

4

Modelo de Perdas Operacionais

“When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.”

William Thompson, Lord Kelvin

Este é o principal capítulo deste trabalho, pois não há como controlar, e, por conseguinte, gerenciar algo, sem medi-lo. Sendo assim, o capítulo apresenta o Modelo de Distribuição de Perdas (LDA) como uma métrica para o risco operacional.

Portanto, em primeiro lugar formaliza-se a abordagem clássica de perdas operacionais, prosseguindo em discussões técnicas a cerca de métodos intermediários, cujo objetivo final está no cálculo e na alocação do capital regulatório destinado à proteção contra o risco operacional. Para tanto, é necessário determinar os modelos de severidade e de frequência, peças fundamentais na composição do modelo de perdas agregadas ou perdas operacionais.

4.1.

Modelo de Distribuição de Perdas (LDA)

Decidiu-se utilizar o LDA para a mensuração do risco por sua objetividade, uma vez que não necessita da interpretação gerencial dos indicadores de risco, dependendo obrigatoriamente do histórico de perdas observadas para previsão da distribuição das perdas futuras; e por permitir fazer predições, análise de cenários, teste de estresse e análise de custo-benefício – inclusive para transferência de risco (Coleman, 2000) [20].

Adicionalmente, sob o ponto de vista metodológico este modelo parece ser menos complicado para construção que os modelos internos para risco de crédito e de mercado (Frachot et al., 2001) [42].

Além disso, o LDA é mais sensível ao risco, pois os dados internos são usados para calibrar o processo. Isso torna esta métrica mais ligada ao real risco de cada banco, e, portanto ao seu gerenciamento, podendo minimizar o capital alocado em detrimento aos métodos do tipo “*one-size-fits-all methods*”.

No apêndice 6, o Comitê da Basileia (2001a) [6] apresenta o LDA conforme transcrito a seguir: “*Under the Loss Distribution Approach, the bank estimates, for each business line/risk type cell, the probability distribution functions of the single event impact and the event frequency for the next (one) year using its internal data, and computes the probability distribution function of the cumulative operational loss.*”.

Alguns comentários sobre a transcrição acima:

- O LDA é uma modalidade de AMA, pois cabe ao banco estimar suas próprias distribuições de probabilidade.
- A Basileia espera que a medida do risco operacional seja feita por linha de negócio e por tipo de evento de perda, conforme visto nos capítulos 2 e 3.
- O LDA se baseia em modelos preditivos para a distribuição de probabilidade da severidade e da frequência, um ano à frente, para cada unidade de risco – por linha de negócio / tipo de evento de perda. Normalmente, o modelo de previsão é ingênuo, considerando as distribuições de probabilidade da severidade e da frequência de uma unidade de risco um ano à frente iguais às respectivas distribuições de probabilidade subjacentes baseadas somente nos dados internos (Coleman, 2003b) [24].
- Pode-se concluir erroneamente que a calibração das distribuições de probabilidade da severidade e da frequência dos eventos seja feita única e exclusivamente por intermédio de dados internos. Conforme visto no capítulo 3, a Basileia permite a conjunção dos dados internos com dados externos e opiniões de especialistas para a AMA. Isso possibilita a criação de modelos de previsão para as distribuições de probabilidade da severidade e da frequência mais robustos e menos ingênuos.
 - Os dados externos e os dados simulados em cenários de estresse tornam o modelo mais robusto, pois captam perdas extremas que comumente não são feitas com os dados internos.

- As opiniões de especialistas se traduzidas em dados permitem uma previsão do comportamento dos controles internos um ano à frente, tornando o modelo preditivo menos ingênuo.
- Além da conjunção dos dados internos com os dados externos e as opiniões de especialistas, uma outra solução para tornar os modelos de previsão menos ingênuos é dar maior peso às perdas mais recentes. Como sugestão, pode-se aplicar um alisamento exponencial aos impactos financeiros nos dados internos. Em toda a revisão da literatura não foi encontrada proposta semelhante.
- Por fim, calcula-se a distribuição de probabilidade das perdas operacionais de uma unidade de risco como sendo o somatório das severidades das perdas operacionais desta unidade de risco. É importante destacar que para cada unidade de risco a severidade é uma variável aleatória, assim como a quantidade de perdas nesta unidade de risco (frequência), sendo suas distribuições de probabilidade um ano à frente estimadas pelos modelos internos de cada banco.

Entretanto, a distribuição de probabilidade das perdas operacionais por si só não representa uma medida simples para o risco operacional. Por isso, o Novo Acordo estabelece que para obtenção do valor de capital alocado para risco operacional, os bancos deverão calcular o VaR, com horizonte de um ano, para um intervalo de confiança de 99,9%.

Em termos práticos, este VaR corresponde ao quantil 99,9% da distribuição de probabilidade das perdas operacionais de uma unidade de risco, que é determinada pela agregação das distribuições de probabilidade da severidade e da frequência desta determinada unidade de risco com horizonte de um ano.

Esta agregação baseia-se em modelos da área de ciências atuariais, mas não pode simplesmente copiá-los, pois tem suas particularidades, como viés nas amostras, e baixa qualidade e quantidade de dados. Entretanto, o custo benefício de tais modelos necessita ser avaliado. Um bom exemplo de processo de escolha da complexidade adequada para modelos é o modelo IRB (Internal Rating Based) para o risco de crédito proposto pela Basileia, que começa com uma versão sofisticada, e vai relaxando-a até torná-la aceitável, implementável e um *proxy* correto para o capital requerido (Frachot et al., 2003) [44].

O VaR é uma medida já utilizada por outros tipos de risco, principalmente pelo risco de mercado. Todavia, existem alguns aspectos que diferenciam os

modelos de VaR para risco mercado, dos modelos de VaR para risco operacional, chamados de OpVaR:

- O processo estocástico gerador das perdas operacionais quase nunca determina uma distribuição normal subjacente, ao contrário da hipótese Gaussiana geralmente assumida pelo modelo de VaR para risco de mercado⁶⁷ (Cruz, 2002) [26].
- O modelo de VaR para o risco de mercado não utiliza a frequência dos eventos, pois os preços seguem um processo estocástico contínuo (geralmente movimento Browniano), isto é, assume-se que sempre existe uma cotação de preço disponível para um papel, enquanto o mercado estiver aberto. Logo, não é possível contar a quantidade de eventos em um período para o risco de mercado, o que não ocorre com as perdas operacionais. Assim as quantidades de perdas operacionais em períodos pré-determinados seguem um processo estocástico discreto, tais como: Poisson, Poisson misturado, Cox etc (Cruz, 2002) [26].
- Tendo em vista a hipótese de continuidade dos preços para o risco de mercado, os dados são coletados somente quando há mudanças no valor MTM (*mark-to-market*), diferentemente das perdas operacionais que são coletadas quando observadas (Cruz, 2002) [26].

4.1.1.

Distribuição de Perdas Operacionais

Em ciências atuariais existem dois modelos de risco básicos para estimar o total de perdas para um período: modelos de risco individual ou coletivo. Conforme mencionado, estes modelos⁶⁸ também podem ser utilizados para risco operacional, bastando efetuar pequenas adaptações.

Resumidamente, a agregação das perdas em um período pode ser escrita como a soma (S) de um número (N) de perdas individuais (X_1, \dots, X_N), onde cada X_j , $j=1, \dots, N$, é uma variável aleatória contínua que representa a severidade (impacto financeiro) de cada j-ésima perda dentro do período determinado.

⁶⁷ Mais recentemente, a hipótese Gaussiana tem sido bastante questionada, inclusive para o risco de mercado, e técnicas de TVE (teoria dos valores extremos) são utilizadas aonde fatos estilizados indicam que a distribuição de probabilidade da população tem caudas pesadas.

⁶⁸ Mais detalhes podem ser vistos em (Klugman et al., 1998) [58].

O modelo de risco coletivo assume que as perdas X_j são variáveis aleatórias iid (independentes e identicamente distribuídas). Neste modelo, N (frequência) é considerado uma variável aleatória independente de cada X_j (severidade).

Já o modelo de risco individual caracteriza-se por um número fixo de perdas em um período, isto é, N é uma constante. Além disso, assume-se que as perdas X_j são variáveis aleatórias independentes, porém não necessariamente identicamente distribuídas, como no modelo de risco coletivo. Este modelo é comumente usado em seguros para a soma das perdas de um contrato em particular. Em risco operacional este tipo de modelo pode ser usado para um número fixo de eventos em uma unidade de risco – por linha de negócio e tipo de evento de perda (Cruz, 2002; Klugman et al., 1998) [26,58].

Especificamente em relação ao LDA, seguem alguns conceitos necessários à formulação matemática da distribuição de perdas operacionais:

- Os índices i e j correspondem respectivamente a uma determinada linha de negócio e um tipo de evento de perda, conforme preconiza o Novo Acordo da Basileia.
- $X_k(i,j)$ é uma variável aleatória contínua que representa a severidade (impacto financeiro) da k -ésima perda operacional para a linha de negócio i e para o tipo de evento de perda j .
- $N(i,j)$ é uma variável aleatória discreta que corresponde ao número de eventos de perda para linha de negócio i e para o tipo de evento de perda j .
- A função de probabilidade da frequência $N(i,j)$ é chamada de $p_{i,j}$, onde $p_{i,j}(n)=\Pr(N(i,j)=n)$.
- O índice k varia de 0 até $N(i,j)$.
- $N(i,j)$ e $X_k(i,j)$ dependem do período de interesse, geralmente esse período corresponde ao horizonte de um para o cálculo do VaR.

Assim, a perda operacional total para uma linha de negócio i e para um tipo de evento de perda j em um determinado período é:

$$S(i, j) = \sum_{k=0}^{N(i,j)} X_k(i, j) \quad (1)$$

Com base na equação (1), no Teorema de Bayes⁶⁹, no fato dos eventos $N(i,j)=n$ serem mutuamente excludentes para todo n ⁷⁰, e sendo $G_{i,j}$ a função de distribuição de probabilidade acumulada da variável aleatória $S(i,j)$, tem-se:

$$G_{i,j}(x) = \Pr(S(i, j) \leq x) = \sum_{n=0}^{\infty} \Pr(S(i, j) \leq x | N(i, j) = n) \times p_{i,j}(n) \quad (2)$$

Portanto, $G_{i,j}$ é uma média da distribuição de probabilidade acumulada da variável aleatória $S(i,j)$ condicionada a $N(i,j)=n$, ponderada por $p_{i,j}(n)$. Logo quanto maior a probabilidade de $N(i,j)=n_a$, maior a parcela de contribuição da distribuição de probabilidade acumulada da variável $S(i,j)$ condicionada a $N(i,j)=n_a$ para $G_{i,j}$, onde n_a varia de 0 a +8. Esta observação pode ser muito útil para aproximações e otimizações computacionais.

Todavia, é extremamente difícil fazer inferências sobre o modelo descrito em (2), sem que algumas hipóteses sejam arbitradas. Sendo assim, considera-se $S(i, j)$ um modelo de risco coletivo com algumas adaptações. Estas hipóteses são descritas abaixo, definindo $S(i,j)$ como um processo estocástico de contagem composta (Nyström & Skoglund, 2002) [68]:

- X_1, X_2, \dots, X_n são variáveis aleatórias iid condicionadas a $N=n$.
- A distribuição das variáveis aleatórias X_1, X_2, \dots, X_n independe de n , condicionadas a $N=n$.
- N e X_k são independentes para todo o k .
- X_k é uma variável aleatória não-negativa.
- Todos os itens comentados acima valem para cada linha de negócio i e tipo de evento de perda j .

Das hipóteses acima, conclui-se que X_1, X_2, \dots, X_n são variáveis aleatórias iid, unificadas pela notação $X(i,j)$, onde a função de distribuição de probabilidade acumulada desta severidade é chamada de $F_{i,j}$ com domínio real não-negativo. Por exemplo, Log-Normal, Weibull, Gamma, Pareto etc.

Ainda de acordo com as hipóteses mencionadas, no apêndice A é demonstrado que $\Pr(S(i, j) \leq x | N(i, j) = n)$ é igual a 1 quando $x = 0$, e igual a

⁶⁹ Teorema de Bayes: sejam A e B eventos pertencentes a um conjunto de Borel,

$$\Pr(A) = \Pr(A | B) \times \Pr(B) \quad (3)$$

⁷⁰ Seja B um conjunto de Borel composto por eventos mutuamente excludentes B_n (partições), onde n varia de 0 até +8, logo:

$$\Pr(B) = \sum_{n=0}^{+\infty} \Pr(B_n) = 1 \quad (4)$$

$F_{i,j}^{n*}(x)$ quando $x > 0$, onde $*$ é o operador convolução⁷¹ sobre funções de distribuição, e $F_{i,j}^{n*}(x)$ é a n-ésima auto-convolução⁷² de $F_{i,j}(x)$.

Portanto, a distribuição composta $G_{i,j}$ tem a seguinte expressão⁷³:

$$G_{i,j}(x) = \begin{cases} \sum_{n=1}^{+\infty} p_{i,j}(n) F_{i,j}^{n*}(x) & x > 0 \\ p_{i,j}(0) & x = 0 \end{cases} \quad (5)$$

Assim a modelagem da distribuição de perda agregada $G_{i,j}$ para o LDA é feita em duas etapas. Em primeiro lugar, encontra-se a distribuição da severidade e da frequência das perdas. Em seguida, obtém-se a distribuição de perda agregada compondo-as.

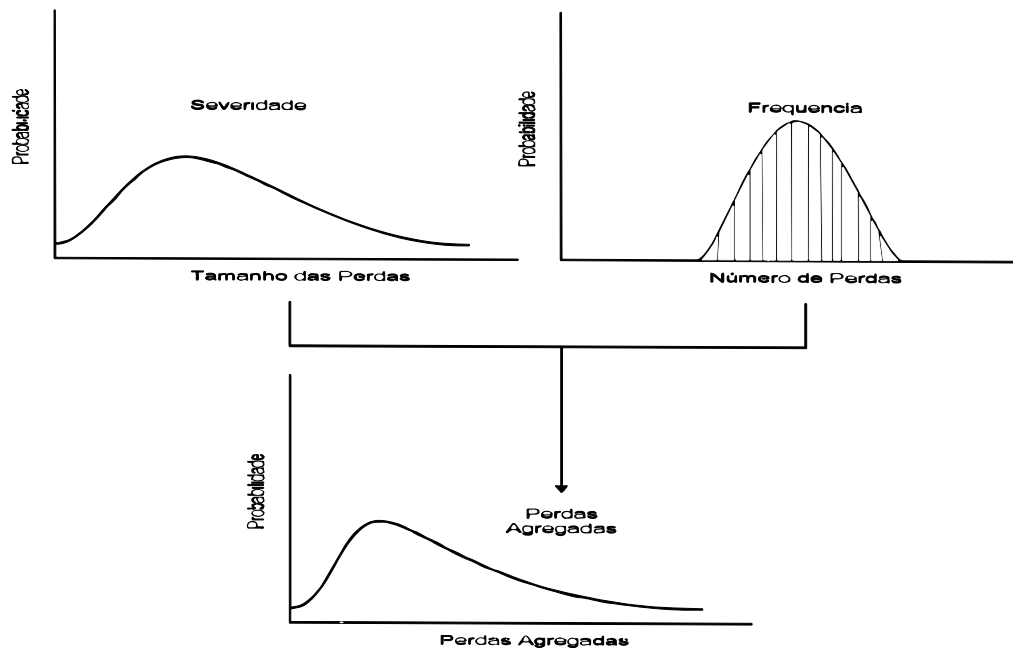


Figura 2 – Modelo de agregação da severidade e da frequência.

⁷¹ Segundo Feller (1971) [40]: a convolução de uma função limitada pontual φ com uma função de distribuição de probabilidade acumulada F é a função em x definida por:

$$u(x) = \int_{-\infty}^{+\infty} \varphi(x - y) dF(y) \quad (6)$$

A convolução é representada por $u = F * \varphi$. Quando F tem uma função de densidade de probabilidade f pode-se alternativamente escrever $u = f * \varphi$, e tem-se:

$$u(x) = \int_{-\infty}^{+\infty} \varphi(x - y) f(y) dy \quad (7)$$

⁷² Estas relações são importantes para o cálculo da auto-convolução:

$$\begin{aligned} F^{1*} &= F \\ F^{n*} &= F^{(n-1)*} * F \end{aligned} \quad (8)$$

⁷³ É importante destacar que quando $n=0$, não existe x definido, e quando $x=0$, $n=0$, pois por hipótese a variável aleatória da severidade é não-negativa.

Entretanto, geralmente não existe expressão analítica para distribuição composta apresentada em (5). Nestes casos são utilizados métodos numéricos implementados em computadores ou métodos analíticos aproximados (Frachot et al., 2001; Nyström & Skoglund, 2002) [42,68].

Conforme mencionado anteriormente, o maior objetivo da determinação da distribuição composta é encontrar o seu VaR. Todavia, a determinação de VaR com métodos numéricos altamente acurados é difícil, ou pelo menos demanda muito tempo de processamento. Isto se deve a típica baixa frequência de eventos e a alta variância das distribuições de severidade. Logo, uma simulação da distribuição composta eficiente requer uma quantidade enorme de números aleatórios (Nyström & Skoglund, 2002) [68].

Assim soluções analíticas fechadas para a distribuição composta, ou aproximações dessas soluções de forma rápida e acurada são desejadas, não só para reduzir o tempo de processamento, mas também para viabilizar análise da sensibilidade dos parâmetros (Nyström & Skoglund, 2002) [68].

4.1.1.1.

Métodos de Determinação da Distribuição de Perdas Operacionais

Todos os métodos, que serão vistos adiante, se baseiam em uma ou mais fórmulas apresentadas abaixo. Estas fórmulas derivam dos conceitos matemáticos de convolução, das equações (1) e (5), e das hipóteses assumidas pelo modelo LDA.

Das equações (6) e (8):

$$F^{*k}(x) = \int_{-\infty}^{+\infty} F^{*(k-1)}(x-y)dF(y) \quad (9)$$

Como X é por hipótese uma variável aleatória contínua não-negativa, pelas equações (7) e (9) tem-se:

$$F^{*k}(x) = \int_0^{+\infty} F^{*(k-1)}(x-y)f(y)dy \quad (10)$$

Diferenciando a equação (10) obtém-se f, que é a função de densidade de probabilidade (f.d.p) de X:

$$f^{*k}(x) = \int_0^{+\infty} f^{*(k-1)}(x-y)f(y)dy \quad (11)$$

Caso a variável aleatória X seja discreta com probabilidades definidas no conjunto dos naturais, as equações (10) e (11) se reduzem a:

$$F^{*k}(x) = \sum_{y=0}^x F^{*(k-1)}(x-y)f(y), \quad x \in \mathbb{N} \quad (12)$$

$$f^{*k}(x) = \sum_{y=0}^x f^{*(k-1)}(x-y)f(y), \quad x \in \mathbb{N} \quad (13)$$

A variável aleatória S tem como f.d.p a função g , que é obtida por diferenciação a partir da equação (5):

$$g(x) = \sum_{n=0}^{+\infty} p(n)f^{*n}(x) \quad (14)$$

As funções geradoras de probabilidade (f.g.p) das variáveis aleatórias S , N e X , denominadas respectivamente P_S , P_N e P_X , seguem a relação:

$$P_S(z) = P_N[P_X(z)] \quad (15)$$

Já as funções características⁷⁴ (f.c) das variáveis aleatórias S e X , denominadas respectivamente f_S e f_X , seguem a relação:

$$j_S(z) = P_N[j_X(z)] \quad (16)$$

Panjer & Willmot (1992) [71] utilizam a transformada de Laplace das variáveis aleatórias⁷⁵ S e X , denominadas respectivamente L_S e L_X :

$$L_S(z) = P_N[L_X(z)] \quad (17)$$

As funções geradoras dos momentos (f.g.m) das variáveis aleatórias S e X , denominadas respectivamente M_S e M_X , seguem a relação:

$$M_S(z) = P_N[M_X(z)] \quad (18)$$

Da equação (15) encontram-se relações entre momentos⁷⁶ de S , N e X :

$$E(S) = E(N)E(X) \quad (19)$$

$$Var(S) = E(N)Var(X) + Var(N)[E(X)]^2 \quad (20)$$

$$Ass(S) = E(N)Ass(X) + 3Var(N)E(X)Var(X) + Ass(N)[E(X)]^3 \quad (21)$$

4.1.1.1.1.

Método Analítico – Solução Fechada

Apesar da maioria das combinações de distribuições de severidade e de frequência só apresentar soluções aproximadas ou numéricas, há algumas

⁷⁴ Nem todas as funções geradoras existem para todas as variáveis aleatórias. Exceção feita à função característica.

⁷⁵ Sempre existe para variáveis aleatórias não-negativas, que é justamente o caso conforme hipótese do modelo.

⁷⁶ Seja a variável aleatória Y : $Var(Y)=E\{[Y - E(Y)]^2\}$ e $Ass(Y)=E\{[Y - E(Y)]^3\}$.

combinações aonde a solução analítica da distribuição composta apresentada na equação (5) é possível, como por exemplo:

- $X \sim \text{Exp}(\mathbf{q}), N \sim \text{Geom}(\mathbf{b})$

$$\circ g(x) = \begin{cases} \frac{1}{1+\mathbf{b}}, & x=0 \\ \frac{\mathbf{b}}{\mathbf{q}(1+\mathbf{b})^2} \exp\left[-\frac{x}{\mathbf{q}(1+\mathbf{b})}\right], & x>0 \end{cases}$$

$$\circ G(x) = 1 - \frac{\mathbf{b}}{1+\mathbf{b}} \exp\left[-\frac{x}{\mathbf{q}(1+\mathbf{b})}\right], \quad x \geq 0$$

- $X \sim \text{Exp}(\mathbf{q}), N \sim \text{Qualquer}$

$$\circ G(x) = 1 - e^{-x/\mathbf{q}} \sum_{j=0}^{+\infty} \bar{P}_j \frac{\left(\frac{x}{\mathbf{q}}\right)^j}{j!}, \quad x \geq 0$$

$$\circ \bar{P}_j = \sum_{n=j+1}^{+\infty} p(n), \quad j \in \mathbf{N}$$

- $X \sim \text{Exp}(\mathbf{q}), N \sim \text{NBin}(r, \mathbf{b}) \quad r \in \mathbf{Z}$

$$\circ G(x) = 1 - \sum_{n=1}^r \binom{r}{n} \left(\frac{\mathbf{b}}{1+\mathbf{b}}\right)^n \left(\frac{1}{1+\mathbf{b}}\right)^{r-n} \times H_n \quad x \geq 0$$

$$\circ H_n = \sum_{j=0}^{n-1} \frac{\{x\mathbf{q}^{-1}(1+\mathbf{b})^{-1}\}^j e^{-x\mathbf{q}^{-1}(1+\mathbf{b})^{-1}}}{j!}$$

Normalmente utiliza-se a equação (18) na busca por uma solução fechada para equação (5).

O procedimento consiste em:

1. Calcular M_X e P_N . Geralmente estas funções estão tabeladas em livros de estatística para as principais distribuições.
2. Determinar M_S com base na equação (18).
3. Fazer uma analogia de M_S com outras f.g.m tabeladas objetivando encontrar uma distribuição conhecida cujos parâmetros são funções dos parâmetros das distribuições de X e N .

Vantagens:

- Resultado rápido e exato.
- Possibilita análise acurada da sensibilidade dos parâmetros.

Desvantagens:

- Normalmente não existe solução fechada para várias composições de distribuições de severidade e de freqüência.

4.1.1.1.2.

Método da Distribuição Aproximada

Este método consiste em aproximar a distribuição composta para alguma distribuição conhecida, e em seguida aplicar alguma técnica de estimação de parâmetros, como por exemplo: máxima verossimilhança, mínimos quadrados ordinários, método dos momentos etc. Com os parâmetros estimados tem-se a distribuição composta aproximada.

O método de estimação mais utilizado é o dos momentos pela facilidade de emprego. Portanto, as equações (19), (20) e (21) são fundamentais nesta técnica.

Quando $E(N)$ é muito grande, a distribuição Normal é uma boa aproximação. Em particular, se N tiver uma distribuição Poisson, Binomial ou Binomial Negativa, uma versão do teorema central do limite indica que os parâmetros μ , m ou r respectivamente tendem ao infinito, e a distribuição de S torna-se Normal.

Vantagens:

- Simplicidade e facilidade de uso.
- Não há necessidade de se conhecer as distribuições de severidade e de frequência.
- Evita o cálculo direto da equação (5) por intermédio da equação (10). Geralmente este cálculo é muito dispendioso em termos de processamento computacional.

Desvantagens:

- Não existe meio de se saber o quão boa é a aproximação. Distribuições aproximadas distintas podem apresentar resultados muito diferentes, particularmente na cauda das distribuições.
- A distribuição aproximada pode falhar na acomodação de características especiais da distribuição verdadeira. Por exemplo, se a variável aleatória aproximada de S for contínua, e probabilidade para determinados valores da variável aleatória verdadeira de S forem diferentes de zero.

4.1.1.1.3.

Método do Cálculo Direto

Dado que se conhecem as distribuições de severidade e de frequência, este método baseia-se na determinação das auto-convoluções da distribuição de severidade para $n = 2, 3, 4, \dots$. Após isso, aplica-se a equação (5).

Estas auto-convoluções podem ser calculadas numericamente por intermédio da equação (10). Todavia, embora X seja não-negativa, pode ser uma variável aleatória mista (contínua e discreta). Logo, a equação (10) é re-escrita na forma Lebesgue-Stieltjes⁷⁷:

$$F^{*k}(x) = \int_{-0}^x F^{*(k-1)}(x-y)f(y)dy \quad (22)$$

O cálculo da equação (22) requer métodos numéricos de integração. Como o primeiro termo da integral depende de x , a equação (22) precisa ser calculada para todos os possíveis valores de x , para cada valor de k , respeitando a recursividade. Isto torna esta forma de cálculo tecnicamente impossível, pois x pertence ao conjunto dos reais não-negativos.

Uma forma simples de contornar este problema é substituir a distribuição da severidade por uma distribuição discreta⁷⁸ definida em múltiplos $0, 1, 2, 3, \dots$ de alguma unidade monetária conveniente, como por exemplo 1000. Assim, podem-se utilizar as equações (12) e (13), ao invés da equação (22).

Quanto menor a unidade monetária, mais a distribuição discreta se aproxima da verdadeira distribuição. Entretanto, maior o número de possíveis valores de x , o que aumenta o esforço computacional.

Apesar de tornar viável o cálculo das auto-convoluções, este método requer um número de multiplicações da ordem n^3 para obtenção de valores da distribuição de S para $x = 0$ até $x = n$, isto se $\Pr(X = 0) = 0$.

Se $\Pr(X = 0) > 0$, é necessário um número infinito de cálculos para obtenção de uma única convolução.

Vantagens:

- Utiliza a distribuição de frequência, e a equação (5) sem nenhum artifício de aproximação.

⁷⁷ $\int g(y)f(y)dy$ é calculado por integração de $g(y)f(y)$ sobre os valores de y aonde X é uma variável contínua, somando-se a $g(y_i)\Pr(X = y_i)$ para os valores y_i aonde $\Pr(X = y_i) > 0$, isto é, valores de y para os quais X é discreta. Esta forma de integração pode então ser usada para X contínua, discreta ou mista.

⁷⁸ Os métodos de discretização são vistos no anexo B.

Desvantagens:

- Necessita que a distribuição de severidade seja discretizada.
- Para garantir uma distribuição de S acurada, a unidade monetária deve ser a menor possível. Em contra partida, a quantidade de possíveis valores de x cresce. Isso aumenta o maior valor de N, denominado n. Como o número de multiplicações neste método é da ordem n^3 , pode-se tornar inviável computacionalmente obter uma distribuição de S acurada.
- Se $\Pr(X = 0) > 0$ não é possível encontrar solução por este método.

4.1.1.1.4.

Método Recursivo

Este método restringe o universo de distribuições de freqüência à família classe $(a, b, 1)^{79}$, que satisfaz à relação:

$$p(k) = \left(a + \frac{b}{k}\right)p(k-1), \quad k = 2, 3, 4, K \quad (23)$$

Sendo X uma variável aleatória contínua não-negativa, a p.d.f de S é:

$$g(x) = p(1)f(x) + \int_0^x \left(a + \frac{by}{x}\right)f(y)g(x-y)dy \quad (24)$$

A equação (24) representa uma integral de Volterra de segundo tipo. Solução numérica para esta integral é encontrada em Baker (1977) [13].

Portanto, para fugir da aplicação do método de Baker, que é bastante complicado, X é discretizada⁸⁰, tal que f(x) seja definida em 0, 1, 2, ..., r, sendo estes números múltiplos de alguma unidade monetária conveniente. O número r representa a maior perda possível, e pode até ser infinito.

Assim, a p.d.f de S reduz-se a:

$$g(x) = \frac{[p(1) - (a+b)p(0)]f(x) + \sum_{y=1}^{\min(x,r)} \left(a + \frac{by}{x}\right)f(y)g(x-y)}{1 - af(0)} \quad (25)$$

Sendo a distribuição de freqüência pertencente à subclasse $(a, b, 0)^{81}$ da classe $(a, b, 1)$, tem-se:

⁷⁹ Ver anexo D.

⁸⁰ Os métodos de discretização são vistos no anexo B.

⁸¹ Ver anexo C.

$$p(k) = \left(a + \frac{b}{k}\right)p(k-1), \quad k = 1, 2, 3, 4, K \quad (26)$$

Neste caso a p.d.f de S reduz-se a:

$$g(x) = \frac{\sum_{y=1}^{\min(x,r)} \left(a + \frac{by}{x}\right) f(y) g(x-y)}{1 - af(0)} \quad (27)$$

O valor inicial para as equações recursivas (25) e (27) é:

$$g(0) = P_N [f(0)] \quad (28)$$

O número de multiplicações para obtenção de valores da distribuição de S para $x = 0$ até $x = n$ é de ordem n^2 .

Sundt & Jewell (1981) [83] mostraram que a classe $(a, b, 0)$ somente possui 4 distribuições possíveis: Poisson, Binomial, Negativa Binomial e Geométrica⁸².

Como estas são as distribuições de frequência mais utilizadas para o LDA, a restrição imposta pelo método não causa perda significativa.

Vantagens:

- Reduz o tempo de processamento em relação ao método do cálculo direto.
- Apresenta valores exatos caso a distribuição de severidade seja discreta.
- A única fonte de erro está na discretização da distribuição de severidade.
- Exceto para a distribuição de frequência Binomial, os cálculos são numericamente estáveis (Klugman et al., 1998) [58].
- Método de fácil implementação computacional.

Desvantagens:

- O método só funciona para as distribuições de frequência pertencentes à classe $(a, b, 1)$ ou à classe $(a, b, 0)$.
- Produz a distribuição agregada inteira, não sendo possível determinar a probabilidade para um único valor de S antes que todos os cálculos sejam feitos.

⁸² Informações mais detalhadas sobre estas distribuições são apresentadas no anexo C.

4.1.1.1.5.

Métodos de Inversão

O objetivo dos métodos de inversão é obter numericamente a p.d.f por intermédio de uma expressão conhecida de uma transformada de uma função desejada, tal como f.g.p, f.g.m ou f.c.

Esta estratégia de solução é natural para as distribuições perdas agregadas, uma vez que suas transformadas são facilmente compostas, conforme visto nas equações de (15) a (18), dado que as expressões das distribuições de severidade e frequência são conhecidas.

A seguir será mostrado um método⁸³ cuja transformada objeto é a f.c (função característica). Isto se deve ao fato de que uma distribuição sempre tem uma única f.c. Por outro lado, para uma dada f.c, sempre existe uma única distribuição.

4.1.1.1.5.1.

Transformada Rápida de Fourier – *Fast Fourier Transform (FFT)*

A transformada de Fourier de uma função contínua $f(x)$ qualquer é:

$$\tilde{f}(z) = \int_{-\infty}^{+\infty} f(x)e^{izx} dx \quad (29)$$

A transformada inversa que recupera a função original é:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{f}(z)e^{-izx} dz \quad (30)$$

Quando $f(x)$ é uma p.d.f, a sua transformada de Fourier representa a sua função característica (f.c).

Se f_x for uma função periódica de tamanho n ($f_{x+n} = f_x$) definida para todos os valores x pertencente ao conjunto dos Inteiros, a transformada discreta de Fourier (*Discrete Fourier Transform – DFT*) de f_x é:

$$\tilde{f}_k = \sum_{j=0}^{n-1} f_j \exp\left(\frac{2\pi i}{n} jk\right), \quad k = K - 1, 0, 1, K \quad (31)$$

⁸³ Em Klugman et al (1998) [58] e Robertson (1992) [80] encontram-se outro exemplo de método de inversão baseado em f.c. Esse método utiliza algoritmos de Heckman e Meyers para aproximação de integral numérica. A grande vantagem deste método é a possibilidade determinar a probabilidade para um único valor de S sem precisar levantar a distribuição agregada inteira. Todavia, este benefício não é tão importante para o cálculo do VaR, e portanto, este método não será apresentado neste trabalho.

A DFT de f_x também é uma função periódica de tamanho n , cuja transformada inversa que recupera a função original é:

$$f_j = \frac{1}{n} \sum_{k=0}^{n-1} \tilde{f}_k \exp\left(-\frac{2\pi i}{n} jk\right), \quad j = K - 1, 0, 1, K \quad (32)$$

Pela equação (31) observa-se que para obter n valores para a DFT de f_x , o número de termos a serem calculados é da ordem n^2 .

Todavia, o algoritmo FFT (*Fast Fourier Transform*) reduz o número de cálculos à ordem $n \log_2 n$.

Este algoritmo baseia-se na relação:

$$\tilde{f}_k = \sum_{j=0}^{m-1} f_{2j} \exp\left(\frac{2\pi i}{m} jk\right) + \exp\left(\frac{2\pi i}{n} k\right) \sum_{j=0}^{m-1} f_{2j+1} \exp\left(\frac{2\pi i}{m} jk\right), \quad m = \frac{n}{2} \quad (33)$$

Logo, a equação (33) pode ser re-escrita de uma forma recursiva:

$$\tilde{f}_k = \tilde{f}_k^a + \exp\left(\frac{2\pi i}{n} k\right) \tilde{f}_k^b \quad (34)$$

Se $n = 2^r$, onde r é um número inteiro positivo, pode-se dividir a transformada r vezes até encontrarem transformadas de tamanho 1. O conhecimento das transformadas de tamanho 1 permite composições sucessivas de transformadas de tamanho 2, 2^2 , ..., 2^r por simples adição, usando-se a equação (34).

Assim, o método segue o procedimento apresentado abaixo:

1. Discretizar a distribuição de severidade usando algum método apresentado no anexo B, obtendo um vetor de valores que representa a distribuição discretizada da severidade: $f_x(0), f_x(1), \dots, f_x(n-1)$, onde $n = 2^r$ para r inteiro positivo, e n o número de pontos desejados para a distribuição agregada g .
2. Aplicar o algoritmo FFT ao vetor de valores, obtendo a f.c. da distribuição de severidade discretizada. O resultado é um vetor complexo com $n = 2^r$ valores.
3. Transformar o vetor complexo na f.c. de g por intermédio da equação (16). Esta f.c. também é um vetor complexo com $n = 2^r$ valores.
4. Aplicar o algoritmo IFFT⁸⁴ (*Inverse Fast Fourier Transform*) no vetor complexo resultante do passo anterior. Este novo vetor é real, possui $n = 2^r$ valores, e representa a distribuição agregada discretizada: $g_s(0), g_s(1), \dots, g_s(n-1)$.

⁸⁴ Este algoritmo utiliza o mesmo princípio de divisões sucessivas do algoritmo FFT. Portanto, equação similar a (34) pode ser obtida partindo-se da equação (32).

Algumas observações referentes ao método FFT:

- Quando o número de pontos da distribuição de severidade discretizada for menor que $n = 2^l$, o vetor que representa esta distribuição deve ser completado com “zeros” até que o seu tamanho seja igual a n .
- Quando a distribuição de severidade possui probabilidade para valores além de $x = n$, a premissa de que f_x é periódica de tamanho n introduz erro no cálculo da transformada de Fourier, e por conseguinte, na determinação da distribuição agregada g_S . Uma solução para minimizar este problema é considerar que a probabilidade restante concentra-se em $x = n$, sendo n suficientemente grande.

Vantagens:

- Reduz o tempo de processamento em relação ao método recursivo para n grande.
- O método FFT pode ser estendido para distribuição de severidade com valores negativos.
- Método de fácil implementação computacional, sendo que a maioria dos pacotes de softwares estatísticos dispõe de algoritmos FFT e IFFT.

Desvantagens:

- Produz a distribuição agregada inteira, não sendo possível determinar a probabilidade para um único valor de S antes que todos os cálculos sejam feitos.

4.1.1.1.6.

Método de Simulação

De forma geral, este método segue três passos objetivando determinar a distribuição de S :

1. Construir um modelo para S em função das variáveis aleatórias $N, X_1, \dots, X_N, Y, Z, \dots$, onde suas distribuições e dependências são conhecidas.
2. Para $i = 1, \dots, m$ gerar valores pseudo-aleatórios $n_i, x_{1i}, \dots, x_{Ni}, y_i, z_i, \dots$, e então calcular s_i usando o modelo elaborado no passo 1.
3. A distribuição de S é aproximada à distribuição empírica baseada na amostra pseudo-aleatória s_1, \dots, s_m .

Vantagens:

- O erro causado pela simulação pode ser tão pequeno quanto se queira, bastando para isso, entre outras coisas, aumentar o número de cenários.
- O método não é restrito as hipóteses de independência das variáveis aleatórias N, X_1, \dots, X_N e distribuições idênticas das variáveis $X_{j|s}$. Isso torna a medida do risco operacional mais fiel ao controle executado pela gerência de risco. Por exemplo, caso exista uma política de risco que estipule um limite para as operações bancárias atrelado às perdas acumuladas, a hipótese de independência das variáveis aleatórias N, X_1, \dots, X_N cai por terra. Provavelmente, esta hipótese geralmente é levada em consideração por ser mais conservadora. Todavia, a AMA permite que modelos internos sejam utilizados desde que sua consistência seja aprovada.
- Facilidade de implementação, inclusive por intermédio de planilhas eletrônicas largamente utilizadas por usuários sem muita experiência em programação.
- Muitas planilhas ou pacotes estatísticos dispõem de geradores de número pseudo-aleatórios⁸⁵ para várias distribuições de probabilidade.

Desvantagens:

- Produz a distribuição agregada inteira, não sendo possível determinar a probabilidade para um único valor de S antes que todos os cálculos sejam feitos.
- Dependendo da capacidade de processamento do computador e da quantidade de cenários simulados o processo por ficar extremamente lento.

4.1.2.

Cálculo do Capital Regulatório (OPVaR)

No modelo LDA, o capital regulatório é uma medida de VaR chamada OPVaR (*Operational Value at Risk*). O VaR operacional com horizonte de um ano e um intervalo de confiança q (99,9% pelo Acordo da Basileia) é definido como:

$$OPVaR(q) = G^{-1}(q) = \inf \{z \mid G(z) \geq q\} \quad (35)$$

⁸⁵ Para encontrar informações mais detalhas sobre geradores de número pseudo-aleatórios consulte o apêndice H de Klugman et al. (1998) [58].

, onde $G(z)$ é a distribuição de perdas operacionais prevista para o próximo ano, e é determinada por intermédio de algum dos métodos vistos anteriormente.

Vale ressaltar que aqueles métodos executam a agregação das distribuições de severidade e frequência em um ano. Portanto, pode ser ingênuo afirmar que a distribuição resultante seja a mesma para o próximo ano. Assim, a AMA permite que métodos de “envelhecimento” da carteira de riscos operacionais também sejam utilizados. Considerar-se-á que $G(z)$ já é a distribuição de perdas operacionais prevista para o próximo ano.

Para se entender a medida OPVaR é importante conhecer como as instituições financeiras classificam suas perdas com base na distribuição G , e como elas se protegem dessas perdas.

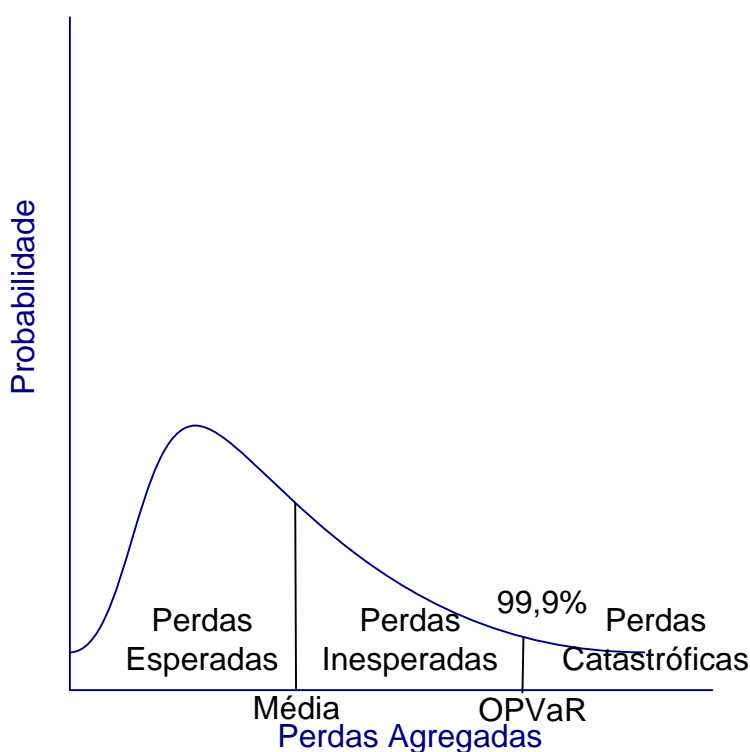


Figura 3 – Classificação das Perdas.

Perdas esperadas: são perdas até a perda esperada, que é a média⁸⁶ da distribuição G . As instituições financeiras absorvem estas perdas com o lucro proveniente das receitas. Visando não diminuir demais o lucro, divide-se a perda

⁸⁶ Segundo Coleman (2003b) [24], deve-se utilizar a mediana ao invés da média para determinação da perda esperada.

esperada pelas transações, e embute-se este valor na taxa administrativa de cada transação.

Perdas Inesperadas: estão entre a média e o OPVaR. Estas perdas são cobertas pelo capital regulatório destinado à proteção contra o risco operacional (OPVaR).

Perdas Catastróficas: estão além do OPVaR, e necessitam de capital extra ou seguro visando a proteção.

Primeiro será apresentado o cálculo do OPVaR para uma linha de negócio e um tipo de evento, e depois para o banco como um todo.

4.1.2.1.

Para uma linha de negócio (i) e um tipo de evento (j)

A perda esperada $EL(i, j)$, e a perda inesperada $UL(i, j; q)$ para um intervalo de confiança q são definidas como:

$$EL(i, j) = E[S(i, j)] = \int_0^{+\infty} x dG_{i,j}(x) \quad (36)$$

$$UL(i, j; q) = G_{i,j}^{-1}(q) - E[S(i, j)] = \inf \{x \mid G_{i,j}(x) \geq q\} - \int_0^{+\infty} x dG_{i,j}(x) \quad (37)$$

O Novo Acordo da Basileia [12] propõe que o OPVaR seja igual a soma da perda inesperada com a perda esperada:

$$OPVaR(i, j; q) = UL(i, j; q) + EL(i, j) \quad (38)$$

Isso se justifica, pois de (37) tem-se:

$$UL(i, j; q) + EL(i, j) = G_{i,j}^{-1}(q) \quad (39)$$

Baseando-se nas hipóteses apresentadas no item 4.1.1, $EL(i, j)$ é calculado sem a necessidade de se determinar a distribuição agregada G .

$$E[S(i, j)] = E[E[S(i, j) \mid N(i, j)]] = E[X(i, j)] \times E[N(i, j)] \quad (40)$$

Já para o cálculo de $UL(i, j; q)$ é necessária a determinação de G , cujo quantil estimado pode apresentar um viés muito grande em relação ao quantil verdadeiro, especialmente quando a distribuição de severidade tem caudas pesadas, e o intervalo de confiança (q) é muito elevado. Por isso, é muito importante controlar a acurácia do OPVaR, e isto pode ser feito confrontando os

quatro primeiros momentos estimados de G com seus valores teóricos⁸⁷ (Frachot et al., 2001) [42].

4.1.2.2.

Para o banco como um todo

Como o LDA é uma abordagem dentro da AMA que possibilita uma alocação de capital econômico seguindo o conceito “*bottom-up*”, o OPVaR para o banco como um todo é composto pelos OPVaR’s de todas as combinações possíveis de linhas de negócio e tipos de eventos de risco operacional.

Esta composição é função da estrutura de dependência entre as unidades de risco (linha de negócio e tipo de evento).

Uma solução mais apropriada e geral para representar estruturas de dependência é a utilização de cópulas. Cópulas são as funções que relacionam a distribuição conjunta de vetores aleatórios com suas distribuições marginais uniformes padrões. Como consequência, cópulas pode ser uma solução para modelar as estruturas de dependências entre unidades de riscos. Existem numerosas famílias de cópulas, e cada uma com suas especificações⁸⁸ (Chapelle et al., 2004) [18].

A seguir serão apresentados alguns casos particulares dessas estruturas de dependência.

4.1.2.2.1.

Método do Somatório

O Comitê da Basileia sugere que o OPVaR para o banco como um todo seja igual a soma dos OPVaR’s para cada linha de negócio e tipo de evento, como é feito na IMA:

$$OPVaR(q) = \sum_{i=1}^I \sum_{j=1}^J OPVaR(i, j; q) \quad (41)$$

Esta sugestão corresponde a assumir que diferentes unidades de risco (par linha de negócio e tipo de risco) são totalmente dependentes positivas, ou

⁸⁷ As equações (19) a (21) apresentam os três primeiros momentos teóricos de G.

⁸⁸ Mais informações sobre cópulas e a sua utilização na determinação da estrutura de dependência das unidades de risco podem ser vista no apêndice B de (Frachot et al., 2001) [42].

melhor, perfeitamente correlatadas⁸⁹. Isto superestima o OPVaR total, forçando o banco reservar mais capital que o necessário.

4.1.2.2.

Método das Unidades de Risco Independentes

Uma suposição mais realista consideraria as unidades de risco independentes⁹⁰. Neste caso, seja S a perda total de um banco, logo:

$$S = \sum_{i=1}^I \sum_{j=1}^J S(i, j) \quad (42)$$

Se $S(i,j)$ independentes, a distribuição do banco como um todo G é a auto-convolução das distribuições $G_{i,j}$:

$$G(x) = \underset{i=1}{\overset{I}{*}} \underset{j=1}{\overset{J}{*}} G_{i,j}(x) \quad (43)$$

Assim, similarmente às equações (38) e (39) tem-se:

$$OPVaR(q) = G^{-1}(q) \quad (44)$$

4.1.2.2.3.

Regra da Raiz Quadrada

A idéia desta regra é encontrar $UL(q)$ diretamente por intermédio das perdas inesperadas individuais $UL(i,j;q)$, sem ter de determinar a distribuição de perdas operacionais do banco como um todo (G).

$$OPVaR(q) = EL + \sqrt{\sum_{i=1}^I \sum_{j=1}^J (OPVaR(i, j; q) - EL(i, j))^2} \quad (45)$$

Observa-se que esta regra corresponde a uma aproximação a uma Normal.

Apesar da fórmula (45) facilitar computacionalmente o cálculo do OPVaR, pode produzir erros significantes.

⁸⁹ Corresponde ao caso aonde a função dependência – ou a cópula – é o limite superior de Fréchet.

⁹⁰ Cópulas produto.

4.1.2.2.4.

Efeito da Diversificação

A medida do efeito da diversificação – composição das unidades de risco operacional para o cálculo do OPVaR – é chamada de razão de diversificação, e é definida por:

$$c(q) = \frac{OPVaR^+(q) - OPVaR(q)}{OPVaR^+(q)} \quad (46)$$

, onde $OPVaR^+(q)$ é dado pela equação (41).

A razão de diversificação varia de 0 a 1.

Portanto, quando esta razão é superestimada, isto é, tende a 1, o OPVaR para o banco como um todo é subestimado, e vice-versa.

4.1.3.

Validação do Modelo de Perdas Operacionais

Conforme visto anteriormente, o OPVaR depende das distribuições de severidade e frequência da(s) unidade(s) de risco(s), do(s) método(s) de agregação e da estrutura de dependência, no caso do OPVaR de um banco com um todo. Uma combinação destas alternativas define um modelo de distribuição de perdas.

Entretanto, para efeito de controle interno e exigência regulatória, cada unidade de risco deve possuir um único modelo de perda operacional.

Além da unicidade, os modelos devem ser avaliados internamente, garantindo a confiança nas medidas apresentadas pelo modelo, e a qualidade dos controles internos e das políticas de risco que derivam destas medidas; e externamente pelos órgãos reguladores, assegurando que os critérios qualitativos e quantitativos estabelecidos pelo Novo Acordo da Basileia estejam sendo observados.

Portanto, técnicas de validação de modelos de perdas operacionais são de extrema importância para seleção e avaliação de modelos.

A validação pode ocorrer em dois momentos:

- Na determinação do modelo. Neste caso, delimita-se um conjunto de perdas operacionais para teste e um outro conjunto para treinamento. O conjunto de treinamento serve para estimação do modelo, e o de

teste serve para a validação do modelo. Exemplos: as técnicas de “*Cross Validation*” e “*k-fold Cross Validation*”.

- Auditoria interna ou externa no modelo em uso. Neste caso, compara-se o previsto (calculado pelo modelo) com o realizado (perda operacional real).

Para a validação dos modelos de risco operacional utilizam-se análises básicas, baseadas simplesmente no número de violações (realizado ou teste excedendo o previsto), e análises estatísticas. Como análises estatísticas destacam-se as mesmas utilizadas para o risco de mercado: teste de Kupiec e o critério de Crnkovic-Drachman.

Outra medida importante é o índice extremal, que verifica o grau de agrupamento de eventos extremos. Este índice varia de 0 a 1, sendo que quanto menor o seu valor, mais agrupada a série temporal. Ele é muito útil principalmente quando a distribuição de severidade é uma distribuição de valor extremo, pois estas distribuições são muito sensíveis ao parâmetro de forma. A observação durante a validação de muitas violações temporalmente juntas pode indicar que o processo estocástico escolhido para determinar a distribuição de severidade não foi o ideal. Além disso, o índice extremal oferece uma alternativa para sobrepujar a deficiência de predição causada por uma forte estrutura de dependência de eventos, identificada por um agrupamento de eventos extremos (Cruz, 2002) [26]. Esta alternativa será vista mais adiante.

A seguir são apresentados de forma sucinta alguns detalhes das técnicas indicadas acima. Em todos os casos, é considerado que as distribuições de severidade e de frequência também passaram por seus critérios de seleção e validação.

4.1.3.1.

Técnica de *Cross Validation*

O método de *holdout* é o mais simples e mais comum quando se utiliza esta técnica.

1. Dividi-se o conjunto total de dados em conjunto de treinamento e conjunto de teste.
2. Ajusta-se o modelo com os dados de treinamento.
3. Utiliza-se o modelo para prever os valores do conjunto de teste.
4. Acumulam-se os erros encontrados, calcula-se a média e o erro absoluto dos erros.

Esta técnica pode apresentar alta variância, pois depende muito do conjunto de testes escolhido (Guimarães, 2003) [46].

4.1.3.2.

Técnica *k-fold Cross Validation*

Esta técnica contorna o problema descrito no item anterior.

1. O conjunto total de dados é dividido k subconjuntos.
2. O método de *holdout* é aplicado em k iterações. Cada iteração tem um conjunto de teste dentre os k existentes, e os demais constituem o conjunto de treinamento. Os erros vão sendo acumulados nas k iterações.
3. Finda as iterações, todos os k conjuntos foram utilizados como conjunto de teste, e estatísticas são realizadas sobre os erros acumulados: média, desvio padrão, erro absoluto etc. Estas estatísticas visam medir a qualidade do modelo.

Esta técnica não é tão dependente de como os subconjuntos são formados, sendo que a variância do resultado estimado reduz-se drasticamente como o aumento de subdivisões k.

Outra vantagem é fato de todos os dados serem utilizados tanto para ajuste do modelo como para teste (Guimarães, 2003) [46].

4.1.3.3.

Teste de Kupiec

Utiliza-se este teste para verificar a probabilidade de aceitar um modelo incorreto ou rejeitar um bom modelo.

Conceito básico do teste: tenta verificar se a diferença entre razão de violação (número de exceções/tamanho da amostra) do modelo e o nível de confiança especificado é estatisticamente não significativa.

Seja:

V = número de violações, isto é, número de perdas acima do VaR

T = número total de observações

p = nível de confiança especificado para o VaR. Exemplo: VaR 99%, p =

1%

$$H_0 : p = p^* \\ LR_{UC} = 2 \ln \left[(1 - p^*)^{T-V} p^{*V} \right] + 2 \ln \left[\left(1 - \frac{T}{V} \right)^{T-V} \left(\frac{V}{T} \right)^V \right] \quad (47)$$

, onde $LR_{UC} \sim \chi^2(1)$ sob H_0 .

4.1.3.4.

Critério de Crnkovic-Drachman

Este critério é utilizado para encontrar uma medida do ajuste das predições aos impactos financeiros reais.

Conceito básico do critério: comparar a função de distribuição de probabilidade (fdp) da predição (g) com a distribuição uniforme.

1. Para cada dia (ou período), verificar que percentil (p) de g corresponde à perda real.
2. Repetir o passo 1 em N dias (períodos) consecutivos.
3. O desvio da distribuição dos percentis em relação à distribuição uniforme é uma primeira medida de qualidade do modelo. Menor o desvio, maior a qualidade.
4. Seja $n(t)$ o número de percentis observados menores ou iguais a t, onde t pertence $[0,1]$. Seja ainda $f(t) = n(t)/N$ e $g(t) = t$. A qualidade do modelo pode ser medida por:

$$K(f(x), g(x)) = \max_{0 \leq x \leq 1} \{f(x) - g(x)\} - \max_{0 \leq x \leq 1} \{g(x) - f(x)\} \quad (48)$$

K é conhecido como estatística de Kuiper, e mede basicamente a distancia entre duas funções de distribuição acumuladas. Então K é a medida de quão longe a distribuição dos percentis está da distribuição uniforme.

4.1.3.5.

Índice Extremal

Existem alguns métodos para estimar o índice extremal, mas serão vistos apenas dois: tamanho médio do agrupamento e blocos.

4.1.3.5.1.**Método do Tamanho Médio do Agrupamento**

Sejam d_1, \dots, d_k observações excedentes, isto é, acima de um limite pré-determinado u . Estas observações estão ordenadas no tempo, sendo t_1, \dots, t_k os respectivos instantes de tempo.

Dado um inteiro positivo r , qualquer conjunto de pelo menos r observações consecutivas abaixo do limite u define dois agrupamentos excedentes. Logo, r é o tamanho mínimo entre dois agrupamentos excedentes consecutivos.

Seja $n(u,r)$ o número de agrupamentos excedentes.

Assim, o tamanho médio do agrupamento excedente é definido como:

$$m(u, r) = \frac{k}{n(u, r)} \quad (49)$$

O índice extremal é o inverso do tamanho médio do agrupamento excedente:

$$IE(u, r) = \frac{1}{m(u, r)} \quad (50)$$

4.1.3.5.2.**Método dos Blocos**

Em primeiro lugar, separa-se a série temporal das diferenças entre os OPVaR's e as perdas operacionais reais em B blocos.

Em seguida, determina-se o limite u .

Sendo assim, seja K o número de blocos com uma ou mais observações excedentes, e N o número total de observações excedentes. O índice extremal estimado é definido como:

$$IE(B, u) = \frac{K}{N} \quad (51)$$

4.1.3.5.3.**Uso Prático do Índice Extremal**

Conforme dito anteriormente, o índice extremal oferece uma alternativa para sobrepujar a deficiência de predição causada por uma forte estrutura de dependência de eventos. A correção na predição pode ser feita pela fórmula:

$$q_F = (q_I)^q \quad (52)$$

, onde q é o nível de confiança inicial, por exemplo 95%, q_f é o nível de confiança final, isto é, o ajustado para correção na predição, e α é o índice extremal.

Cruz (2002) [26] ressalta que: como quanto menor o valor de α mais agrupado os valores extremos, a correção na predição é obtida com valores de nível de confiança maiores, de acordo com a equação (52).

4.1.4.

Alocação do Capital Econômico

Após a validação do modelo, o cálculo do OPVaR para o banco como um todo ganha credibilidade.

Entretanto, esta medida por si só não adianta muita coisa. Além de servir para o mecanismo de controle, peça fundamental no gerenciamento de risco, o principal objetivo do OPVaR é subsidiar a alocação econômica de capital por linha de negócio e tipo de evento (unidade de risco). Assim, cada unidade tem sua reserva para conter as perdas inesperadas.

A Figura apresenta um esquema simplificado de alocação de capital econômico⁹¹.

⁹¹ Métodos mais complexos podem ser vistos em Delbaen & Denault (2000) [33].

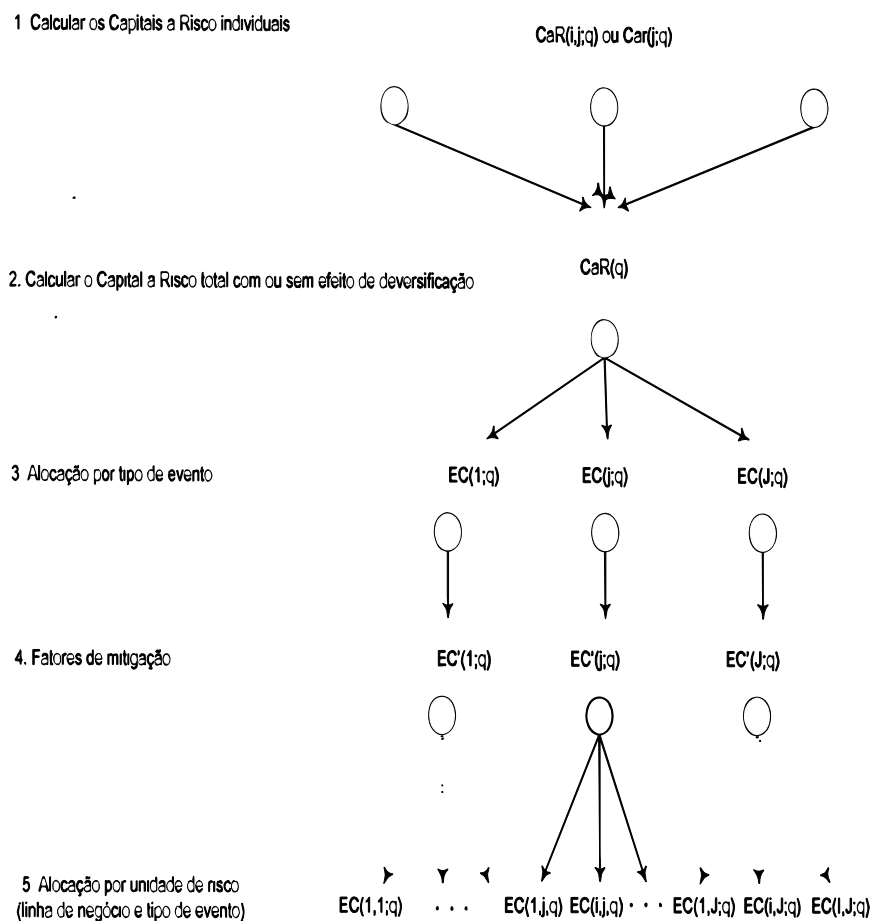


Figura 4 – Efeito bottom-up na alocação de capital econômico com o LDA.

1. Calcular os OPVaR's individuais

Vale destacar que geralmente não existem dados suficientes para considerar linhas de negócio e tipos de evento. Por este motivo, os bancos preferem utilizar somente os tipos de evento para obter resultados mais robustos.

2. Calcular o OPVaR total com ou sem efeito de diversificação

O cálculo sem o efeito de diversificação não leva em consideração a estrutura de dependência que possa haver entre as unidades de risco.

Todavia, o Novo Acordo da Basileia aprova tal procedimento, até porque ele é mais conservador, superestimando o OPVaR.

3. Alocação por tipo de evento

O método mais simples para esta alocação é o linear:

$$EC(j;q) = \delta(j;q)OPVaR(q) \quad (53)$$

onde

$$\mathbf{d}(j; q) = \frac{OPVaR(j; q)}{\sum_{j=1}^J OPVaR(j; q)} \text{ ou } \mathbf{d}(j; q) = \frac{\sum_{i=1}^I OPVaR(i, j; q)}{\sum_{i=1}^I \sum_{j=1}^J OPVaR(i, j; q)} \quad (54)$$

4. Fatores de mitigação

A AMA permite que fatores de mitigação sejam adicionados aos modelos internos. Um exemplo de mitigação permitida pelo Novo Acordo é a cobertura por seguro.

Assim, a alocação por tipo de evento $EC(j; q)$ leva em consideração que fatores de mitigação foram utilizados.

5. Alocação por unidade de risco

$$EC(i, j; q) = \mathbf{d}(i, j; q) EC(j; q) \quad (55)$$

onde

$$\mathbf{d}(i, j; q) = \frac{OPVaR(i, j; q)}{\sum_{i=1}^I OPVaR(i, j; q)} \quad (56)$$

4.1.5.

Extensões ao Modelo de Perdas Operacionais

A AMA permite que extensões sejam feitas ao modelo de distribuição de perdas apresentado anteriormente. Estas extensões visam prever a ação dos controles e da organização interna, bem como dos fatores externos que influenciarão as perdas operacionais em um horizonte de um ano. Um bom exemplo é a inclusão da taxa de recuperação de perdas operacionais ao modelo. A descrição completa deste modelo encontra-se em Guimarães (2003) [46].

Além disso, políticas de risco ou informações de especialistas podem gerar restrições para a estimação dos parâmetros do modelo. Por exemplo, uma unidade operacional deixa de exercer suas funções quando uma perda acumulada limite for atingida. Outro exemplo, uma perda igual ou superior a um valor acontece apenas uma vez a cada d anos. Isso deve ser levado em consideração para que as estimativas sejam não-viesadas. Este tratamento é feito via Análise de Cenários e Auto-Avaliação.

Serão vistos mais adiante os métodos básicos (sem restrições) de estimação dos parâmetros das distribuições de severidade e de frequência.

4.1.6.

Metodologia do Modelo de Perdas Operacionais

A seguir é apresentada metodologia básica para trabalho com modelo de distribuição de perdas:

1. Modelar base de dados por unidade de risco (linha de negócio e tipo de risco).
2. Coletar os dados de perdas operacionais para a base de dados modelada em um período de tempo pré-determinado. Lembrar que é permitido utilizar dados externos e dados simulados para compor a base de dados quando a abordagem AMA é adotada.
3. Modelar os dados de perda para cada sub-base, isto é, analisar os dados, identificando fatos estilizados que indiquem a adoção de modelos para severidade e para frequência por unidade de risco.
4. Estimar os parâmetros das distribuições de severidade e de frequência.
5. Validar a aderência das distribuições paramétricas aos respectivos dados:
 - a. Modelo rejeitado: Vai para 3 ou 2.
 - b. Modelo não rejeitado: Continuar.
6. Realizar teste de independência entre a distribuição de severidade e de frequência.
7. Agregar as distribuições de severidade e frequência segundo algum dos métodos apresentados em 4.1.1.1.
 - a. Se as distribuições forem independentes, pode-se utilizar qualquer um dos métodos.
 - b. Caso contrário, utilizar o método de simulação incluindo a estrutura de dependência na geração pseudo-aleatória.
8. Validar os modelos:
 - a. Modelo rejeitado: Vai para 6 ou para 3.
 - b. Modelo não rejeitado: Continuar.
9. Escolha o quantil (q) para o cálculo do OPVaR.
10. Calcule OPVaR(q).

4.2

Modelos de Severidade

O objetivo desta seção é apresentar uma metodologia para encontrar a distribuição de probabilidade da variável aleatória que melhor se ajuste aos dados de severidade⁹² da base de dados de perdas operacionais.

A Figura 5 descreve de forma sucinta esta metodologia.

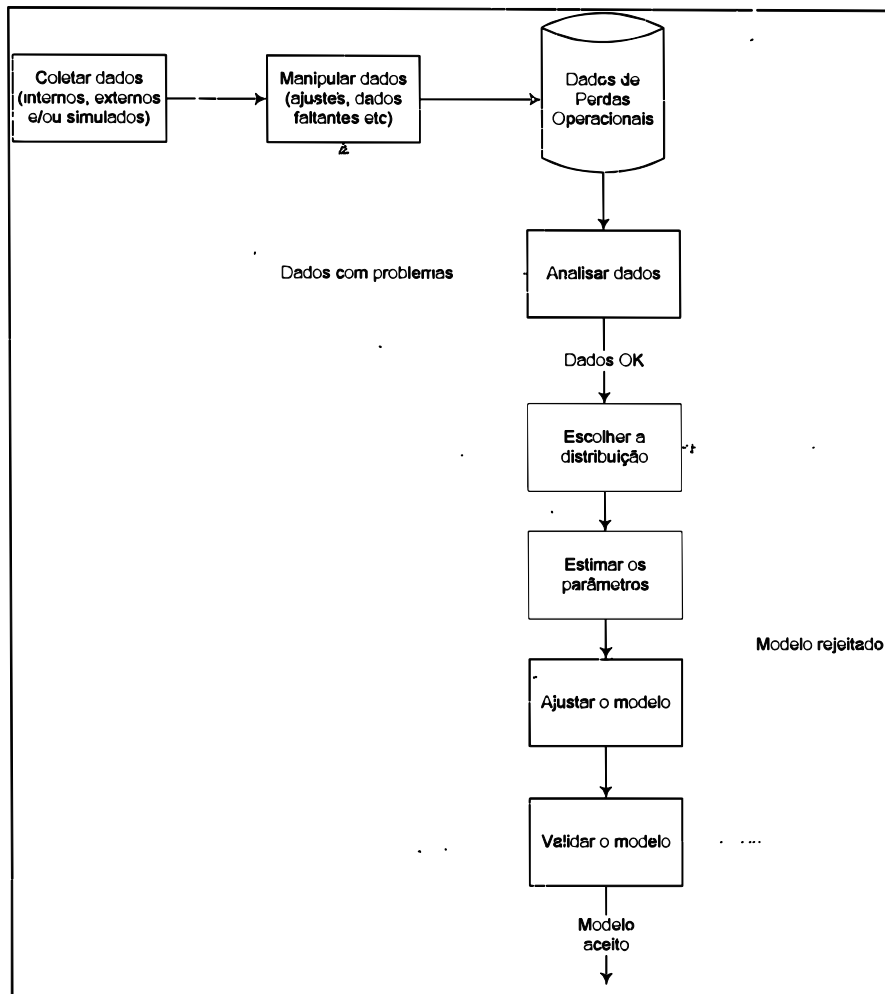


Figura 5 – Abordagem geral para ajuste de distribuição estatística.

Logo, a idéia central consiste em se estimar os parâmetros de uma função de distribuição com base nos dados de perdas operacionais da base de dados.

Todavia, a maioria das unidades de risco apresenta eventos extremos, que são eventos de baixa frequência e alto impacto.

⁹² Segundo Guimarães (2003) [46], a severidade ou impacto é o valor monetário que a instituição financeira perde, caso um evento de risco operacional se materialize.

Portanto, a utilização de técnicas estatísticas tradicionais para estes casos não levará em consideração este comportamento nos extremos, e subestimarão o OPVaR.

Apesar da desconfiança de vários estatísticos em relação à confiabilidade de estimativas baseadas em eventos que nunca foram observados, ferramentas matemáticas da teoria dos valores extremos (TVE)⁹³ apresentam resultados estatísticos significantes, extrapolando eventos extremos do passado para prever o futuro. A Figura 6 ilustra bem este conceito.

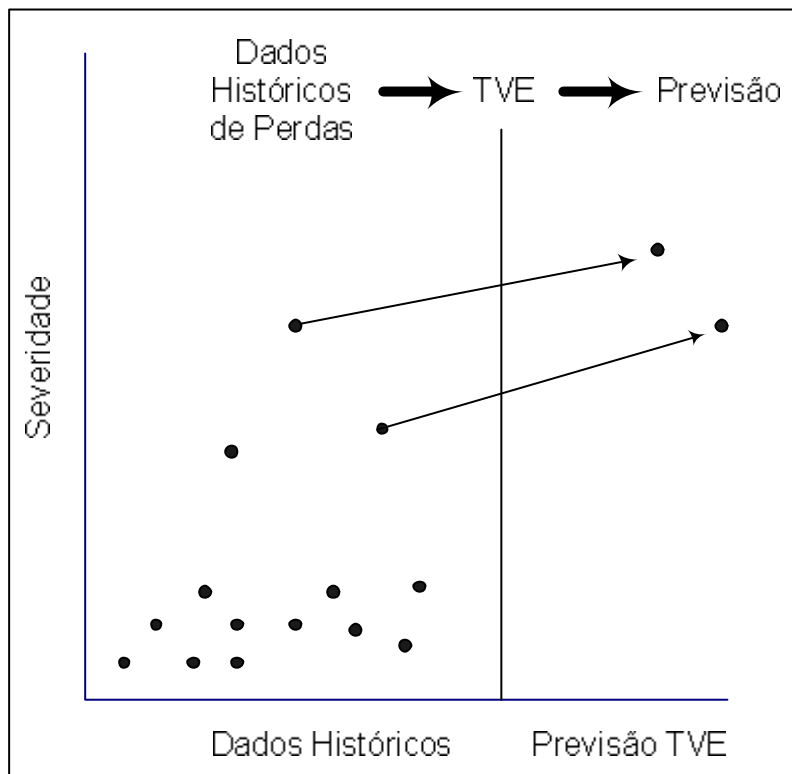


Figura 6 – Ilustração básica da TVE. Fonte: Cruz (2002) [26].

O maior objetivo na determinação da distribuição agregada é o cálculo do VaR em um quantil elevado (região extrema na cauda da distribuição). A distribuição agregada é função da distribuição da severidade, e, portanto, a cauda da distribuição da severidade é de suma importância para o cálculo do VaR.

Quando a unidade de risco apresenta eventos extremos, há escassez de dados para modelagem da cauda da distribuição de severidade, sendo

⁹³ Informações detalhadas sobre TVE podem ser vistas em Mendes (2004) [64].

necessário utilizar uma distribuição assintótica limite da estatística de ordem que represente o máximo na amostra de severidade de perdas operacionais, ou então uma distribuição limite do excesso das perdas operacionais em relação a um limiar muito alto.

A Tabela 4 apresenta uma ordenação do peso da cauda⁹⁴ de algumas distribuições.

| Distr. | Weibull ($a > 1$) | Exponencial | Pearson | Weibull ($a < 1$) | Lognormal | GEV, GPD |
|---------------------|-----------------------------------|-------------|-----------------|------------------------|-----------------|-------------|
| f(x) p/ x grande | $\text{Exp}(-x^A)$ ($A > 1$) | $\exp(-x)$ | $x^A \exp(-Bx)$ | $\exp(-x^A)$ | $x^{-A} \log x$ | x^{-A} |

peso da cauda →

Tabela 4 – Comparação entre o peso das caudas de algumas distribuições. A e B são constantes. Fonte: Cruz (2002) [26].

Em última instância, a análise de dados indicará os fatos estilizados que apontarão qual distribuição subjacente paramétrica utilizar para descrever o comportamento da variável aleatória severidade.

4.2.1.

Distribuições de Probabilidade

Tendo em vista a característica de eventos extremos em grande parte das unidades de risco operacional, esta seção apresenta uma descrição básica da modelagem univariada dos extremos e dos excessos para a severidade. Estas modelagens fazem parte da TVE.

Assim, são apresentadas duas distribuições, que de acordo com o valor dos parâmetros estimados se degeneram em seis, aumentando a possibilidade de um melhor ajuste aos dados de severidade.

⁹⁴ Não existe comum acordo sobre a definição de uma distribuição de cauda pesada. Um exemplo de definição baseia-se no momento máximo de uma distribuição – $\sup\{r : E(x^r) < \infty\}$. Nesta definição, uma distribuição é dita de cauda leve se tiver momento finito para todas as ordens r , e cauda pesada caso contrário (de Fontnouvelle & Rosengren, 2004) [31].

Entretanto, outras distribuições devem ser utilizadas na fase de escolha da distribuição, permitindo que na validação dos modelos seja selecionado aquele que melhor se ajuste aos dados.

Além das distribuições da TVE, costuma-se utilizar as seguintes distribuições⁹⁵:

- Exponencial, Pareto;⁹⁶
- Weibull, Gumbel;⁹⁷
- Lognormal;
- Gamma;
- Loggamma;
- Burr;
- Logistic;

4.2.1.1.

GEV (Generalized Extreme Value)

A GEV é a distribuição limite para a estatística de ordem que define o máximo de uma amostra iid.

Seja X uma variável aleatória com distribuição GEV (ξ, μ, s) , onde μ é o parâmetro de localização, s de escala, e ξ de forma. Seja $z = (x - \mu) / s$, então:

$$P(X \leq x) = F_{\xi, \mu, s}(x) = F_{\xi, 0, 1}(z) \quad (57)$$

$$= \begin{cases} \exp[-\exp(-z)], & \forall z \quad (\xi = 0) \\ \exp[-(1 + \xi z)^{-1/\xi}], & \forall z \mid 1 + \xi z \geq 0 \quad (\xi \neq 0) \end{cases}$$

Quando $\xi \rightarrow 0$ ($\xi > 0$) ou $\xi = 0$, a GEV se degenera em uma distribuição Gumbel. Para $\xi > 0$ e $\xi < 0$ a GEV representa as distribuições Fréchet e Weibull respectivamente.

A amostra utilizada para estimação dos parâmetros baseia-se no método do máximo dos blocos. Nele, M_1, \dots, M_m são variáveis aleatórias iid que representam os máximos coletados em blocos de tamanho n ⁹⁸. A amostra de m

⁹⁵ de Fontnouvelle & Rosengren (2004) [31], nas páginas 8-11, apresenta várias características destas distribuições que servem de subsídio para escolha da distribuição com base nas informações da análise dos dados.

⁹⁶ Estas distribuições são degenerações da GPD, como será visto adiante.

⁹⁷ Estas distribuições são degenerações da GEV, como será visto adiante.

⁹⁸ Mesmo que as variáveis originais X_1, X_2, \dots possuam alguma forma de dependência, é razoável supor que os blocos de tamanho n sejam variáveis aleatórias iid em \mathbb{R}^1 , e que os máximos em cada bloco sejam também iid. De fato, n deve ser escolhido de tal forma que essas suposições sejam satisfeitas (Mendes, 2004) [64].

máximos (m_1, \dots, m_m) é utilizada nos métodos de estimação de parâmetros, que serão vistos mais adiante. A Figura 7 representa esta técnica de amostragem.

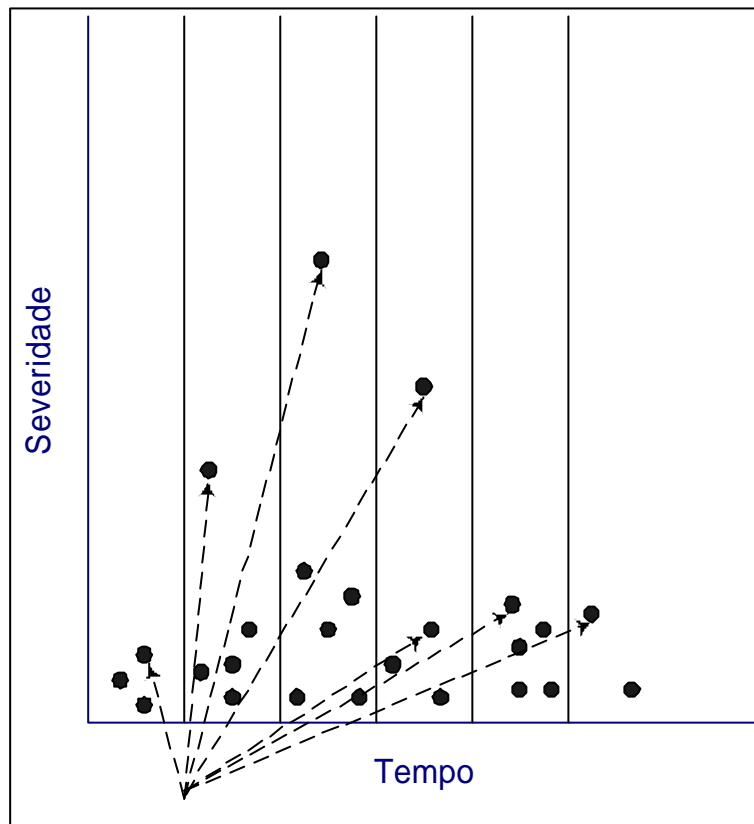


Figura 7 – Método do Máximo dos Blocos: eventos de maior severidade são coletados em determinados períodos de tempo. Fonte: Cruz (2002) [26].

4.2.1.2.

GPD (*Generalized Pareto*)

A GPD é a distribuição limite do excesso das perdas em relação a um limiar muito alto.

Seja X uma variável aleatória com distribuição GPD (α, μ, s) , onde μ é o parâmetro de localização, s de escala, e α de forma. Seja $z = (x - \mu) / s$, então:

$$\begin{aligned}
 P(X \leq x) &= G_{\alpha, \mu, s}(x) = G_{\alpha, 0, 1}(z) \\
 &= \begin{cases} 1 - \exp(-z), & \forall z \geq 0 & (\alpha = 0) \\ 1 - (1 + \alpha z)^{-1/\alpha}, & \begin{cases} \forall z \geq 0 & (\alpha > 0) \\ 0 < z < -1/\alpha & (\alpha < 0) \end{cases} \end{cases} \quad (58)
 \end{aligned}$$

Quando $\alpha \rightarrow 0$ ($\alpha < 0$) ou $\alpha = 0$, a GPD se degenera em uma distribuição Exponencial. Para $\alpha > 0$ e $\alpha < 0$ a GPD representa as distribuições Pareto e Beta respectivamente.

A amostra utilizada para estimação dos parâmetros baseia-se no método POT (*Peak Over Threshold*). Determina-se um limiar u suficientemente alto para satisfazer as suposições teóricas e não gerar viés nas estimativas. Por outro lado, u não pode ser alto demais, o que implicaria em um número pequeno de observações na cauda, resultando em maior variabilidade dos estimadores⁹⁹. A Figura 8 exemplifica esta técnica.

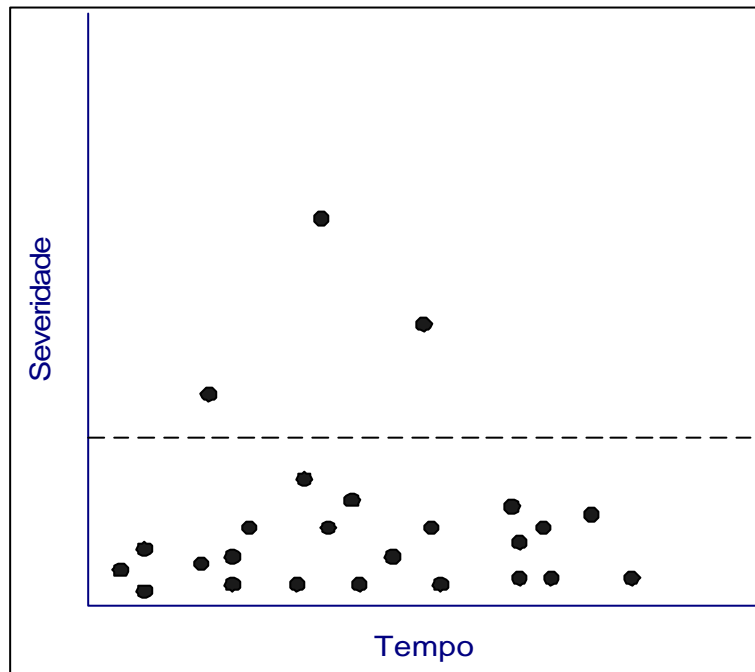


Figura 8 – Método POT (Peak Over Threshold): eventos são selecionados acima de um limite (linha tracejada), ignorando o tempo. Fonte: Cruz (2002) [26].

Cuidado, pois a GPD apresenta a distribuição dos excessos, e na verdade o que se quer é a distribuição da severidade. Para u suficientemente alto, pode-se aproximar a cauda da distribuição F da severidade para (Hidetoshi, 2002) [51]:

$$F(x) = (1 - F(u))G_{\alpha,0,s}(x-u) + F(u), \quad x > u \quad (59)$$

⁹⁹ Métodos para determinação do limiar ótimo encontram-se em Pickands (1975;1994) [72,73].

Assim é necessário estimar dois parâmetros (θ , s). Entretanto, este trabalho pode ser simplificado caso os dados sejam transformados. Ao se aplicar o logaritmo natural aos dados da amostra, e ao limiar u , a distribuição dos excessos para essas novas variáveis sempre terá caudas leves, e então θ será igual a 0. Portanto, a variável aleatória transformada só possuirá um parâmetro para estimação (s) (de Fontnouvelle et al., 2003) [30].

4.2.2.

Análise dos Dados

Esta seção apresenta algumas ferramentas para escolha da distribuição subjacente a ser utilizada para estimação dos parâmetros do modelo de severidade.

4.2.2.1.

Distribuição Empírica e Gráfico de Cauda

A relação entre estatística de ordem e a função de distribuição empírica cumulativa é direta:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}} \quad x \in \mathfrak{R} \quad (60)$$

Em termos práticos, o gráfico da distribuição empírica pode ser feito utilizando a fórmula:

$$p_{k,n} = \frac{n - k + 0.5}{n} \quad (61)$$

onde k é o número de ordem do dado, e n o número de eventos da amostra (Cruz, 2002) [26].

A forma empírica é mais usada para análise de dados, dando indicações sobre a forma paramétrica a ser utilizada no ajuste aos dados.

Esta análise pode ser feita visualmente, comparando o gráfico da distribuição empírica com gráficos de outras distribuições, ou por intermédio do gráfico da cauda.

O gráfico da cauda é obtido pelo par ordenado $(\log(x_i), \log(1-F(x_i)))$ para as observações $\{x_i\}$.

Segundo de Fontnouvelle & Rosengren [31], para os gráficos de cauda que apresentam um comportamento linear, a sugestão é que a distribuição

subjacente seja uma distribuição Pareto Tipo I. Além disso, a inclinação do gráfico (a) indica uma estimativa heurística para o índice de cauda¹⁰⁰:

$$\log(1 - F(x_i)) = -a \log(x_i) + c \quad (62)$$

, onde c é uma constante.

Índices de cauda maiores que 1 sugerem que a perda esperada seja muito alta, o que torna o OPVaR praticamente inviável de ser alocado. Contudo, de Fournouelle & Rosengren (2004) [31] afirmam que as distribuições subjacentes da severidade não têm caudas tão pesadas assim, e que portanto, é possível obter OPVaR razoáveis.

4.2.2.2.

Gráfico da Média dos Excessos

A média dos excessos para um dado limiar é definida como a média de todas as perdas superiores ao limiar menos o valor do limiar:

Portanto, a média dos excessos é uma função do limiar.

Segundo de Fournouelle & Rosengren (2004) [31], a forma deste gráfico depende do tipo da distribuição subjacente:

- Linear com inclinação positiva e maior que 1 → Pareto
- Horizontal (inclinação zero) → Exponencial
- Côncava com inclinação positiva e maior que 1 → Lognormal ou Weibull

4.2.3.

Estimação dos Parâmetros

O objetivo desta seção é apresentar detalhes dos métodos de estimação dos parâmetros das distribuições de severidade escolhidas como distribuições subjacentes.

Entretanto, há que se ter cuidado com a amostra utilizada para estimação dos parâmetros. Dependendo das questões envolvidas na amostragem, o modelo de severidade sofre alguns ajustes.

¹⁰⁰ O índice de cauda será tratado com mais detalhes no item 4.2.3.2.4 deste trabalho.

4.2.3.1.

Amostragem

Existem dois problemas fundamentais relacionados às amostras de perdas operacionais: dados truncados e dados agregados.

Estes problemas, bem como suas soluções, são vistas nos dois próximos tópicos.

4.2.3.1.1.

Dados Truncados

Conforme visto na seção 3.2, é necessário fazer o tratamento adequado do limiar na amostra de perdas operacionais. Assim, garante-se que os estimadores serão não-viesados, e que o OPVaR não será superestimado.

Mesmo que só sejam utilizados dados internos para estimação da distribuição de severidade, sempre haverá um limiar. Logo, para a estimação dos parâmetros da distribuição subjacente, a amostra deve ser considerada truncada, e isso requer um tratamento estatístico apropriado para evitar o viés.

Como o grau de incerteza a respeito do limiar é grande, uma vez que cada área de negócio coleta os dados de forma diferente, este limiar passa a ser considerado um parâmetro a ser estimado no modelo de severidade. É claro que se houver certeza sobre este limiar, a estimação dos parâmetros do modelo é simplificada.

Além disso, quando os dados externos (públicos ou consórcio) são usados para complementar os dados de perdas, principalmente com o objetivo de melhorar a previsão de eventos extremos, deve-se utilizar um método para torná-los compatíveis¹⁰¹ com os dados internos. Como no caso dos dados internos, sempre existe um limiar para os dados externos, e na maioria das vezes também é desconhecido.

De forma geral, a distribuição de severidade deve ser estimada com base em m conjuntos de dados de perdas de m diferentes provedores, onde provedor pode ser uma unidade de negócio interna ao banco, público externo, consórcio ou simulação (análise de estresse). Assim, existem dois métodos possíveis para a calibração da severidade:

¹⁰¹ A incompatibilidade está relacionada a problemas de escala e de viés.

- Método 1: As distribuições de probabilidade subjacentes das m fontes de dados de perdas são as mesmas, mas os limiares de cada fonte podem ser diferentes.
- Método 2: As m fontes de dados de perda são provenientes de diferentes distribuições subjacentes, e logo devem ser re-escaladas. Além disso, os limiares também podem ser diferentes como no método anterior.

Apesar do método 2 ser mais geral e realístico, a sua complexidade não vale o esforço de implementação (Frachot et al., 2003) [44].

Para o método 1, consideram-se dois casos mutuamente excludentes:

- Caso 1: O limiar corresponde a um vetor não-estocástico e desconhecido, onde cada coordenada representa o limiar de uma fonte de dados. É claro que se houver fontes cujos limiares sejam conhecidos, a estimação do modelo é simplificada.
- Caso 2: O limiar é estocástico, sendo representado por uma distribuição contínua e não-negativa. Este caso representa a situação aonde o número de fontes é muito grande, como por exemplo, nos dados públicos.

O caso 2 é bem mais sofisticado, e necessita que seja arbitrada uma distribuição para o limiar¹⁰².

Portanto, este trabalho só apresentará o método 1 e o caso 1 (Baud et al., 2002a; 2002b; Frachot et al., 2003) [14,15,55].

O conceito geral reside no fato da distribuição da amostra total ser fundamentalmente diferente da distribuição verdadeira por conta dos diferentes limiares das fontes de dados. Em termos estatísticos, a distribuição da amostra total está relacionada com a distribuição verdadeira condicionada a perdas maiores que o limiar. É neste ponto que a máxima verossimilhança se encaixa.

A máxima verossimilhança é um método assintoticamente eficiente desde que a função de verossimilhança tenha sido corretamente especificada, isto é, a distribuição da amostra seja correta. Denomina-se $f(\cdot; \theta)$ a distribuição (verdadeira) da severidade, onde θ é o vetor de parâmetros que caracterizam essa distribuição.

Exemplificando o conceito acima para o caso de uma única fonte de dados de perda, como todos os dados são coletados acima de um único limiar H , a

¹⁰² Informações detalhadas sobre o caso 2 são vistas em Baud et al. (2002^a) [14] e de Fontnouvelle (2003) [30].

distribuição da amostra da severidade $f^*(\cdot; \theta)$ é igual a distribuição verdadeira condicionada às perdas que excedem H . Logo:

$$f^*(x; \mathbf{q}, h) = f(x; \mathbf{q} | H = h) = \mathbf{1}\{x \geq h\} \cdot \frac{f(x; \mathbf{q})}{\int_h^{+\infty} f(x; \mathbf{q}) dx} \quad (63)$$

, onde $\mathbf{1}\{x \geq h\}$ é uma variável *dummy* igual a 1 se $x \geq h$, e 0 caso contrário.

Já para o caso de m fontes de dados de perdas, os parâmetros adicionais não são somente os m limiares h_1, \dots, h_m (representados pelo vetor \mathbf{h}), mas também os pesos p_1, \dots, p_m de cada provedor (representados pelo vetor \mathbf{p}), isto é, o número de dados de perdas de um provedor sobre o número total de perdas. Finalmente, a função de verossimilhança deve se basear na distribuição de probabilidade da amostra total:

$$f^*(x; \mathbf{q}, \mathbf{h}, \mathbf{p}) \equiv f(x; \mathbf{q} | H_1 = h_1) \propto p_1 + K + f(x; \mathbf{q} | H_m = h_m) \propto p_m \quad (64)$$

Logo, aplicando o método da máxima verossimilhança, os parâmetros devem ser calculados com base na seguinte fórmula:

$$(\hat{q}, \hat{h}, \hat{p}) = \arg \max \left\{ \left[\sum_{i \in \text{fonte}_1} \ln f^*(x_i; \mathbf{q}, h_1, p_1) \right] + \right. \\ \left. + K + \left[\sum_{i \in \text{fonte}_m} \ln f^*(x_i; \mathbf{q}, h_m, p_m) \right] \right\} \quad (65)$$

Alguns comentários a respeito do método apresentado:

- Quando $m = 1$ (obviamente $p = 1$), θ é estimado para diferentes h , o que estabelece curvas $h = h(\theta)$. H é determinado graficamente no ponto de inflexão destas curvas.¹⁰³
- Mesmo para $m = 1$, este método lida com severa perda de acurácia já que todas as perdas abaixo do ponto de inflexão são desconsideradas para o processo de calibração. Logo, existe um *trade-off* entre vício e acurácia.
- Sempre haverá viés quando $m > 1$.
- Os casos aonde $m > 1$ são mais propensos a acontecer na prática, embora requeiram algoritmos de otimização de alto nível para resolução da equação (65).

¹⁰³ Baud et al. (2002b) [15] garante que existe prova matemática para este procedimento, quando $m = 1$.

- Devem-se aplicar testes estatísticos rigorosos para decidir se um peso é igual a zero. Caso haja aceitação da hipótese nula, este peso é retirado do processo de estimação. Na prática, aumenta-se m , até que o logaritmo natural da função de verossimilhança pare de crescer. Entretanto, é possível usar um teste estatístico que aponte o m ótimo, isto é, a partir desse valor não é esperado algum crescimento estatisticamente significativo para verossimilhança. Como exemplo, utiliza-se o teste da razão de verossimilhança, cujo esforço computacional não muito grande.
- Como f^* é uma função não-negativa, quanto maior o número de dados, maior será o valor da verossimilhança. Logo, em uma primeira análise:
$$\hat{h}_j = \min_{i \in \text{fonte}_j} x_i.$$
 Entretanto, os dados da fonte j podem ter sido coletados de forma errada. Como exemplo, uma fonte externa pode apresentar poucos dados com severidades muito baixas. Assim, o limiar verdadeiro será significativamente subestimado, sendo preferível considerar que a amostra está contaminada por dados aberrantes e não-informativos, que devem ser excluídos do processo de estimação.

4.2.3.1.2.

Dados Agregados

Normalmente as bases de dados de perdas contêm eventos individuais e eventos agregados. Isto ocorre principalmente quando a base de dados de perdas operacionais é criada, pois nem sempre é possível coletar todos os eventos individuais do passado. Na maioria das vezes, somente as somas das severidades destes eventos agregados estão disponíveis.

O efeito dos dados agregados está no fato dos métodos de estimação não poderem ser utilizados diretamente sobre a amostra. Assim, duas soluções são apresentadas, partindo-se do pressuposto que as perdas agregadas são conhecidas (Frachot & Roncalli, 2001) [42].

A primeira: eliminar todas as observações agregadas da amostra. Geralmente, esta solução não é satisfatória, pois existem poucos dados históricos de perdas.

A segunda: suponha que sejam conhecidos os números de eventos de cada perda agregada. Seja ζ a severidade de uma perda individual e ξ a severidade de uma perda agregada. Considere t o índice da observação. Logo, a

estrutura da base de dados é (t, n_t, ξ_t) onde t é a observação, n_t é o número de eventos correspondentes a esta observação e ξ_t é a severidade da perda agregada observada. Se a observação for uma perda individual, então o registro na base de dados é $(t, 1, \zeta_t)$. Assim, a perda individual pode ser considerada um caso especial de perda agregada. Dessa forma, ξ_t é definida como:

$$\mathbf{x}_t = \sum_{n=1}^{n_t} \mathbf{z}_n \quad (66)$$

Assim a equação (66) representa o modelo a ser utilizado para estimação dos parâmetros da distribuição de severidade.

Pelo fato de na maioria das vezes não haver expressão analítica para a distribuição de ξ , não é possível utilizar o método da máxima verossimilhança para estimação dos parâmetros.

Uma solução é o método generalizado dos momentos (GMM – *generalized method of moments*). Sabendo-se os momentos teóricos da distribuição de severidade (ζ), encontra-se uma expressão para os momentos de ξ ¹⁰⁴:

$$\begin{aligned} E[\mathbf{x}_t] &= n_t E[\mathbf{z}] \\ E[\mathbf{x}_t^2] &= n_t E[\mathbf{z}^2] + n_t(n_t - 1)E^2[\mathbf{z}] \\ E[\mathbf{x}_t^3] &= n_t E[\mathbf{z}^3] + 3n_t(n_t - 1)E[\mathbf{z}]E[\mathbf{z}^2] + n_t(n_t - 1)(n_t - 2)E^3[\mathbf{z}] \end{aligned} \quad (67)$$

Parâmetros da distribuição de severidade são facilmente estimados por intermédio do GMM. Quando existem dificuldades de se implementar o GMM, métodos mais complicados podem se utilizados, como os métodos dos momentos simulados e o da inferência indireta.¹⁰⁵

Este assunto tem duas implicações com o limiar:

- Com o passar do tempo, uma vez estabelecido o limiar para coleta dos eventos, o efeito causado pelos dados agregados irá diminuindo (Frachot et al., 2003) [44].
- Para base de dados com muitos eventos agregados com altas severidades, além do modelo descrito pela fórmula (66), deve-se levar em consideração também o modelo para dados truncados. Isso complica e muito o processo de estimação, e normalmente estes dois fenômenos não são tratados em conjunto.

¹⁰⁴ Utilizar os resultados apresentados no apêndice A de Frachot et al. (2001) [42] para encontrar as expressões para os momentos de ordens superiores.

¹⁰⁵ Estes métodos são detalhados no apêndice D de Frachot et al. (2001) [42].

4.2.3.2.

Métodos de Estimação

Esta seção apresenta alguns dos principais métodos de estimação utilizados nos modelos de risco operacional. O objetivo não é descrever no detalhe os métodos, mas sim indicar as suas particularidades em relação ao risco operacional. Vários livros de estatística tratam destes métodos de forma didática e completa, ficando a cargo do leitor um aprofundamento nos métodos de seu interesse.

4.2.3.2.1.

Método dos Momentos Truncados

Este método apresenta uma solução alternativa ao problema dos dados truncados visto na seção 4.2.3.1.1.

Não é possível aplicar o método dos momentos (incondicionais) para se estimar os parâmetros da distribuição subjacente, pois esta é uma distribuição condicionada ao limiar. Neste caso, é possível estimar estes parâmetros por intermédio do método dos momentos truncados, visto em Álvarez (2001) [1]:

1. Determine os n primeiros momentos teóricos da distribuição de probabilidade (verdadeira) truncada, onde n é o número de parâmetros da distribuição.
2. Calcule os n primeiros momentos empíricos da amostra.
3. Iguale cada expressão de momento truncado com seu respectivo valor de momento empírico.
4. Resolva as n equações simultâneas.
5. A solução destas equações apresentará os n parâmetros da distribuição incondicional.

Apesar da facilidade de implementação em relação a solução apresentada na seção 4.2.3.1.1, este procedimento apresenta duas desvantagens:

- Considera o limiar único. Quando existem várias fontes de dados, para não gerar viés nos parâmetros, o limiar único deverá ser igual ao maior limiar de todas as fontes. Isso diminui a acurácia.
- Considera o limiar conhecido. Conforme visto anteriormente, isso permite que dados aberrantes e não-informativos façam parte do processo de estimação.

4.2.3.2.2.

Método da Máxima Verossimilhança

Apesar do método da máxima verossimilhança (ML – *Maximum Likelihood*) ser um dos mais populares, ele apresenta dois sérios problemas em relação às especificidades dos modelos de risco operacional:

- Conforme já foi visto, geralmente a severidade apresenta eventos extremos. Assim, existe um maior interesse na cauda da distribuição subjacente, sendo que o método da máxima verossimilhança pondera cada valor da distribuição igualmente.¹⁰⁶
- Hidetoshi (2002) [51] observa que tendo em vista a alta concentração dos dados em uma faixa reduzida, e o pequeno tamanho da amostra, geralmente, o OPVaR baseado na distribuição de severidade estimada por este método não é robusto em relação ao valor do limiar.

4.2.3.2.3.

Método dos Momentos Ponderados

O método dos momentos ponderados (PWM – *Probability Weighted Method of Moments*) foi desenvolvido por Hosking & Wallis (1997) [54]. O objetivo foi possibilitar a aplicação do método dos momentos para distribuições com momentos inexistentes. A descrição deste método encontra-se no anexo E desta trabalho.

Analisando a descrição do método, observa-se que ele atribui um maior peso aos valores da cauda da distribuição, resolvendo um problema apontado no método ML.

Além desta vantagem sobre o ML, Cruz (2002) [26] apresenta outras, destacando uma melhor aplicabilidade do PWM em pequenas amostras¹⁰⁷ - ver Landwehr et al. (1979) [59]:

- Embora os estimadores PWM sejam assintoticamente ineficientes comparados aos estimadores ML, nenhuma deficiência é detectável em amostras de tamanho 100 ou menos – ver Hosking et al. (1985) [53].

¹⁰⁶ O efeito da ponderação igualitária é resolvido por intermédio do método da máxima verossimilhança penalizada.

¹⁰⁷ Esta situação representa bem o caso do risco operacional, principalmente quando se utilizam técnicas de amostragem POT ou Máximo dos Blocos para estimar distribuições de caudas pesadas.

- Os vieses dos estimadores PWM são pequenos, e decrescem rapidamente com o aumento do tamanho da amostra. Além disso, Coles & Dixon (1998) [25] confirmaram que os estimadores PWM são menos viciados que os estimadores ML para pequenas amostras.
- Os desvios padrões dos estimadores PWM são comparáveis aos dos estimadores ML para amostras de tamanho moderado ($n = 50, 100$), e sempre são significativamente menores para pequenas amostras.

Hidetoshi (2002) [51] observa que os valores dos parâmetros estimados pelo PWM, bem como o OPVaR correspondente são relativamente estáveis a variações de limiar. Além disso, ele observa também que o ajuste na cauda para a distribuição GPD apresenta os melhores resultados quando estimada pelo PWM.

4.2.3.2.4.

Métodos de Estimação do Parâmetro da Forma (?)

A estimação do parâmetro da forma (?) é fundamental para inferência estatística da cauda da distribuição, especialmente se a distribuição subjacente for a GEV ou a GPD.

O índice da cauda (?) é o inverso do parâmetro da forma (?):

$$g = 1/x \quad (68)$$

Estimadores do índice de cauda não assumem que as observações extremas seguem exatamente uma distribuição assintótica, e então eles podem ser mais eficientes que os estimadores ML.

Hill (1975) [52] propôs um conjunto de estimadores para ?, cuja consistência para distribuições com caudas pesadas foi provada por Pictet et al. (1996) [74].

Para as estatísticas de ordem $x_{1:n} > x_{2:n} > \dots > x_{n:n}$ de uma amostra aleatória de tamanho n , e supondo $? > 0$ (caudas pesadas), os estimadores de Hill para o parâmetro da forma são:

$$\hat{x}_k = \frac{1}{k-1} \sum_{j=1}^{k-1} (\ln x_{j:n} - \ln x_{k:n}) \quad , k = 2, K, n \quad (69)$$

, onde k representa o número de observações da cauda. Particularmente, quando a distribuição subjacente for uma GPD, k representa o número de observações que excedem um determinado limiar.

O gráfico de Hill k ? \hat{x}_k auxilia na seleção do limiar e do parâmetro da forma ao mesmo tempo. O parâmetro deve ser escolhido onde o gráfico estiver mais estável (*flat*), e o k deve ser o maior possível para este parâmetro (aumenta a acurácia).

Coleman (2003) [23] utiliza um método analítico empírico para a determinação do parâmetro da forma, mas só vale para amostras pequenas:

$$\hat{x} = \frac{1}{n-6} \sum_{k=4}^{n-3} \hat{x}_k \quad (70)$$

Apesar da fórmula apresentada acima, tradicionalmente, a estimação de β depende fortemente da escolha de k . Entretanto Huisman, Koedijk, Kool e Palm (2001) [55], a partir de agora HKKP, propuseram uma melhoria baseada em uma regressão no estimador de Hill para minimizar o papel da seleção de k .

HKKP notaram que o estimador de Hill é viciado em pequenas amostras, e este viés é aproximadamente linear em k , então:

$$E[\hat{x}_k] = \beta + ck \quad (71)$$

, onde β representa o verdadeiro valor do parâmetro da forma, e c uma constante.

Assim, motivados pela equação (71), HKKP propuseram a seguinte regressão:

$$x_k = \beta_0 + \beta_1 k + e_k \quad (72)$$

Esta regressão é estimada para k variando de 1 a K . HKKP sugerem que K seja igual a metade do tamanho da amostra N . Apesar disso, de Fontnouvelle & Rosengren (2004) [31] indicaram que deve-se procurar um K ótimo, uma vez que para $K = N/2$ (janela grande dependendo do tamanho de N) pode não haver linearidade.

Apesar disto, HKKP e de Fontnouvelle & Rosengren (2004) [31] concluíram que a estimativa de β_0 é robusta em relação a escolha de K , onde β_0 é interpretado como a estimativa de β com viés corrigido.

Como HKKP é uma melhoria do estimador de Hill, e o estimador de Hill é construído sob a hipótese de $\beta > 0$, então não se pode utilizar estes dois métodos para rejeitar a hipótese de distribuição de cauda leve. Esta é uma hipótese muito importante, pois distribuições de caudas leves podem ter diferentes implicações para o OPVaR em relação a distribuições de caudas pesadas.

Como solução para este problema, podem ser utilizados os estimadores de Pickands (1975) [72] ou de Dekkers, Einmahl e de Haan (1989) [32] – DEdH¹⁰⁸, para os quais θ pertence aos Reais.

Para as estatísticas de ordem $x_{1:n} > x_{2:n} > \dots > x_{n:n}$ de uma amostra aleatória de tamanho n , podendo θ assumir qualquer valor nos Reais, os estimadores de Pickands para o parâmetro da forma são:

$$\hat{\alpha}_k = \frac{1}{\ln 2} \ln \left(\frac{x_{k:n} - x_{2k:n}}{x_{2k:n} - x_{4k:n}} \right) \quad (73)$$

Para as estatísticas de ordem $x_{1:n} > x_{2:n} > \dots > x_{n:n}$ de uma amostra aleatória de tamanho n , podendo θ assumir qualquer valor nos Reais, os estimadores de DEdH para o parâmetro da forma são:

$$\hat{\alpha}_k = 1 + H_n^{(1)} + \frac{1}{2} \left(\frac{H_n^{(1)2}}{H_n^{(2)}} - 1 \right)^{-1} \quad (74)$$

$$H_n^{(j)} = \frac{1}{k} \sum_{j=1}^k (\ln x_{j:n} - x_{k:n})^j$$

Logo, os estimadores de Pickands e de DEdH dependem do valor de k .

A maneira mais simples¹⁰⁹ de se determinar o estimador do parâmetro da forma para estes dois estimadores é feita com auxílio de gráfico, assim como no caso do estimador de Hill.

4.2.3.2.5.

Método dos Mínimos Quadrados

A estimação baseada no método dos mínimos quadrados (LS – Least Square) é dada pela resolução de um problema de minimização da soma dos quadrados das diferenças entre as probabilidades da distribuição acumulada empírica e as correspondentes probabilidades obtidas da distribuição a se estimar.

Logo, a distribuição estimada por este método apresenta o melhor ajuste aos dados da amostra, considerando a distribuição como um todo. Isso é facilmente visto no gráfico P-P *plot*.

¹⁰⁸ É uma extensão ao estimador de Hill.

¹⁰⁹ Hall (1990) [48] sugere um critério de minimização do erro quadrático médio para o estimador de Hill, usando bootstrap; Danielsson & de Vries (1997b) [28] ajustam o procedimento de Hall. Estes procedimentos dispensam o gráfico.

Entretanto, para o cálculo do OPVaR a preocupação maior deve ser no ajuste junto a cauda da distribuição, e este método geralmente apresenta uma diferença muito grande entre os dados reais e os obtidos na distribuição estimada na cauda.

4.2.4.

Critérios de Seleção

Uma vez analisados os dados da amostra, escolhidos alguns modelos de severidade, estabelecidos os pares: função de distribuição e método de estimação para estes modelos, estimados os parâmetros ajustando a distribuição à amostra para cada par, é necessário selecionar que distribuição estimada melhor representa a severidade da unidade de risco em questão.

Existem dois tipos de critérios que auxiliam esta seleção:

- Critérios baseados em análises gráficas.
- Critérios baseados em testes analíticos.

Esta seção apresenta alguns dos principais critérios de seleção utilizados nos modelos de risco operacional. O objetivo não é detalhar os critérios, mas sim apresentar as suas particularidades em relação ao risco operacional. Vários livros de estatística tratam destes critérios de forma didática e completa, ficando a cargo do leitor um estudo adicional.

4.2.4.1.

PP-plot ou QQ-plot

PP-plot ou QQ-plot são critérios gráficos para verificar quão próxima a distribuição estimada está da distribuição empírica.

Quanto melhor o ajuste, mais próximo da função identidade estarão os pontos “plotados”.

É importante verificar o comportamento do gráfico na parte superior à direita, pois é lá que se encontra a cauda, que representa as maiores severidades (valores extremos). Portanto, não adianta ter um ajuste perfeito no corpo da distribuição, e discrepâncias na cauda. Como o objetivo final é o cálculo do OPVaR em quantis elevados, o ajuste na cauda passa a ser a prioridade número um.

4.2.4.2.

Função de Sobrevivência do Máximo Teórico

Investiga-se a cauda da distribuição de severidade pela observação da relação entre o valor teórico máximo do suporte da distribuição estimada e o maior valor dos dados originais.

Esta relação é obtida pelos pares ordenados: severidade da perda original, e respectivo valor da função de sobrevivência do máximo teórico, para todos os dados originais da cauda.

A função de sobrevivência do máximo teórico para um determinado valor é a probabilidade da variável aleatória, que representa o máximo do suporte da distribuição estimada exceder o referido valor.

Em termos matemáticos, seja F_{\max} a função de distribuição do maior valor dentre N variáveis aleatórias, $M_N = \max(X_1, \dots, X_N)$, a função sobrevivência do máximo teórico é representada pela seguinte fórmula:

$$\bar{F}_{\max}(x) = 1 - F_{\max}(x) = P(M_N > x) \quad (75)$$

Considerando as N variáveis aleatórias das perdas como iid, a formula (75) é simplificada para:

$$\bar{F}_{\max}(x) = P(M_N > x) = [P(X > x)]^N = [1 - F(x)]^N \quad (76)$$

Hidetoshi (2002) [51] apresenta um exemplo de análise gráfica utilizando a função sobrevivência do máximo teórico para quatro tipos de distribuições estimadas por intermédio do método de estimação da máxima verossimilhança.

Quando a extremidade a direita da curva da função sobrevivência do máximo teórico é próxima de 1 (caso LN na Figura 9), julga-se que a cauda da distribuição estimada é mais pesada que os dados originais. Isto significa que o máximo teórico excede a maior perda dos dados originais com uma probabilidade muito alta. Logo, o risco é superestimado.

Por outro lado, no caso da Gamma, na Figura 9, é muito pouco provável que o máximo teórico exceda a sexta maior perda dos dados originais. Logo, o risco é subestimado.

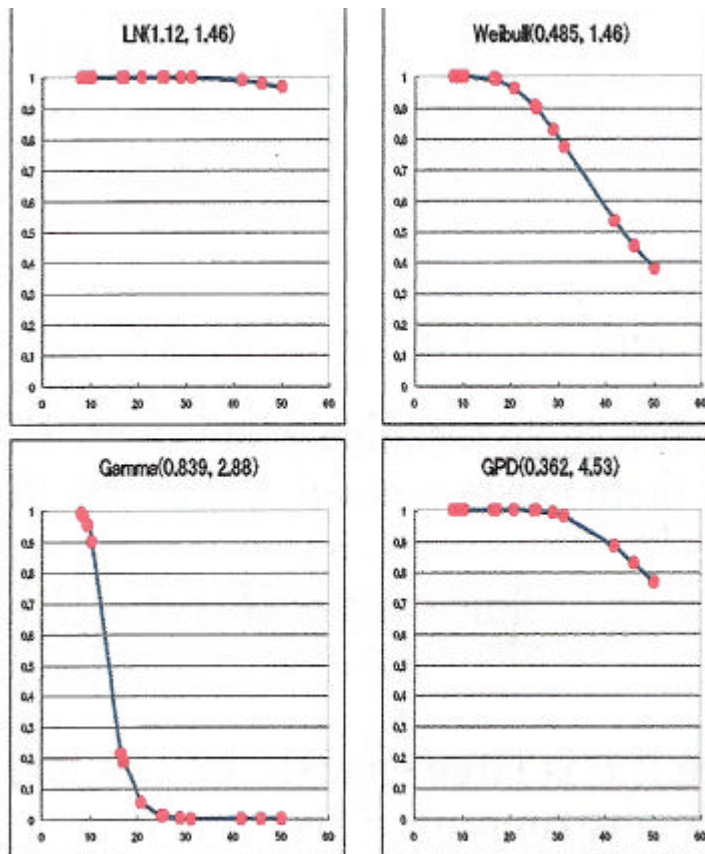


Figura 9 – Função sobrevivência dos dados originais da cauda.

4.2.4.3.

Teste de Kolmogorov-Smirnov (KS)

A estatística de teste do KS¹¹⁰ é o valor da maior discrepância entre a distribuição estimada e a distribuição empírica.

Portanto, este teste não leva em consideração se a distribuição estimada se ajusta bem aos dados como um todo.

Além disso, o poder deste teste é baixo, e existe uma tendência a superajustar os dados, especialmente para pequenas amostras.

Em relação aos modelos de severidade para risco operacional, este teste não é recomendado, uma vez que se foca em apenas um ponto no suporte da distribuição estimada, que pode não estar na cauda da distribuição, principal objetivo do ajuste das distribuições de severidade. Todavia, Chandra et al. (1981) [19] apresenta uma extensão a este teste para a distribuição GEV.¹¹¹

¹¹⁰ Este teste é apresentado em Cruz (2002) [26] 47-48.

¹¹¹ A descrição deste teste encontra-se em Cruz (2002) [26] 80-81.

4.2.4.4.

Teste de Anderson-Darling (AD)

Segundo Álvarez (2001) [1], este teste pode ser utilizado para os modelos de severidade, pois dá maior peso à cauda da distribuição. Este é um importante critério uma vez que a variação na cauda reflete a sensibilidade da distribuição.

Além disso, a poder do teste é maior que para o KS.¹¹²

4.3.

Modelos de Frequência

O objetivo desta seção é apresentar uma metodologia para encontrar a distribuição de probabilidade da variável aleatória que melhor se ajuste aos dados de frequência¹¹³ da base de dados de perdas operacionais.

A Figura 5, apresentada na seção anterior, descreve de forma sucinta esta metodologia, não só para os modelos de severidade, como também para os modelos de frequência.

A variável aleatória da frequência possui uma distribuição de contagem, que é uma distribuição discreta com probabilidade somente nos inteiros não-negativos. Esta distribuição descreve o número de eventos de perda em um determinado período.

Portanto, baseada na frequência dos eventos de perdas em intervalos de tempo de igual período, estimam-se os parâmetros de uma distribuição de contagem.

Entretanto, é necessário executar um ajuste na estimativa dos parâmetros, uma vez que a exposição ao risco varia ao longo do tempo, e, teoricamente, quanto maior a exposição ao risco maior o número de eventos de perdas operacionais. Este fenômeno é chamado de efeito da exposição, e o ajuste necessário será visto mais adiante.

¹¹² Informações mais específicas podem ser vistas em Cruz (2002) [26] 48-49.

¹¹³ A frequência é o número de eventos de perdas operacionais que ocorrem em um determinado período.

4.3.1.

Distribuições de Probabilidade

As distribuições de frequência podem ser divididas em três grupos de acordo com a classe a que pertencem:

- Ordinário: classe $(a, b, 0)$. São as distribuições de contagem usuais: Poisson, Binomial Negativa, Binomial e Geométrica. Ver apêndice C deste trabalho.
- Zero-truncado: classe $(a, b, 1)$ com $p_0 = 0$. São as distribuições aonde não há a possibilidade de frequência nula. Não fazem muito sentido para modelos de risco operacional. Ver apêndice D deste trabalho.
- Zero-modificado: classe $(a, b, 1)$ com $p_0 > 0$. São as distribuições aonde é arbitrado um valor para a probabilidade na frequência zero, e as probabilidades remanescentes são ajustadas de acordo. Ver apêndice D deste trabalho.

Quanto à composição, as distribuições podem ser classificadas em:

- Simples: não há composição de distribuições.
- Compostas¹¹⁴: as distribuições são obtidas para composição de uma distribuição com outra. Exemplo: Polya-Aeppli é a composição da Poisson com a Geométrica¹¹⁵.

Para risco operacional, normalmente se tem trabalhado com as seguintes distribuições ordinárias e simples:

- Poisson
- Binomial Negativa
- Binomial

4.3.1.1.

Poisson

É a distribuição de frequência mais popular nos modelos de risco operacional, apresentando a seguinte função de distribuição de probabilidade:

$$N \sim Po(I), \quad p(k) = \Pr(N = k) = \frac{e^{-I} I^k}{k!} \quad k = 0,1,2,K \quad (78)$$

¹¹⁴ Informações completas são vista em Klugman et al. (1998) [58].

¹¹⁵ A distribuição e os estimadores podem ser vistos em Cruz (2002) [26].

Fácil de estimar, pois tem apenas um parâmetro que é igual à média e à variância.

$$E[N] = Var[N] = I \quad (79)$$

A distribuição de contagem gerada por um processo estocástico homogêneo de Poisson apresenta duas características muito úteis¹¹⁶:

- Seja N_1, N_2, \dots, N_n variáveis independentes de Poisson com parâmetros $\lambda_1, \lambda_2, \dots, \lambda_n$. Então, $N = N_1 + N_2 + \dots + N_n$ forma uma distribuição de Poisson com parâmetro $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$. Assim, é fácil adicionar mais dados à base sem mudar estruturalmente as análises.
- Suponha que o número de eventos N é uma variável aleatória de Poisson com média λ . Considere que cada evento é classificado em uma das m classes de probabilidade p_1, p_2, \dots, p_m independente de todos os demais eventos. Logo, o número de eventos N_1, N_2, \dots, N_m , correspondentes aos eventos do tipo 1, 2, ..., m respectivamente, são variáveis aleatórias de Poisson com médias $\lambda p_1, \lambda p_2, \dots, \lambda p_m$. Logo, se uma distribuição de Poisson se ajustar à base de dados de perdas, ajustar-se-á também a mesma base truncada (com parâmetro diferente).

Entretanto, a hipótese de se assumir que o processo estocástico é homogêneo de Poisson implica que a probabilidade de uma perda ocorrer não depende do tempo passado desde a última perda.

Danielsson et al. (2001) [29] e Embrechts et al. (2002) [37] sugerem que esta probabilidade varia no tempo para perdas operacionais muito grandes. Neste caso, a assunção da Poisson não é tecnicamente correta.

Todavia, segundo de Fournouvel et al. (2003) [30] a variação no tempo da taxa de eventos de perdas resultaria em uma distribuição agregada de cauda pesada. Logo, o OPVaR estimado seria muito conservativo.

4.3.1.2.

Binomial Negativa

Esta distribuição tem sido usada alternativamente à distribuição de Poisson, pois possui dois parâmetros, tendo maior flexibilidade na forma.

A função de distribuição de probabilidade Binomial Negativa é dada por:

¹¹⁶ A demonstração destas características é encontrada em Klugman et al. (1998) [58].

$$N \sim BN(r, \mathbf{b}), \quad p(k) = \Pr(N = k) = \binom{k+r-1}{k} \left(\frac{1}{1+\mathbf{b}} \right)^r \left(\frac{\mathbf{b}}{1+\mathbf{b}} \right)^k, \quad (80)$$

$$k = 0, 1, 2, \dots, \quad r > 0, \mathbf{b} > 0$$

A média e variância da Binomial Negativa são apresentadas abaixo:

$$E[N] = r\mathbf{b} \quad \text{Var}[N] = r\mathbf{b}(1 + \mathbf{b}) \quad (81)$$

Da equação (81), e sendo r e β positivos, a variância é sempre maior que média.

Em Klugman et al. (1998) [58] é visto que se a taxa de eventos de perda for linearmente crescente ao longo do tempo, a distribuição de frequência deve ser a Binomial Negativa.

4.3.1.3.

Binomial

A distribuição Binomial descreve uma situação física na qual uma coleção de m fontes de risco está sujeita, cada uma, a um certo evento de perda. Assim, sejam m fontes de risco independentes e idênticas, onde para cada uma existe uma probabilidade q do evento de perda ocorrer. A probabilidade de ocorrerem exatamente k eventos de perda para esta coleção de fontes de risco é a função de distribuição de probabilidade Binomial, e é dada por:

$$N \sim B(m, q) \quad p(k) = \Pr(N = k) = \binom{m}{k} q^k (1-q)^{m-k}, \quad (82)$$

$$k = 0, 1, \dots, m, \quad 0 < q < 1$$

Esta distribuição tem o um suporte finito. Portanto, é uma boa opção para os casos aonde existe um limite para a quantidade de perdas. Por exemplo, um grande processamento é feito uma vez por mês. Logo, é impossível que este processamento trave mais que 12 vezes no ano.

A média e a variância da distribuição Binomial são dadas por:

$$E[N] = mq \quad \text{Var}[N] = mq(1-q) \quad (83)$$

Da equação (83), sendo m inteiro positivo e $0 < q < 1$, a variância é sempre menor que média.

Geralmente m é conhecido e fixo.

4.3.2.

Análise dos Dados

Esta seção apresenta uma ferramenta básica para escolha da distribuição subjacente a ser utilizada para estimação dos parâmetros do modelo de frequência. Todavia, a sofisticação do modelo de frequência está intimamente relacionada com o modelo de severidade, quando o objetivo é o cálculo do OPVaR.

Frachot et al. (2001) [42] apresenta uma análise de sensibilidade generalizada da relação do OPVaR e a distribuição de frequência. De acordo com sua experiência, existem dois tipos de riscos operacionais: aqueles com perdas potenciais de alta severidade, e aqueles com perdas de alta frequência e baixa severidade. Um erro na estimativa dos parâmetros da distribuição de frequência produz um pequeno impacto no OPVaR no primeiro caso, e um resultado totalmente diferente no segundo. A explicação desta análise é a seguinte:

- Caso 1: Se a distribuição de severidade for de cauda pesada, a probabilidade de dois (ou mais) eventos extremos de perdas é muito pequena (exceto se a frequência for muito alta).
- Caso 2: O OPVaR não é relacionado a um evento extremo de perda, mas sim a várias perdas de baixa severidade. Neste caso, a frequência assume o papel principal no cálculo do OPVaR.

Álvarez (2001) [1] confirma esta análise quando apresenta praticamente os mesmos resultados para o OPVaR, utilizando as distribuições Poisson, Binomial Negativa e Binomial ambas com a mesma média, e com algumas variâncias diferentes para cada distribuição (exceto Poisson, pois a média é igual a variância neste caso).

Desta forma, para o caso 1 o princípio da parcimônia prevalece, logo, a melhor escolha é a distribuição de Poisson.

Por outro lado, maior cuidado na modelagem da frequência deve ser despendido para o caso 2.

4.3.2.1.

Estatística Descritiva

As principais ferramentas utilizadas neste caso são as estatísticas descritivas da amostra, principalmente a média e a variância, pois apresentam indicadores na escolha da distribuição paramétrica.

A diferença básica entre as três distribuições descritas anteriormente reside na relação entre a média e variância teórica:

- Poisson: média e variância iguais.
- Binomial Negativa: variância maior que a média.
- Binomial: variância menor que a média.

4.3.3.

Estimação dos Parâmetros

O objetivo desta seção é apresentar detalhes dos métodos de estimação dos parâmetros das distribuições de frequência escolhidas como distribuições subjacentes.

Entretanto, há que se ter cuidado com a amostra utilizada para estimação dos parâmetros. Dependendo das questões envolvidas na amostragem, o modelo de frequência sofre alguns ajustes.

4.3.3.1.

Amostragem

Existem dois problemas fundamentais relacionados aos dados de frequência de eventos de perdas operacionais: efeito da exposição ao risco e mistura de dados internos e externos.

Estes problemas, bem como suas soluções, são vistas nos dois próximos tópicos.

4.3.3.1.1.

Efeito da Exposição

Considere que a unidade de risco consiste de n fontes. Seja N_j o número de eventos de perdas da j^{th} fonte. Então $N = N_1 + N_2 + \dots + N_n$. Assumindo-se que N_j são iid, logo:

$$P_N(z) = [P_{N_1}(z)]^n \quad (84)$$

Agora suponha que a unidade de risco espera expansão para n^* fontes, com a frequência N^* . Então:

$$P_{N^*}(z) = [P_{N_1}(z)]^{n^*} = [P_N(z)]^{n^*/n} \quad (85)$$

Logo, se N for infinitamente divisível¹¹⁷, as distribuições de N^* e N terão a mesma forma, apenas com diferenças nos parâmetros.

Todos os membros da classe $(a, b, 0)$, exceto a Binomial, têm esta propriedade. Para a classe $(a, b, 1)$, nenhum membro tem esta propriedade. Para as distribuições compostas, basta que a distribuição primária tenha a propriedade, que a composição também o terá.

Um segundo ajuste possível é quando os dados são coletados de fontes com exposições diferentes (geograficamente ou temporalmente).

Seja N a distribuição de frequência para uma exposição unitária, com seu vetor de parâmetros η . Suponha e_j um indicador da exposição (por exemplo, a receita bruta) para j^{th} fonte. Se a fonte j tem e_j exposições unitárias, e considerando $N \sim Po(\eta)$ para simplificar as contas, o número de perdas para esta fonte terá uma distribuição de Poisson com parâmetro ηe_j . Se n_j for o número de perdas para a fonte j , tem-se:

$$N_j \sim Po(\eta e_j) \quad p_j(k) = \frac{e^{-\eta e_j} (\eta e_j)^{n_j}}{n_j!} \quad (86)$$

Assim, utiliza-se a equação (86) para estimar o parâmetro da Poisson levando-se em consideração o efeito da exposição.

¹¹⁷ O conceito de divisibilidade infinita é muito importante não só para o efeito da exposição, mas também para justificar o fato de normalmente não se encontrar solução fechada para distribuição agregada. Este conceito é visto em detalhes em Klugman et al. (1998) [58].

4.3.3.1.2.

Mistura de Dados Internos e Externos

Considerando um banco específico, os seus dados internos para a frequência evidenciam a sensibilidade ao risco e as práticas de gerenciamento de risco dentro da instituição. Por exemplo, um banco cujas frequências dos eventos passados foram particularmente pequenas em relação a sua exposição, alocarão um OPVaR abaixo da média dos bancos. Todavia, nada garante que esta frequência abaixo da média resultou de um competente gerenciamento de risco ao invés de mera sorte no histórico de negócios. Se a distribuição se baseasse em frequências abaixo da média ocasionadas por sorte, o baixo OPVaR poderia estar mal dimensionado.

Portanto, a comparação dos dados internos com dados externos é uma solução para separar o que pode ser atribuído a uma boa prática de gerenciamento de risco, do que advém do decurso de sorte nos negócios.

Mas como tornar os dados compatíveis para uma comparação? A resposta é baseada em um modelo estatístico chamado de Teoria da Credibilidade¹¹⁸.

Segundo Klugman et al. (1998) [58], Credibilidade é um procedimento pelo qual os dados externos são combinados com os dados internos com o objetivo de obter melhor estimativa para perda esperada (ou qualquer outra estatística).

A seguir são apresentados três tópicos que descrevem o método para mistura dos dados internos e dos dados externos.

4.3.3.1.2.1.

Cálculo da Frequência Esperada dos Eventos

Considere uma unidade de risco de um determinado banco, e seja N_t o correspondente número de eventos do ano t .

Os dados históricos internos (em relação a dados de frequência) são representados por $\underline{N}_t = \{N_t, N_{t-1}, \dots\}$.

Seja EI o indicador de exposição desta unidade de risco neste banco, e assumamos que N_t é uma Poisson com parâmetro λEI , onde λ é um parâmetro (não observado) que mede o grau de risco do banco.

¹¹⁸ Informações detalhadas sobre a Teoria da Credibilidade são vistas em Klugman et al. (1998) [58].

O número de eventos esperados para o ano t deve ser $E[N_t | \underline{N}_{t-1}]$, e não $E[N_t]$, pois o histórico deve ser levado em consideração.

Considerando o risco, os bancos diferem uns dos outros pelo parâmetro θ e o indicador de exposição EI . De acordo com a teoria da credibilidade, assume-se que θ é uma variável aleatória não observável, que varia dos bancos de baixo risco (θ próximo de 0) a bancos de alto risco. Cada banco tem um específico grau de risco não observado θ , e seus dados internos \underline{N}_{t-1} fornecem as informações sobre este θ . Assuma que θ tem uma distribuição Gamma $\theta(a,b)$ dentro do segmento bancário:

$$f(\mathbf{I}) = \frac{\mathbf{I}^{a-1} e^{-\mathbf{I}/b}}{b^a \Gamma(a)} \quad (87)$$

Segundo Frachot & Roncalli (2002) [43], apesar da escolha da distribuição Gamma ser arbitrária neste caso, existem dois bons motivos:

- Esta classe de distribuições é grande e flexível o suficiente para capturar o tipo de heterogeneidade esperada.
- Além disso, facilita os cálculos e provê soluções fechadas.

Logo, a esperança incondicional (número de eventos esperados para unidade de risco do banco em questão com base em informações do segmento bancário) é dada por:

$$\begin{aligned} \mathbf{p}_t^0 &= E[N_t] \\ &= E_I [E[N_t | \mathbf{I}]] \\ &= EI \times E[\mathbf{I}] \end{aligned} \quad (88)$$

E a esperança condicional (número de eventos esperados para a unidade de risco do banco em questão) é dada por:

$$\begin{aligned} \mathbf{p}_t &= E[N_t | \underline{N}_{t-1}] \\ &= E[E[N_t | \mathbf{I}, \underline{N}_{t-1}] | \underline{N}_{t-1}] \\ &= EI \times E[\mathbf{I} | \underline{N}_{t-1}] \end{aligned} \quad (89)$$

De acordo com as equações (87), (88) e (89), e o apêndice A de Frachot & Roncalli (2002) [43], as equações (88) e (89) são explicitadas da seguinte forma:

$$\mathbf{p}_t^0 = a \times b \times EI \quad (90)$$

$$\mathbf{p}_t = w \times \mathbf{p}_t^0 + (1-w) \times \left(\frac{1}{t} \sum_{k=1}^t N_{t-k} \right) \quad (91)$$

, onde w é o fator de credibilidade que depende do indicador de exposição EI , do tempo t e do parâmetro b .

$$w = \frac{1}{1 + t \times b \times EI} \quad (92)$$

Pelas equações (90), (91) e (92) verifica-se que o número de eventos esperados é próximo da média histórica do número de eventos observados pelo banco, somente quando t é muito grande, isto é, a série histórica interna é bem extensa. Neste caso, o supervisor pode ficar suficientemente seguro em avaliar o risco do banco com base única e exclusivamente nos dados internos.

4.3.3.1.2.2.

Determinação da distribuição de probabilidade da frequência futura

A distribuição de probabilidade condicional de N_t (a posteriori) é a probabilidade de um banco com uma serie histórica \underline{N}_{t-1} experimentar n perdas em um tempo t – ver Klugman et al. (1998) [58]:

$$\Pr\{N_t = n \mid \underline{N}_{t-1}\} = \frac{\Gamma(\tilde{a} + n)}{\Gamma(\tilde{a})n!} (1 + \tilde{b})^{-\tilde{a}} \left(\frac{\tilde{b}}{1 + \tilde{b}} \right)^n \quad (93)$$

onde

$$\tilde{a} = a + \sum_{k=1}^t N_{t-k} \quad (94)$$

e

$$\tilde{b} = \frac{b \times EI}{1 + t \times b \times EI} \quad (95)$$

A distribuição de probabilidade incondicional de N_t (a priori) é a probabilidade de um banco com uma serie histórica desconhecida experimentar n perdas em um tempo t :

$$\Pr\{N_t = n\} = \frac{\Gamma(a + n)}{\Gamma(a)n!} (1 + b^0)^{-a} \left(\frac{b^0}{1 + b^0} \right)^n \quad (96)$$

onde

$$b^0 = b \times EI \quad (97)$$

4.3.3.1.2.3.

Calibração dos parâmetros

Os parâmetros a e b devem ser estimados com base nos dados de frequência do segmento bancário (exemplo, QIS^[4]). Seja $\{N_t^i\}$ um conjunto de

observações de dados de frequência, e EI o indicador de exposição ambos do i^{th} banco do segmento bancário. Então, o método da máxima verossimilhança pode ser utilizado na distribuição (incondicional) do numero de eventos, dada pela equação (96), para a estimação dos parâmetros a e b :

$$(a, b) = \arg \max \sum_i \left[\ln \Gamma(a + N_t^i) - \ln \Gamma(a) - (a + N_t^i) \ln(1 + b \times EI^i) + N_t^i \ln b \right] \quad (98)$$

Encontrados a e b , ajustam-se estes parâmetros (\tilde{a}, \tilde{b}) levando em consideração as frequências internas do passado, o tempo t e o indicador de exposição EI do banco em questão, utilizando para isto as equações (94) e (95).

4.3.3.2.

Métodos de Estimação

A estimação dos parâmetros da distribuição de frequência não apresenta nenhuma novidade ou particularidade em relação ao risco operacional.

Os métodos mais utilizados são o método da máxima verossimilhança e o método dos momentos.

4.3.4.

Critérios de Seleção

Os critérios de seleção para os modelos de frequência seguem o mesmo conceito básico dos critérios para os modelos de severidade.

A idéia é comparar as discrepâncias entre os dados originais e o que a distribuição estimada está predizendo.

4.3.4.1.

Teste Qui-Quadrado de Pearson

É usado para testar se uma amostra é proveniente de uma distribuição de probabilidade especificada.

O teste qui-quadrado é muito popular, e basicamente compara o resultado real com a previsão. A estatística de teste é dada por:

$$Q = \sum_{k=0}^n \frac{(n_k - E_k)^2}{E_k} \quad E_k = n\hat{p}_k = n\Pr(N = k; \hat{\mathbf{q}}) \quad (99)$$

, onde E_k é o valor esperado da quantidade de eventos k , e n_k é a quantidade real de eventos k .

A hipótese nula é rejeitada se Q exceder $\chi^2_{d,\alpha}$ com $d = k - r - 1$ graus de liberdade, onde k é o tamanho da amostra, r o número de parâmetros e α é o nível de significância do teste.