

3

Modelos para o Cálculo de IBNR

3.1

O Método de Mack

Tomas Mack em (24) propõe um modelo probabilístico para o método Chain Ladder que fornece estimativas de provisão idênticas à técnica *Chain Ladder* e permite prever erros nas estimativas realizadas. Essa abordagem é dita “livre de distribuição”, ou seja, não assume distribuição específica para os dados. Porém, ela assume hipóteses sobre as variáveis aleatórias $A_{i,j}$ correspondentes aos pagamentos acumulados dos sinistros.

3.1.1

Hipóteses de Mack

Para Mack (24), os dados $A_{i,j}$, para os diferentes anos de ocorrência i , são variáveis aleatórias independentes, isto é,

$$\{A_{i,1}, \dots, A_{i,n}\}, \{A_{k,1}, \dots, A_{k,n}\}, i \neq k, \text{ são independentes.} \quad (3-1)$$

A média e a variância das variáveis $A_{i,j}$ correspondentes aos períodos futuros são dadas por:

$$E[A_{i,j+1}|A_{i,0}, \dots, A_{i,j}] = E[A_{i,j+1}|A_{i,j}] = f_j A_{i,j} \quad (3-2)$$

$$Var[A_{i,j+1}|A_{i,0}, \dots, A_{i,j}] = Var[A_{i,j+1}|A_{i,j}] = \sigma_j^2 A_{i,j} \quad (3-3)$$

Na Equação 3-2, $f_0, f_1, \dots, f_j > 0$ correspondem aos fatores de desenvolvimento e na equação 3-3 $\sigma_0^2, \dots, \sigma_{j-1}^2 > 0$ correspondem aos parâmetros de variância para $0 \leq i \leq n$ e $1 \leq j \leq n - 1$.

As Equações 3-1, 3-2 e 3-3 são as chamadas *hipóteses de Mack* e são inerentes à técnica *Chain Ladder* padrão. Mack demonstrou em seu artigo que, sob certas condições, os estimadores \hat{f}_j dados pela Expressão 2-1 são não-viesados e de variância mínima.

De maneira semelhante, pode ser mostrado que os parâmetros de variância σ_j^2 , para $1 \leq j \leq n - 2$, estimados pela equação:

$$\hat{\sigma}_j = \frac{1}{n-j-1} \sum_{i=1}^{n-j} A_{i,j} \left(\frac{A_{i,j+1}}{A_{i,j}} - \hat{f}_j \right)^2 \quad (3-4)$$

são também não-viesados.

Dessa maneira, falta um estimador para σ_{n-1}^2 . Se $\hat{f}_{j-1} = 1$, então pode-se colocar $\sigma_{j-1}^2 = 0$. Se isso não ocorre, pode-se explorar a série exponencial decrescente, $\sigma_0^2, \dots, \sigma_{j-2}^2$, adicionando mais um termo a ela por meio da regressão log-linear. Ou então, de forma mais simples, pode-se fazer uma aproximação para $\hat{\sigma}_{j-1}^2$, exigindo-se que $\frac{\sigma_{j-3}^2}{\sigma_{j-2}^2} = \frac{\sigma_{j-2}^2}{\sigma_{j-1}^2}$ pelo menos enquanto $\sigma_{j-3}^2 > \sigma_{j-2}^2$. Essa condição conduz ao estimador $\hat{\sigma}_{j-1}^2 = \min\left(\frac{\hat{\sigma}_{j-2}^4}{\hat{\sigma}_{j-3}^2}, \min(\hat{\sigma}_{j-3}^2, \hat{\sigma}_{j-2}^2)\right)$, o que completa a relação de todos os estimadores de σ_j^2 , para $1 \leq j \leq n-1$. Essa aproximação é chamada *aproximação de Mack*.

Na Equação 3-4, $n-j$ é o número de resíduos obtidos na estimação dos valores futuros. No denominador, o valor um é subtraído do número de resíduos $n-j$ para completar a demonstração de que os estimadores sejam também não-viesados. Para Verrall em (9), o parâmetro de variância σ^2 é estimado como uma média ponderada de resíduos e os parâmetros estimados $\hat{\sigma}^2$ são utilizados no cálculo de erros de previsão nas estimativas dos pagamentos futuros.

3.1.2

Erro de Previsão na Estimativa das Reservas

Para avaliar a variabilidade e a incerteza sobre as estimativas é comum utilizar como medida de referência o erro médio quadrático. Esse tipo de erro é conhecido pela sigla *EMQ*.

No caso das estimativas para pagamentos futuros dos sinistros, o *EMQ* é definido por:

$$EMQ(\hat{A}_{i,n}) = E((\hat{A}_{i,n} - A_{i,n})^2 | D), \text{ onde } D = \{A_{i,j} | i + j \leq n + 1\} \quad (3-5)$$

Nessa expressão, D representa o conjunto de todas as informações conhecidas. De maneira análoga, o erro médio para as estimativas das reservas por data de ocorrência i , \hat{R}_i , é equivalente a:

$$EMQ(\hat{R}_i) = E((\hat{R}_i - R_i)^2 | D) = E((\hat{A}_{i,n} - A_{i,n})^2 | D) = EMQ(\hat{A}_{i,n}) \quad (3-6)$$

Assim, usando a Equação 3-6, $A_{i,n} \in D$, e o fato de que $E[X - a]^2 = Var(X) + (E[X] - a)^2$, o erro médio quadrático da estimativa \hat{R}_i pode ser reescrito da seguinte forma:

$$EMQ(\hat{R}_i) = Var(R_{i,n} | D) + (E(\hat{A}_{i,n} | D) - A_{i,n})^2 \quad (3-7)$$

De acordo com a Expressão 3-7 é necessário estimar também o processo de variância envolvido no EMQ . De acordo com Mack em (24), o processo de variância por ano de ocorrência e seu estimador são dados pelas seguintes expressões:

$$\left. \begin{aligned} Var(R_i) &\approx \hat{A}_{i,n}^2 \sum_{j=n-i+1}^{n-1} \frac{\sigma_{j+1}^2}{\hat{f}_{j+1} \hat{A}_{i,j}} \\ Var(\hat{R}_i) &\approx \hat{A}_{i,n}^2 \sum_{j=n-i+1}^{n-1} \frac{\sigma_{j+1}^2}{\hat{f}_{j+1} \sum_{q=1}^{n-j} \hat{A}_{q,j}} \end{aligned} \right\} \quad (3-8)$$

Com junção do estimador $Var(\hat{R}_i)$ exposto na Equação 3-8 e das hipóteses de Mack mencionadas na Seção 3.1.1, é possível demonstrar que:

$$EMQ(\hat{R}_i) = \hat{A}_{i,n}^2 \sum_{j=n-i+1}^{n-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left(\frac{1}{\hat{A}_{i,j}} + \frac{1}{\sum_{t=1}^{n-j} A_{t,j}} \right) \quad (3-9)$$

onde $\hat{A}_{i,j} = A_{i,n-i+1} \hat{f}_{n-i+1} \cdots \hat{f}_{j-1}$, $j > n - i + 1$, correspondem aos valores estimados dos pagamentos futuros e $\hat{A}_{i,n-i+1} = A_{i,n-i+1}$.

A Equação 3-9 é o principal resultado obtido por Tomas Mack em (24) e a partir dele obtém-se o erro padrão para a reserva IBNR total. Define-se como desvio padrão, $d.p.$, de \hat{R}_i a raiz quadrada do estimador do erro médio quadrático, isto é, $s.e.(\hat{R}_i) = \sqrt{EMQ(\hat{R}_i)}$.

O desvio padrão $d.p.$ da reserva total, $\hat{R}_{Total} = \hat{R}_2 + \cdots + \hat{R}_n$, não é obtido simplesmente pela soma dos erros padrões obtidos das reservas por ano de ocorrência i , com $2 \leq i \leq n$, pois eles estão correlacionados por meio de seus estimadores em comum \hat{f}_j e $\hat{\sigma}_j^2$. Sendo assim, pelo resultado 3-9, o erro médio quadrático da reserva IBNR total pode ser estimado pela Equação 3-10:

$$EMQ(\hat{R}_{Total}) = \sum_{i=2}^n \left\{ (s.e.(\hat{R}_i))^2 + \hat{A}_{i,n} \left(\sum_{k=i+1}^n \hat{A}_{k,n} \right) \sum_{j=n+1-i}^{n-1} \frac{\frac{2\sigma_j^2}{\hat{f}_j^2}}{\sum_{t=1}^{n-j} A_{t,j}} \right\} \quad (3-10)$$

Exemplo 3.1.1 *Este exemplo apresenta os resultados do método Mack Chain Ladder disponível no pacote **Chain Ladder** da plataforma R, para o triângulo exposto na Tabela 2.3.*

A Tabela 3.1 apresenta a estimativa dos fatores f_j e dos parâmetros σ_j e a Tabela 3.2 mostra os erros de previsão das reservas por data de ocorrência \hat{R}_i e da reserva total IBNR \hat{R}_{Total} . Resultados esses obtidos por meio dos estimadores desenvolvidos por Mack.

Parâmetros	Desenvolvimento j									
	1	2	3	4	5	6	7	8	9	10
f_j	2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009	1.000000
σ_j	166.983	33.295	26.295	7.825	10.929	6.389	1.159	2.808	0.803	

Tabela 3.1: Fatores de desenvolvimento e parâmetros de variância para o Exemplo 3.1.1.

Ano de origem	Reserva	Erro de previsão	Erro de previsão%
1	-	-	-
2	153.95	143	91%
3	617.37	592	96%
4	1636.14	713	47%
5	2746.74	1452	53%
6	3649.10	1995	56%
7	5435.30	2204	41%
8	10907.19	5354	49%
9	10649.98	6332	60%
10	16339.44	24566	150%
<i>Total</i>	52135.2	26880.74	52%

Tabela 3.2: Reservas e erros de previsão para o Exemplo 3.1.1.

3.2

Método Bootstrap Chain Ladder

Este modelo corresponde a abordagem feita por *England* e *Verrall* em (9). Ele está baseado na estrutura proposta por *Kremer* em (17) e no modelo proposto por *Renshaw* e *Verrall* em (28). O objetivo principal dessa abordagem é reduzir a variância do método Chain Ladder proposto por *Mack*. O método *Bootstrap Chain Ladder* é paramétrico, pois assume uma certa distribuição para os dados.

A grande vantagem desta abordagem é que nela está incorporada a teoria dos modelos lineares generalizados (MLG), que é muito utilizada na modelagem estatística. Além disso, este método utiliza a técnica de *bootstrap* para estimar erros de previsão quando os erros de previsão analíticos envolvem fórmulas complexas. Assim, resíduos podem ser usados para prever o erro padrão produzido pelo *bootstrap*, desde que a definição dos resíduos esteja apropriada ao modelo em questão.

Basicamente, o método denominado *BootChainLadder* utiliza 2 estágios:

1. *Bootstrapping*: calcula-se resíduos de Pearson sobre os dados iniciais e aplica-se a técnica *bootstrap* sobre eles para assim obter um novo conjunto de dados incrementais. A partir daí se faz a previsão para os incrementos futuros dos pagamentos dos sinistros via *Chain Ladder* padrão e conseqüentemente se obtém uma nova amostra de dados acumulados e

uma estimativa para a reserva IBNR. Isso significa que, fazendo N “reamostragens” se obtém N pseudo-conjuntos de dados e a partir é possível fazer estimativas para as estatísticas de interesse;

2. *Simulação*: simula-se o processo de erro, onde os valores obtidos por meio de *bootstrap* são utilizados como médias, assumindo-se um processo de distribuição e um modelo teórico adequado.

O conjunto das reservas obtidas dessa maneira forma uma amostra de dados da distribuição prevista, de forma que medidas estatísticas como média, erro de previsão e quantis podem ser avaliados.

A seguir estão expostos conceitos e pressupostos inerentes ao desenvolvimento do método *Bootstrap Chain Ladder*. São eles: Modelo de Kremer, Modelos Lineares Generalizados, Modelo de Renshaw e Verrall e a técnica *Bootstrap*. No Apêndice A está descrito o passo-a-passo desse modelo de acordo com (10).

Modelo de Kremer

Kremer [1982] em (17) desenvolveu um dos primeiros estudos sobre modelos loglineares. Para ele, em cada pagamento futuro previsto pelo método *Chain Ladder* dado por $\hat{A}_{i,j+1} = A_{i,j}\hat{f}_j$ está implícito a interação de dois fatores: um parâmetro relacionado com a coluna β_j (presente em f_j) e um parâmetro relacionado a linha α_i (que reflete a influência do valor $A_{i,j}$). Ele considera o logaritmo dos valores incrementais $I_{i,j}$ como variáveis respostas de um modelo log-normal, ou seja, as variáveis aleatórias $I_{i,j}$ seguirão uma distribuição de probabilidade lognormal.

Considere $I_{i,j}$ os valores incrementais decorrentes da data de origem i e data de aviso j , $Y_{i,j} = \log(I_{i,j})$ com $Y_{i,j} = m_{i,j} + \varepsilon_{i,j}$, para os quais sejam válidas as seguintes equações:

$$Y_{i,j} \sim N(m_{i,j}, \sigma^2), \quad (3-11)$$

$$\varepsilon_{i,j} \sim N(0, \sigma^2), \quad (3-12)$$

$$m_{i,j} = \eta_{i,j}, \quad (3-13)$$

$$\eta_{i,j} = c + \alpha_i + \beta_j, \quad \alpha_1 = \beta_1 = 0 \quad (3-14)$$

As Equações 3-11, 3-12, 3-13 e 3-14 definem o modelo introduzido por Kremer. Os parâmetros α_i e β_j são estimados por máxima verossimilhança ou pelo método dos mínimos quadrados. A variância desconhecida σ^2 é estimada

pela soma dos quadrados dos resíduos $\hat{\varepsilon}_{i,j}$ divididos pelos graus de liberdade. Assim, com o uso dos resultados de modelos de regressão linear é possível estimar o valor esperado e o erro quadrático médio das variáveis aleatórias $Y_{i,j} = \log(I_{i,j})$.

Dadas as estimativas dos parâmetros, a transformação dos valores estimados da escala logarítmica para escala real pode ser obtida considerando que $I_{i,j} = \exp(Y_{i,j})$, porém esse tipo de transformação produz estimativas enviesadas. Informações adicionais podem ser encontrados em (27) e (32).

Apesar de não produzir estimativas idênticas ao método *Chain Ladder*, este modelo pode ser estendido para outras abordagens devido as várias alternativas possíveis para o preditor linear dado por $\eta_{i,j} = c + \alpha_i + \beta_j$. Esse fato, permite a construção de novos modelos loglineares que reflitam de forma mais adequada os dados em análise e produzam melhores estimativas.

Modelos Lineares Generalizados

Os Modelos Lineares Generalizados (MLG) têm sua origem nos modelos lineares de regressão. Com uma estrutura bem flexível, estes modelos fornecem maneiras de estimar parâmetros, ajustar medidas e definir adequadamente os resíduos a serem utilizados no problema em questão. Esses modelos de regressão são utilizados em variáveis aleatórias com distribuição de probabilidade pertencente à família das exponenciais (Poisson, Normal, Gama, Over-dispersed Poisson, etc.).

A estrutura de um MLG pressupõe a existência de um preditor linear para cada variável aleatória. Este preditor está relacionado com a média por meio de uma função de ligação g , monótona e diferenciável em seu domínio. Formalmente:

$$\left. \begin{aligned} E[Y_i] &= \mu_i \\ \eta_i &= g(\mu_i) \\ \eta_i &= \gamma_0 + \gamma_1 X_{i,1} + \dots + \gamma_k X_{i,k} \end{aligned} \right\} \quad (3-15)$$

A estimação dos parâmetros $\gamma_0, \gamma_1, \dots, \gamma_k$ é feita por máxima verossimilhança ou pelo método dos mínimos quadrados com recursos de algoritmos de aproximação para resolver problemas otimização não-linear, como por exemplo IRLS, *Iteratively Reweighted Least Squares*. Uma vez feito isso e conhecendo a distribuição dos dados, é possível escrever de forma analítica uma expressão para os dois primeiros momentos das variáveis aleatórias Y_i . Com isso se obtém também a respectiva matriz de variância-covariância resultante do processo de

estimação e conseqüentemente uma estimativa para os valores para as variáveis Y_i .

Abordagem de Renshaw e Verrall (1994)

Esta abordagem é baseada na estrutura de Kremer exposta neste capítulo. Nela os dados incrementais $I_{i,j}$ são considerados diretamente como variáveis resposta e a função de logarítmica é utilizada como a função de ligação g entre a média e o preditor linear. Além disso, ainda é assumido que os dados seguem uma distribuição “over-dispersed” *Poisson* de acordo com (10). Formalmente:

$$\left. \begin{aligned} E[I_{i,j}] &= m_{i,j} \\ \text{Var}[I_{i,j}] &= \phi E[I_{i,j}] = \phi m_{i,j} \\ \eta_{i,j} &= \log(m_{i,j}) \\ \eta_i &= c + \alpha_i + \beta_j, \quad \alpha_1 = \beta_1 = 0 \end{aligned} \right\} \quad (3-16)$$

A estrutura 3-16 define um modelo linear generalizado, onde as variáveis respostas $I_{i,j}$ são modeladas de forma que a função logarítmica seja a função de ligação entre a média e o preditor linear, e a variância seja proporcional a média. O parâmetro ϕ é desconhecido será estimado durante o procedimento. Atualmente, existem softwares estatísticos para MLG preparados para estimar os parâmetros α_i e β_j e prever então os valores em questão, o que torna este modelo muito difundido, como os pacotes **MASS** e **stats**, disponíveis na plataforma R.

É importante salientar que este modelo é robusto para um número pequeno de incrementos negativos, visto que as variáveis resposta são os próprios incrementos e a função logaritmo é aplicada a valores estritamente positivos. Então, pela forma como este modelo foi estruturado e os parâmetros estimados, é necessário impor a condição de que a soma dos incrementos seja sempre positiva. Isso limita o modelo e o torna inadequado para alguns casos.

Outras versões para este modelo podem ser encontradas na literatura. Uma que se destaca é o trabalho de Mack (1991), que propõe que as variáveis resposta sigam uma distribuição de probabilidade *Gama*. Informações complementares estão descritas no artigo (23).

Técnica Bootstrap

Esta técnica foi introduzida por Efron em 1979 (8) e consiste na “reamostragem” de resíduos, com reposição. Ela se baseia na geração de repetições da amostra inicial, por meio de simulações, permitindo assim a obtenção de resultados para as estatísticas de interesse, bem como a possibilidade de inferir sobre a variabilidade das estimativas obtidas. Resumidamente a técnica consiste na descrição abaixo, considerando que $X = (X_1, \dots, X_n)$ seja uma amostra contendo n observações:

1. Constrói-se m amostras, X^*_1, \dots, X^*_m , independente e identicamente distribuídas, de maneira que cada uma contenha n observações;
2. As m amostras *iid*, construídas a partir da população finita $X = (X_1, \dots, X_n)$ correspondem a “reamostrar” a população, permitindo que alguns dados se repitam.

O *Bootstrap* surgiu como abordagem ao cálculo de intervalos de confiança e parâmetros, em circunstâncias onde outras técnicas não são aplicáveis, em particular no caso em que o tamanho da amostra é reduzido. Ele pode ser utilizado sem a necessidade de assumir uma distribuição de probabilidade para os dados. Contudo quando esta técnica é ajustada a um modelo teórico, que assume alguma distribuição, é possível inferir sobre sua adequação ao problema, por meio de comparações entre resultados obtidos por simulações por *bootstrap* e as estimativas teóricas. Geralmente esta técnica é associada a modelos de previsão, de forma que as estimativas para os valores esperados dos pagamentos futuros sejam obtidas de forma analítica e as medidas de variabilidade sejam obtidas por meio simulação.

Em problemas de regressão, é comum usar a técnica de *bootstrap* sobre um conjunto de resíduos gerados pelo modelo e não diretamente sobre dados observados. Isso porque os resíduos podem ser utilizados para explorar adequação do modelo ajustado no que diz respeito à escolha da função de ligação, da variância e de termos do preditor linear, entre outras possibilidades. Assume-se que os resíduos são independentes e identicamente distribuídos e a partir do conjunto residual são feitas “reamostragens” e em cada uma delas, é gerado um conjunto de pseudo-dados que é usado no cálculo das estatísticas pretendidas.

A Figura 3.1 ilustra as etapas das reservas estimadas via *Bootstrapping* para o modelo de proposto por *England* e *Verrall* em (9). No diagrama, o triângulo (1) representa os dados iniciais e (2) representa os dados ajustados

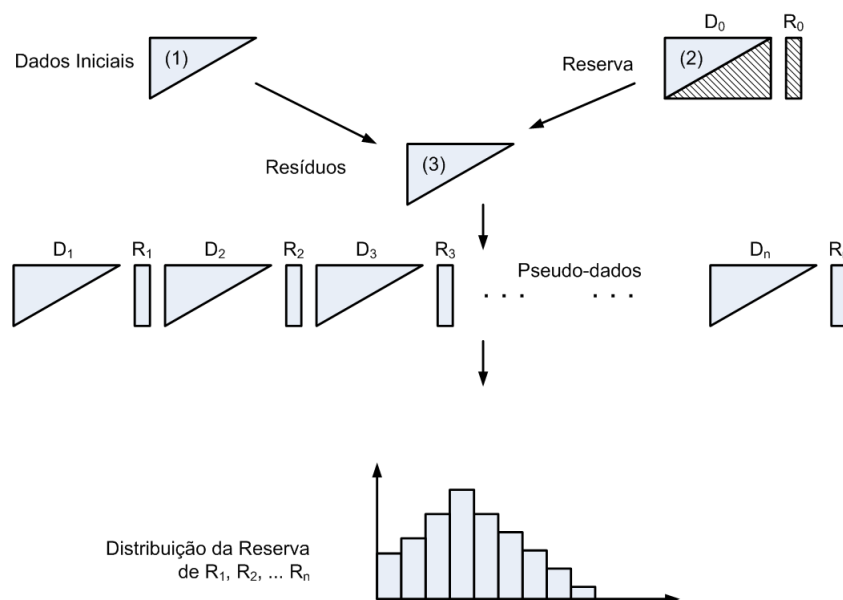


Figura 3.1: Etapas do procedimento *Bootstrapping* para cálculo de reservas. (22).

pelo modelo com sua respectiva reserva calculada via CL padrão. A diferença entre (1), os dados reais, e (2), os dados ajustados pelo modelo, fornece um conjunto de resíduos (3). Esse conjunto é considerado como uma representação de em que medida os dados reais e o modelo podem ser diferentes. Em outras palavras, se os dados são apenas uma realização de algum processo aleatório, outra realização desse processo poderia levar a um outro conjunto de dados que é diferente em qualquer ponto do modelo. A estratégia é, então, produzir lotes de outras possíveis realizações dos dados por “reamostragem” residuais e adicioná-los ao modelo para produzir conjuntos de dados possíveis, conhecidos como pseudo-dados. Para cada triângulo assim produzido, o método de reserva é executado, de modo que uma série de estimativas da reserva são produzidas, as chamadas pseudo-reservas. Se esse procedimento for feito um grande número de vezes, será formada uma grande coleção de pseudo-reservas, a qual terá certa distribuição. A variação desses pseudo-reservas que fornece uma medida da variabilidade da estimativa de reservas.

3.2.1

Erro de Previsão na Estimativa das Reservas

O erro médio quadrático de previsão das estimativas para os modelos Log-normal, Gama e “over-dispersed” Poisson pode ser considerado como a soma de duas componentes: a variabilidade dos dados (processo de variância) e a variabilidade das estimativas (variância estimada). Ou seja, o *EMQ* dos incrementos, pode ser escrito da seguinte forma:

$$EMQ(\hat{I}_{i,i}) = E[(I_{i,j} - \hat{I}_{i,j})^2] \cong Var[I_{i,j}] + Var[\hat{I}_{i,j}] \quad (3-17)$$

Uma expressão precisa para a variância é definida pela distribuição de probabilidade assumida pelo modelo. No caso em questão, a distribuição é “over-dispersed” Poisson e a variância é definida por:

$$Var[I_{i,j}] = \phi m_{i,j} \quad (3-18)$$

O fato de que $\hat{I}_{i,j} = m_{i,j} = \exp(\eta_{i,j})$ e a utilização do *Método Delta* 3-19, que fornece uma aproximação para a variância das estimativas, possibilitam a estimação do erro quadrático médio por meio da Expressão 3-20.

$$Var[\hat{I}_{i,j}] \cong \left| \frac{\partial m_{i,j}}{\partial \eta_{i,j}} \right|^2 Var[\eta_{i,j}] \quad (3-19)$$

$$EMQ(\hat{I}_{i,i}) = E[(I_{i,j} - \hat{I}_{i,j})^2] \cong \phi m_{i,j} + m_{i,j}^2 Var[\eta_{i,j}]. \quad (3-20)$$

A variância do preditor linear $\eta_{i,j}$ pode ser estimada por meio de softwares estatísticos sem grandes dificuldades.

Erro de Previsão Analítico

Vale ressaltar aqui que o desvio padrão, $d.p.$, é a raiz quadrada do erro médio quadrático médio EMQ e que

$$R_i = \sum_{j=n-i+1}^n I_{i,j} \quad e \quad R_{Total} = \sum_{i=2, j=n-i+1}^n I_{i,j}.$$

De acordo com England e Verrall em (10), o desvio padrão das estimativas para as reservas por ano de origem R_i e para a reserva total R_{Total} podem ser calculados respectivamente por meio das Equações 3-21 e 3-22, respectivamente.

$$EMQ(\hat{R}_i) = E[(R_i - \hat{R}_i)^2] \cong \sum_{j=n-i+1}^n \phi m_{i,j} + \sum_{j=n-i+1}^n m_{i,j}^2 Var[\eta_{i,j}] + \quad (3-21)$$

$$+ 2 \sum_{\substack{j_1, j_2 = n-i+1, \\ j_1 > j_2}}^n m_{i,j_1} m_{i,j_2} Cov[\eta_{i,j_1} \eta_{i,j_2}]$$

$$\begin{aligned}
EMQ(\hat{R}_{Total}) = E[(R_{Total} - \hat{R}_{Total})^2] \cong & \sum_{i,j=n-i+1}^n \phi m_{i,j} + \sum_{i,j=n-i+1}^n m_{i,j}^2 Var[\eta_{i,j}] + \\
& + 2 \sum_{\substack{i_1, i_2=n-i+1, \\ j_1, j_2=n-i+1, \\ i_1, j_1 \neq i_2, j_2}}^n m_{i_1, j_1} m_{i_2, j_2} Cov[\eta_{i_1, j_1} \eta_{i_2, j_2}]
\end{aligned} \tag{3-22}$$

Erro de Previsão no Bootstrap Chain Ladder

Na situações em que o desvio padrão é muito difícil de se calcular analiticamente é comum a utilização da técnica *Bootstrap* sobre resíduos obtidos nas estimativas dos valores futuros. É importante salientar a necessidade de usar resíduos apropriados para o problema em questão. No caso do modelo definido pelas Equações 3-16, são utilizados os resíduos de *Pearson* descritos pela expressão:

$$r_{i,j}^{(P)} = \frac{I_{i,j} - m_{i,j}}{\sqrt{m_{i,j}}} \tag{3-23}$$

Depois de calculados os resíduos, uma amostra *bootstrap* para os dados incrementais é então criada por meio da inversão da Equação 3-23, utilizando os resíduos já reamostrados $r_{i,j}^{(*)}$ juntamente com os valores já “ajustados” $\hat{m}_{i,j}$, de acordo com a fórmula:

$$I_{i,j}^{(*)} = r_{i,j}^{*} \sqrt{\hat{m}_{i,j}} + \hat{m}_{i,j} \tag{3-24}$$

Após esse procedimento, uma amostra *bootstrap* para os dados incrementais é obtida, o modelo é “reajustado” e as estatísticas de interesse são calculadas de forma analítica. Ou seja, um conjunto de pseudo-dados é obtido para um triângulo incremental I e conseqüentemente para um triângulo acumulado A , o que permite utilizar a técnica *Chain Ladder* padrão para se obter estimativas para as reservas. Esse processo é repetido um grande número N de vezes e em cada repetição uma nova amostra *bootstrap* é obtida, assim como novas estimativas para as reservas, o que permite obter um conjunto de estimativas sobre o qual é possível calcular o erro padrão ou até a distribuição empírica de probabilidade.

O desvio padrão *bootstrap*, *d.p.*, das reservas estimadas é o desvio padrão das N reservas *bootstrap* obtidas por esta técnica, isto é, a raiz quadrada estimada da variância prevista para as reservas. Contudo, é necessário também obter uma estimativa para o parâmetro ϕ que neste modelo, para dar consistência usa-se:

$$\hat{\phi} = \frac{\sum r^{(P)^2}}{n - p}, \quad (3-25)$$

onde n é o número de dados conhecidos da amostra inicial e p é o número de parâmetros estimados no modelo.

Para uma comparação entre a variância estimada pelo processo de *bootstrap* e a variância estimada analiticamente, é necessário fazer ajustes levando-se em conta o número de parâmetros usados no modelo “ajustado”.

Para obter o erro da previsão *bootstrap* é necessário adicionar um termo que corresponde a estimativa do processo de variância, que é dada pela multiplicação entre o escalar ϕ e as reservas estimadas, ao erro padrão. Assim, o erro de previsão *bootstrap* é dado por:

$$DP_{bootstrap} = \sqrt{\phi^{(P)}R} + \frac{n}{n - p}(d.p.bootstrap(R))^2, \quad (3-26)$$

onde R corresponde a reserva por data de origem ou reserva total, e $(d.p.bootstrap(R))$ corresponde ao erro padrão *bootstrap* da reserva estimada.

Exemplo 3.2.1 *Este exemplo mostra a comparação dos resultados obtidos pelos dois métodos abordados até aqui, tanto para as reservas estimadas quanto para os erros de previsão. Para a obtenção desses resultados, foram utilizados os métodos MackChainLadder e BootChainLadder, disponíveis no pacote ChainLadder da plataforma R, aplicados ao triângulo RAA, também disponível no mesmo pacote.*

Data de origem i	Métodos		
	Mack CL	Boot CL	CL padrão
1981	-	-	-
1982	154	177	153.95
1983	617	571	617.37
1984	1636	1607	1636.14
1985	2747	2887	2746.74
1986	3649	3725	3649.10
1987	5435	5560	5435.30
1988	10907	11153	10907.19
1989	10650	10746	10649.98
1990	16339	17427	16339.44
<i>Total</i>	52135.23	53943	52135.2

Tabela 3.3: Reservas Estimadas no Exemplo 3.2.1.

Data de origem i	Métodos	
	Mack CL	Boot CL
1981	0	0
1982	143	645
1983	592	1226
1984	713	1948
1985	1452	2395
1986	1995	2436
1987	2204	3218
1988	5354	5088
1989	6332	5973
1990	24566	24573
<i>Total</i>	26880.74	19087

Tabela 3.4: Erros de previsão das reservas no Exemplo 3.2.1.