J-1$
1	y_1	y_2	y_3	...	y_J
2	y_{J+1}	y_{J+2}	...	y_{2J-1}	y_{2J}
3	y_{2J+1}	y_{2J+2}	...	y_{3J-1}	y_{3J}
⋮	⋮	⋮			
J	$y_{(J-1)J+1}$	$y_{(J-1)J+2}$	...	y_{J^2-1}	y_{J^2}

Figura 2.2: Reparametrização "por linhas" do triângulo.

Ano de Origem	Desenvolvimento d				
w	0	1	2	...	$J-1$
1	y_1	y_2	y_3	...	y_J
2	y_{J+1}	y_{J+2}	...	y_{2J-1}	
⋮	⋮				
J	$y_{(J-1)J+1}$				

Figura 2.3: Reparametrização "por linhas" do triângulo com valores ausentes.

Perseguindo a mesma lógica apresentada na Figura 2.1, alguns elementos da Figura 2.2, na prática, são faltantes, pois correspondem às parcelas do IBNR (vide definição mais adiante). Logo, a figura 2.3 será o verdadeiro objeto de estudo. Prever a Reserva IBNR significa, em termos práticos, completar os valores faltantes do triângulo inferior da figura 2.3. Na abordagem univariada em discussão, a Reserva IBNR propriamente dita consiste na soma não observada desses valores faltantes, ou seja,

$$\text{IBNR} \equiv R = \sum_{t:\, y_t \text{ é ausente}} y_t. \tag{2-6}$$

É no tratamento destes valores ausentes que a formulação em Espaço de Estado apresenta um de seus diferenciais: a estimação se torna natural mediante a utilização do filtro de Kalman. Não só se obtém a previsão pontual de tais valores, mas também sua medida de dispersão (cf. Harvey, 1989; Durbin & Koopman, 2001; e Brockwell & Davis, 2002).

No entanto, enuncia-se um problema:

> Supondo que os valores faltantes são variáveis aleatórias de 2ª ordem e adotando-se como estimador pontual de R a sua esperança condicional dados todos os outros valores conhecidos do triângulo, *calcular seu erro médio quadrático associado*, o qual, sabe-se, é mínimo dentre todos os outros estimadores que são funções dos valores conhecidos.

Neste trabalho, o problema delimitado acima é atacado mediante duas abordagens diferentes, mas que possuem como base a teoria dos modelos em Espaço de Estado e a mesma ordenação por linhas já proposta. Tais temas são explorados plenamente no próximo capítulo.