

# Modelagem semiparamétrica de retornos financeiros através da teoria do valor extremo<sup>☆</sup>

Elsevier Inc<sup>a,b</sup>, Rafael Felipe Bressan<sup>a,1,\*</sup>

<sup>a</sup> Avenida Madre Benvenuta, 2007 - Santa Mônica Florianópolis - SC 88035-901

<sup>b</sup> 360 Park Avenue South, New York

---

## Abstract

This template helps you to create a properly formatted L<sup>A</sup>T<sub>E</sub>X manuscript.

*Keywords:* elsarticle.cls, L<sup>A</sup>T<sub>E</sub>X, Elsevier, template

*2010 MSC:* 00-01, 99-00

---

*Teoria do valor extremo.* EVT da sigla em inglês, é um ramo da estatística que lida diretamente com eventos raros, extremos. Seu objetivo é modelar eventos que se distanciam muito da mediana de uma distribuição. Justamente por esta característica, a EVT está sendo utilizada para modelar riscos que possuem  
5 distribuição com caudas longas, um dos fatos estilizados que apresentamos sobre retornos de ativos financeiros.

Ao utilizarmos a EVT, e mais especificamente o método conhecido como *peaks over threshold* – POT, estamos interessados em modelar apenas a parte da cauda da distribuição das perdas de um ativo financeiro maiores que um  
10 determinado valor de limiar  $u$ . É da modelagem desta cauda, portanto, que faremos as estimativas de risco  $VaR_\alpha$  e  $ES_\alpha$ .

---

<sup>☆</sup>Fully documented templates are available in the elsarticle package on CTAN.

<sup>\*</sup>Corresponding author

Email address: rafael.bressan@edu.udesc.br (Rafael Felipe Bressan)

URL: [www.elsevier.com](http://www.elsevier.com) (Elsevier Inc)

<sup>1</sup>Depto. de Economia/Esag/UDESC

## 1. Função de distribuição de valores extremos generalizada

Consideremos uma amostra de uma variável aleatória cujas observações sejam independentes e igualmente distribuídas (*iid rv*)  $L_i, i \in \mathbb{N}$  que represente as  
15 perdas financeiras de um determinado ativo.

A EVT está interessada em investigar o comportamento da distribuição dos máximos desta *iid rv* dados por  $M_n = \max(L_1, \dots, L_n)$  para vários valores de  $n$  e a medida que  $n \rightarrow \infty$ . A sequência  $M_n$  é chamada de máximos em bloco e é possível demonstrar que a única distribuição para a qual  $M_n$  converge com  
20  $n \rightarrow \infty$  é a distribuição de valores extremos generalizada (GEV, da sigla em inglês).

Para tanto, é necessário normalizarmos esta sequência de máximos de forma que sua distribuição seja convergente para uma distribuição  $H(x)$  não-degenerada<sup>2</sup>. Seja  $F(x)$  a distribuição original de uma variável aleatória *iid*, é  
25 possível normalizar seus máximos em bloco através da relação  $M_n^* = (M_n - d_n)/c_n$  de forma que:

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - d_n}{c_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(c_n x + d_n) = H(x) \quad (1)$$

Em outras palavras, para determinadas sequências  $c_n$  e  $d_n$  a serem escolhidas, existe uma distribuição de  $H(x)$  não-degenerada a qual representa a distribuição dos máximos em bloco de  $F(x)$ .

30 A potenciação de  $F$  em  $n$  deriva diretamente da suposição que a variável aleatória é *iid*, enquanto que a transformação de  $x \rightarrow c_n x + d_n$  é a normalização dos máximos em bloco.

**Definição 1** (MDA). Domínio de máxima atração, se a equação (1) é válida para uma  $H$  não-degenerada, então se diz que  $F \in MDA(H)$ ,  $F$  pertence ao  
35 domínio de máxima atração de  $H$ .

---

<sup>2</sup>Distribuição degenerada é aquela cuja densidade de probabilidade está totalmente concentrada em apenas um ponto de seu domínio.

**Teorema 2** (Fisher-Tippett, Gnedenko). Se  $F \in MDA(H)$  para alguma  $H$  não-degenerada, então  $H$  deve ser uma distribuição do tipo de valores extremos generalizada – GEV.

O teorema 2 foi estabelecido através de três artigos, Fisher & Tippett (1928),  
40 Gnedenko (1941, 1943).

**Definição 3** (GEV). Distribuição de valores extremos generalizada, é definida por sua p.d.f (função densidade de probabilidades) a qual é dada por:

$$H_{\xi}(x) = \begin{cases} \exp(-(1 + \xi x)^{-\frac{1}{\xi}}), & \xi \neq 0, \\ \exp(-e^{-x}), & \xi = 0, \end{cases} \quad (2)$$

O parâmetro  $\xi$  (leia-se qsi) é conhecido como o parâmetro de forma da distribuição e dependendo deste valor teremos diferentes tipos de distribuição  
45 (casos particulares da GEV). Quando  $\xi = 0$  a distribuição resultante é uma Gumbel, quando  $\xi > 0$  uma Fréchet surge, e por fim quando  $\xi < 0$  temos uma Weibull.

Tomemos como exemplo a distribuição exponencial e calcularemos seu MDA e verificaremos se este está entre umas das distribuições GEV. Uma distribuição  
50 exponencial é caracterizada pela seguinte função de distribuição (c.d.f):

$$F(x) = 1 - e^{-\beta x}, \beta > 0 \text{ e } x \geq 0$$

Se escolhermos as sequências  $c_n = 1/\beta$  e  $d_n = \ln n/\beta$  podemos substituir diretamente na equação (1) e calcular  $H(x)$ .

$$F^n(c_n x + d_n) = \left(1 - \frac{1}{n} e^{-x}\right)^n$$

$$\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n} e^{-x}\right)^n = H(x)$$

Fazendo uma simples substituição de variáveis,  $i = -e^{-x}$ , então:

$$H(x) = \lim_{n \rightarrow \infty} \left(1 + \frac{i}{n}\right)^n$$

Que é o limite fundamental  $e^i$ , o qual substituindo novamente  $i$  temos:

$$H(x) = \exp(-e^{-x}) = H_0(x), \text{Distribuição Gumbel}$$

55 Ou seja, a distribuição exponencial pertence ao *MDA* da distribuição Gumbel, a qual por sua vez é um dos casos particulares da GEV quando  $\xi = 0$ .

Para este trabalho não necessitamos calcular a qual *MDA* pertencem nossas distribuições, bastando saber que basicamente todas as distribuições contínuas de utilidade prática estão contidas em  $MDA(H_\xi)$  para algum valor de  $\xi$ .

## 60 2. Excessos acima de um limiar

O método conhecido como POT, para calcular a função de distribuição dos valores que excedem um determinado limiar de um conjunto de dados vem sendo empregado no mundo financeiro para ajustar as caudas das distribuições de retornos, ou perdas, dos ativos. Este método é preferido a teoria clássica de 65 valores extremos (e.g. máximos em bloco), pois, desperdiça uma quantidade menor de dados da série original. Qualquer valor que exceda o limiar pré-determinado é considerado na distribuição dos excessos. Esta distribuição dos valores da série que estão acima de um determinado limiar  $u$  é definida como:

**Definição 4** (Distribuição dos excessos). Seja  $X$  uma variável aleatória com 70 função de distribuição c.d.f  $F$ . A distribuição dos excessos sobre um limiar  $u$  tem a seguinte função de distribuição:

$$F_u(x) = P(X - u \leq x | X > u) = \frac{F(x + u) - F(u)}{1 - F(u)} \quad (3)$$

para  $0 \leq x < x_F - u$ , onde  $x_F \leq \infty$  é o limite direito da distribuição  $F$ .

Ou seja, a função distribuição dos excessos sobre um limiar  $u$  é a probabilidade condicional que um valor  $X$  retirado dos dados subtraído de  $u$  (o excesso) seja

75 menor que um dado quantil  $x$ , sabendo-se que  $X$  é maior que  $u$ . Uma importante distribuição que surge na modelagem dos excessos sobre um limiar é a distribuição de pareto generalizada – GPD, que segue.

**Definição 5** (GPD). Distribuição de Pareto Generalizada é definida por sua função de distribuição:

$$G_{\xi, \beta(u)}(X) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta(u)}\right)^{-\frac{1}{\xi}}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{x}{\beta(u)}\right), & \xi = 0, \end{cases} \quad (4)$$

80 onde  $\beta > 0$ , e  $x \geq 0$  quando  $\xi \geq 0$  ou  $0 \leq x \leq -\beta/\xi$  quando  $\xi < 0$ .

Os parâmetros  $\xi$  e  $\beta$  são conhecidos respectivamente como parâmetros de forma e escala da distribuição. Na figura 1 abaixo, são mostradas três parametrizações para a função de distribuição acumulada (c.d.f) e para a densidade de probabilidades (p.d.f) de GPD com parâmetro  $\xi$  iguais a -0,5, 0 e 0,5 enquanto  
85 que o parâmetro de escala  $\beta$  é mantido constante e igual a 1. Perceba como para  $\xi < 0$  a p.d.f tem um limite direito que é dado por  $-\beta/\xi$  a partir do qual os valores de  $g(x)$  são zero.

A distribuição de Pareto generalizada tem papel fundamental na teoria de valor extremo em função do teorema de Pickands-Balkema-de Haan (Pickands  
90 (1975) e Balkema & de Haan (1974)) conforme abaixo:

**Teorema 6** (Pickands-Balkema-de Haan). Pode ser encontrada uma função  $\beta(u)$  tal que:

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0$$

se e somente se  $F \in MDA(H_\xi)$  para  $\xi \in \mathbb{R}$ .

O que este teorema nos diz é que para distribuições as quais os máximos em bloco normalizados convergem para uma GEV (na forma da equação (2)), então a distribuição dos excessos acima de um limiar destas mesmas distribuições  
95 convergem para uma GPD, dado um valor de limiar  $u$  adequado.

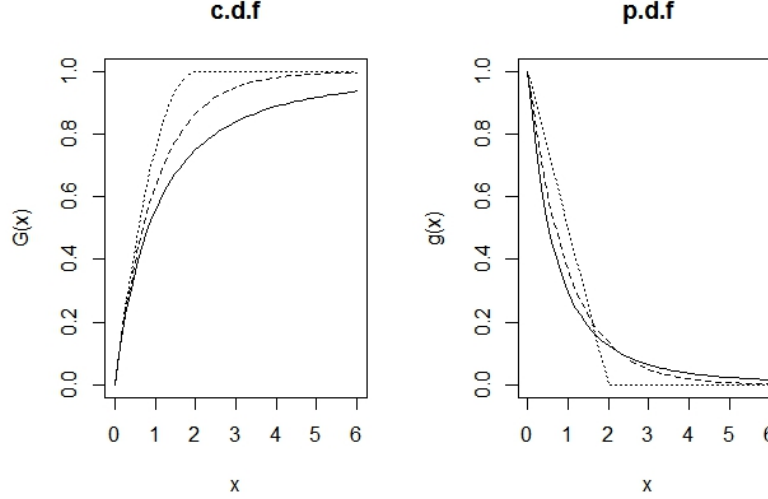


Figure 1: Três parametrizações para uma GPD. A linha sólida corresponde a  $\xi = 0,5$ , a linha pontilhada a  $\xi = -0,5$  e a linha tracejada a  $\xi = 0$ .

**Suposição 7.** Seja  $F$  a distribuição de perdas com limite direito  $x_F$ , assumamos que para um valor limiar alto o suficiente  $u$  nós temos que  $F_u(x) = G_{\xi,\beta}(x)$ , onde  $F_u(x)$  denota a distribuição dos excessos de  $x$  em relação ao valor de limiar  $u$ , para  $0 \leq x < x_F - u$ ,  $\xi \in \mathbb{R}$  e  $\beta > 0$ .

100 Esta é uma suposição, uma vez que a distribuição dos excessos não segue exatamente uma GPD, mas apenas tende a esta distribuição dado um limiar  $u$  alto e uma amostra de dados grande o suficiente.

Dada a parametrização de uma GPD, é interessante sabermos o valor esperado desta distribuição, uma vez que esta medida de valor central nos fornece  
105 importante informação sobre a quantidade de risco que estamos buscando medir, assim como a informação de que a própria distribuição foi ajustada aos dados de forma satisfatória, como será demonstrado adiante.

O valor esperado de uma variável aleatória não negativa pode ser computado através da integral de sua cauda,  $P(X > x) = 1 - P(X \leq x)$ . A cauda da GPD  
110 é, para  $\xi \neq 0$ ,  $(1 + \xi x/\beta(u))^{-1/\xi}$

Bastando, portanto, integrar em relação a  $x$  sobre o domínio deste, que é de

0 a  $\infty$ .

$$\int_0^{\infty} (1 + \xi x / \beta(u))^{-1/\xi} dx$$

Desta forma, o valor esperado de uma GPD  $G_{\xi, \beta(u)}(X)$ , ou seja, sua média, converge para valores de  $\xi < 1$  e é dado pela seguinte equação:

$$E[G_{\xi, \beta(u)}(X)] = \frac{\beta(u)}{1 - \xi} \quad (5)$$

**Definição 8** (função média dos excessos). A função média dos excessos de uma variável aleatória  $X$  com média finita é dada por:

$$e(u) = E(X - u | X > u) \quad (6)$$

Ou seja, a equação (6) representa o valor esperado da função de distribuição dos excessos dada pela Definição 4. Ela representa a média de  $F_u$  como uma função do limiar  $u$ . Esta função por vezes também é conhecida como função média de vida residual (*mean residual life function*), sendo encontrada esta denominação em alguns pacotes de *software* estatísticos.

Para uma variável distribuída na forma de uma GPD, o parâmetro de escala é uma função linear em  $u$  dado por  $\beta(u) = \beta + \xi u$ , (Embrechts et al., 1997, Teorema 3.4.13(e)). Utilizando-se deste fato e da equação (5) chegamos ao cálculo da função média dos excessos para uma GPD, dada por:

$$e(u) = \frac{\beta + \xi u}{1 - \xi} \quad (7)$$

onde  $0 \leq u < \infty$  se  $0 \leq \xi < 1$  e  $0 \leq u \leq -\beta/\xi$  se  $\xi < 0$ . É possível observar que de fato a função média dos excessos em uma GPD é linear em  $u$ . Esta é uma característica importante de uma GPD e que nos auxilia a escolher um valor adequado do limiar  $u$  de tal forma que a Suposição 7 faça sentido.

Assim, quando estamos analisando uma determinada distribuição de perdas  $F$  e desejamos ajustar a cauda desta distribuição, ou seja, as perdas acima de um

dado valor limiar  $u$  a uma GPD  $G_{\xi,\beta}(x)$  precisamos primeiramente determinar um valor adequado de  $u$  de modo que a Suposição 7 de que  $F_u(x) \rightarrow G_{\xi,\beta}(x)$  seja válida. Um método frequentemente utilizado é o gráfico da função média dos excessos com relação a  $u$ . Analisando este gráfico, escolhemos o menor valor  
135 dos excessos com relação a  $u$ . A partir deste ponto a relação  $e(u)$  vs  $u$  torna-se linear.

Desejamos o menor valor de  $u$  para o qual a relação é linear pois, mesmo o método POT implica em grande perda de dados da série temporal, já que apenas os valores acima deste limiar são utilizados para fazer a estimação dos  
140 parâmetros  $\xi$  e  $\beta$  da GPD. Portanto, existe um *trade-off* na escolha do valor limiar  $u$ , escolhendo um valor muito baixo temos uma boa quantidade de dados para estimar os parâmetros da GPD, mas a própria distribuição resultante não será GPD, uma vez que não estaremos trabalhando na região onde a relação  $e(u)$  vs  $u$  é linear. Por outro lado, um valor limiar muito alto nos impõe o  
145 custo de trabalhar com poucos dados para fazer a estimação dos parâmetros da distribuição e por conseguinte, os erros padrões dessas estimativas serão elevados.

**Lema 9.** Sob a Suposição 7 segue que  $F_v(x) = G_{\xi,\beta+\xi(v-u)}(x)$  para qualquer valor limiar  $v \geq u$ .

Logo, a distribuição dos excessos sobre limiares mais altos que  $u$ , também  
150 segue uma GPD com o mesmo parâmetro de forma  $\xi$  e parâmetro de escala que cresce linearmente com este limiar mais alto  $v$ . Se  $\xi < 1$ , a média desta nova GPD converge e é dada por:

$$e(v) = \frac{\beta + \xi(v - u)}{1 - \xi} = \frac{\xi v}{1 - \xi} + \frac{\beta - \xi u}{1 - \xi} \quad (8)$$

Esta é a função média dos excessos sobre limiares mais altos, e está definida para  $u \leq v < \infty$  se  $0 \leq \xi < 1$  e,  $u \leq v \leq u - \beta/\xi$  se  $\xi < 0$ .

155 Esta função é muito útil para calcularmos o  $ES_\alpha$  (*expected shortfall*), considerando que  $VaR_\alpha$  nada mais é que um quantil superior ao limiar  $u$  escolhido.



### 3. Modelando caudas e medidas de risco associadas

Através da modelagem da cauda da distribuição  $F$  de perdas por uma GPD, como feito na seção anterior, é possível calcularmos as medidas de riscos  $160$   $Var_\alpha$  e  $ES_\alpha$  para esta distribuição de perdas em função dos parâmetros da GPD estimada e também fazendo uso da distribuição empírica de  $F$ .

Sob a Suposição 7 nós temos que a cauda da distribuição  $F$ ,  $\bar{F}(x)$ , para  $x \geq u$  é dada por:

$$\begin{aligned}\bar{F}(x) &= P(X > u)P(X > x|X > u) \\ &= \bar{F}(u)P(X - u > x - u|X > u) \\ &= \bar{F}(u)\bar{F}_u(x - u) \\ &= \bar{F}(u) \left(1 + \xi \frac{x - u}{\beta}\right)^{-1/\xi}\end{aligned}\tag{9}$$

Da qual se soubéssemos exatamente a distribuição  $F$  teríamos um modelo  $165$  analítico para as probabilidades de perdas na cauda da distribuição. Aqui  $x$  são os valores a serem observados das perdas, e portanto  $x - u$  são as perdas em excesso ao limiar.

O que fizemos através da equação (9) foi efetivamente separar a distribuição  $F$ , ou melhor, sua cauda, em duas partes. A primeira parte, para valores menores  $170$  que  $u$ , não foi modelado analiticamente e portanto utilizamos a distribuição empírica das perdas, aqui representada por sua cauda  $\bar{F}(u)$ , que nada mais é que o número observado de excessos de  $u$  sobre o número total de observações da amostra.

A segunda parte é justamente a modelagem através de uma GPD com  $175$  parâmetros  $\xi$  e  $\beta$  dado o limiar  $u$ . Por esta modelagem paramétrica podemos conhecer as probabilidades de cauda para valores de  $x$  maiores que  $u$ .

O quantil  $\alpha$  é a inversa da função distribuição e nos retorna o valor para o qual um percentual  $\alpha$  de observações da amostra é menor ou igual. Assim sendo,  $Var_\alpha$  nada mais é que um quantil alto para o qual determinamos que  $\alpha\%$  das

180 perdas devem ser menores ou iguais a este valor.

Como a equação (9) fornece a probabilidade de cauda, então esta é igual a  $1 - \alpha$  para um valor de  $\alpha \geq F(u)$ . Fazendo  $\bar{F}(x) = 1 - \alpha$  na equação (9) o valor de  $x$  representará  $VaR_\alpha$  e nos basta manipular esta equação até isolarmos  $VaR_\alpha$  como função de  $\bar{F}(u)$ ,  $\alpha$  e dos parâmetros da GPD  $\xi$  e  $\beta$ . Que nos garante a  
185 equação abaixo:

$$VaR_\alpha = q_\alpha(F) = u + \frac{\beta}{\xi} \left[ \left( \frac{1 - \alpha}{\bar{F}(u)} \right)^{-\xi} - 1 \right] \quad (10)$$

A medida  $ES_\alpha$  pode ser entendida como a média das perdas que excedem o valor dado por  $VaR_\alpha$ . Como o próprio  $VaR_\alpha$  é um quantil acima do valor de limiar  $u$ ,  $ES_\alpha$  é dado pelo valor do  $VaR_\alpha$  somado a função média dos excessos dada pela equação (8) fazendo  $v = VaR_\alpha$ . Esta média é convergente para valores  
190 de  $\xi < 1$  conforme já demonstrado. Ou seja,  $ES_\alpha = VaR_\alpha + e(VaR_\alpha)$ . A qual nos rende de forma mais geral:

$$ES_\alpha = \frac{VaR_\alpha}{1 - \xi} + \frac{\beta - \xi u}{1 - \xi} \quad (11)$$

Portanto, ambas medidas de risco  $VaR_\alpha$  e  $ES_\alpha$ , para distribuições de perdas que tiveram suas caudas modeladas através de uma GPD da forma  $G_{\xi, \beta(u)}$  com  $\xi < 1$  e  $\beta > 0$ , podem ser calculadas respectivamente através das equações  
195 dadas em (10) e (11). As estimativas destas medidas de risco serão encontradas através das estimativas dos parâmetros da GPD, assim como do limiar utilizado e de uma medida empírica de  $\bar{F}(u)$  que será o número de excessos verificados sobre o total de amostras. É claro que, ao adotarmos esta estimativa para  $\bar{F}(u)$  estamos implicitamente supondo que o número de amostras na série de perdas é  
200 significativa, assim como o número de excessos verificados. Daí a importância de se utilizar um valor  $u$  adequado, conforme explicitado na seção 2.

As estimativas de medidas de risco desenvolvidas nesta seção se qualificam como medidas incondicionais, no sentido que elas não dependem do estado atual das coisas, mas sim de todo o histórico de eventos de forma uniforme. Em

Incluir sobre Smith e  $\bar{F}(u)$

205 outras palavras,  $Var_\alpha$  e  $ES_\alpha$  derivados a partir das equações (10) e (11) são medidas históricas de risco associado ao ativo em análise, entretanto, não levam em consideração se nos eventos mais recentes a volatilidade das perdas pode ser diferente do valor histórico.

De fato, uma das características marcantes das perdas (ou retornos, como  
210 o leitor preferir) dos ativos financeiros é o chamado *clustering* de volatilidade, onde grandes volatilidades (retornos positivos ou negativos) têm tendência a ficarem próximas ao longo da linha temporal. Em geral estas aglomerações de volatilidades surgem a partir da autocorrelação destas, ou seja, a volatilidade em um período  $t$  é dependente das volatilidades verificadas em períodos ante-  
215 riores. Um modelo bastante encontrado na literatura que busca modelar estas dependências é o modelo ARCH e suas variantes como GARCH<sup>3</sup>.

Assim, ao passo que as estimativas de risco desenvolvidas nesta seção são valiosas para prazos mais longos, ainda é necessário desenvolver um modelo que lide com o fato das autocorrelações de volatilidades e portanto, que nossa  
220 variável aleatória não é independente e igualmente distribuída ao longo do tempo. O modelo proposto por McNeil & Frey (2000) pode ser utilizado para encontrar as medidas de risco  $Var_\alpha$  e  $ES_\alpha$  condicionais que desejamos, ainda dentro da metodologia de *peaks over threshold* - POT.

### 3.1. Medidas condicionais de risco

225 Ativos financeiros possuem características de autocorrelação, senão em seus retornos propriamente ditos, ao menos em suas volatilidades ou variações absolutas. Ou seja, dada uma grande variação no momento  $t$  é de se esperar novamente uma grande variação, não necessariamente na mesma direção daquela anterior, para o momento  $t+1$  e posteriores. Desta forma, medidas de risco  
230 incondicionais, conforme aquelas derivadas na Seção 3 podem ser adequadas somente para horizontes temporais mais longos, pois implicitamente tomam em

---

<sup>3</sup>O modelo ARCH foi introduzido primeiramente por Engle (1982) e sua generalização GARCH por Bollerslev (1986)

consideração os fatos mais recentes com o mesmo valor de predição que fatos mais longínquos.

Também já foi bastante estudado e demonstrado no mesmo Capítulo que mod-

Revisão

235 elos que levem em conta riscos condicionais ao incorporarem as autocorrelações nas volatilidades, levam a resultados de testes melhores. Assim, nesta subseção trabalharemos com o modelo proposto por McNeil & Frey (2000) os quais fazem uma adequação dos retornos dos ativos a um modelo GARCH e posteriormente tratam os erros desta modelagem como *iid* e portanto, a metodologia de POT e  
240 ajuste de uma GPD pode ser feito.

Primeiramente vamos estabelecer um modelo GARCH para as perdas do ativo subjacente. Se denotarmos  $L_t$  como sendo a perda observada no período  $t$ ,  $\mu_t$  e  $\sigma_t$  são respectivamente a média e o desvio padrão condicionais e mensuráveis através do conjunto de informações disponíveis em  $t-1$  e seja  $Z_t$  inovações *iid*  
245 com média zero e desvio padrão unitário, então temos que:

$$L_t = \mu_t + \sigma_t Z_t \quad (12)$$

Seja  $F_L(l)$  a distribuição marginal de  $L_t$ , então  $F_{L_{t+1}|\mathcal{G}_t}(l)$  é a distribuição preditiva da perda para o próximo período, onde  $\mathcal{G}_t$  é o conjunto de informações disponíveis no período  $t$ , incluindo-o. Portanto, para o cálculo das medidas condicionais de risco estamos interessados em um quantil  $\alpha$  na cauda de  $F_{L_{t+1}|\mathcal{G}_t}(l)$ .  
250 Este quantil  $\alpha$ , que será o nosso  $Var_\alpha$ , é o ínfimo  $l$  tal que o valor da distribuição preditiva seja maior ou igual a  $\alpha$ . Ao passo que o valor condicional do *ES* será o valor esperado das perdas previstas que sejam maiores que VaR para o mesmo intervalo de confiança. Ou seja:

$$Var_\alpha^t = \inf\{l \in \mathbb{R} : F_{L_{t+1}|\mathcal{G}_t}(l) \geq \alpha\}, \quad (13)$$

$$ES_\alpha^t = E[L_{t+1} | L_{t+1} > Var_\alpha^t] \quad (14)$$

Considerando que nossa distribuição de perdas é dada pela equação (12) e

255 sabendo das propriedades de variáveis aleatórias e do operador de expectância,  
as equações dadas acima subsumem a:

$$VaR_{\alpha}^t = \mu_{t+1} + \sigma_{t+1} z_{\alpha}, \quad (15)$$

$$ES_{\alpha}^t = \mu_{t+1} + \sigma_{t+1} E[Z|Z > z_{\alpha}] \quad (16)$$

onde  $z_{\alpha}$  é o quantil  $\alpha$  das inovações  $Z$ .

Agora nos falta escolher um processo que modele nossa série temporal dada em (12), ou seja, precisamos especificar o comportamento de  $\mu_t$  e  $\sigma_t$ . Por suposição  
260 do modelo, especificamos que o comportamento destas variáveis é dependente de acontecimentos passados, contidos no conjunto de informações  $\mathcal{G}_{t-1}$ . Dentre os diversos modelos já propostos para estimar médias e volatilidades condicionais, está o simples porém efetivo modelo GARCH(1,1) para a volatilidade condicional e o modelo AR(1) para a média condicional. Uma extensão destes modelos  
265 básicos para outros mais complexos pode ser facilmente obtida e é vasta na literatura, como exemplo modelos GARCH-M, *Threshold* GARCH, EGARCH, etc. e modelos do tipo ARMA, ARIMA, entre outros encontrados em Tsay (2010).

Neste trabalho, visando aplicar a teoria do valor extremo para o cálculo das medidas condicionais de risco, não faremos maiores suposições acerca da  
270 distribuição das inovações, como por exemplo assumir uma determinada distribuição (e.g. Normal ou t de Student), mas está implícito que esta pertence ao MDA de uma GEV e portanto a distribuição de seus excessos sobre um limiar segue aproximadamente uma GPD.

Dadas estas considerações, o modelo adotado segue um formato em dois  
275 estágios para ser implementado, como segue.

- Ajustar um modelo AR-GARCH para os dados de perdas, sem fazer suposições sobre a distribuição de  $Z$ , porém, utilizando um método de pseudo máxima verossimilhança (PML). Deste modelo tiramos as estimativas de  $\mu_{t+1}$  e  $\sigma_{t+1}$  e portanto, calculamos as inovações implícitas resultantes

Implementar  
com modelo  
EGARCH

280 através da equação (12).

- Consideramos estas inovações calculadas como sendo as realizações da variável aleatória  $Z$ , a qual por suposição pode ter sua cauda ajustada a uma GPD utilizando o método descrito na seção 3. Encontraremos por fim o valor de  $z_\alpha$  e  $E[Z|Z > z_\alpha]$ , com os quais finalizamos os cálculos de  
285 nossas medidas condicionais de risco dadas em (15) e (16).

Nosso modelo completo para as medidas de risco  $VaR_\alpha$  e  $ES_\alpha$  condicionais dada a distribuição de perdas  $L_t$  de um ativo será, portanto:

$$L_t = \mu_t + \epsilon_t \quad (17)$$

$$\mu_t = \mu + \phi_1 \mu_{t-1} \quad (18)$$

$$\epsilon_t = \sigma_t Z_t \quad (19)$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (20)$$

$$Z_t \sim \mathcal{D}(0, 1) \text{ e } \mathcal{D} \in MDA(H_\xi) \quad (21)$$

com as equações (15) e (16) nos fornecendo os valores das medidas  $VaR_\alpha$  e  $ES_\alpha$  respectivamente, quando utilizadas em conjunto com aquelas dadas por (10) e  
290 (11).

O termo de erro  $\epsilon_t$  será uma série estritamente estacionária se  $\beta_1 + \alpha_1 < 1$ , o que garante que a variância de nosso modelo ( $\sigma_t^2$ ) será finita, apesar de variável no tempo.

Este modelo é ajustado utilizando-se o método de pseudo máxima verossimil-  
295 hança, o que significa que, apesar de não adotarmos a distribuição normal para reger as inovações  $Z_t$ , a utilizamos apenas para estimar os parâmetros do modelo. É demonstrado na literatura, (Gouriéroux, 1997, Capítulo 4), que o método PML fornece estimadores consistentes e assintoticamente normais, assim podemos utilizar tal método para obter nossas estimativas de parâmetros e erros padrão,  
300 mesmo não aceitando a condição de normalidade das inovações.

**Exemplo 10** (Risco condicional da Apple). Como exemplo tomaremos os dados de perdas diárias da Apple Inc. (*ticker* AAPL) compreendidas no período entre 02/01/2009 a 16/09/2016, 1.941 observações.

Seguindo o procedimento em 2 estágios, primeiramente iremos ajustar a  
305 série de perdas diárias da Apple ao modelo AR(1)-GARCH(1,1) descrito acima, considerando a distribuição condicional como uma normal.

A figura 2 abaixo nos mostra alguns resultados desta regressão. Seguindo os gráficos da esquerda para a direita, acima e depois abaixo, temos no primeiro gráfico as perdas observadas envoltas em uma banda composta por  $\pm 2\sigma$ . No  
310 segundo gráfico temos a autocorrelação das próprias observações, ou seja, das perdas. Nota-se que estatisticamente não se pode afirmar que estas sejam correlacionadas e portanto, seria de se esperar que o valor de  $\phi_1$  em nosso modelo não seja significativo. No terceiro gráfico temos as autocorrelações dos quadrados das perdas, e neste caso sim, de acordo com os fatos estilizados de  
315 ativos financeiros, estes são altamente correlacionados.

Enquanto os três primeiros gráficos nos trazem informações da série de perdas original, a sequência na parte de baixo da figura 2 mostra informações acerca do modelo AR-GARCH ajustado. O primeiro deles é um gráfico do tipo QQ (quantil-quantil) muito utilizado para verificar a distribuição de variáveis  
320 aleatórias. Nele temos uma distribuição de referência, neste caso a Normal e, se a variável de interesse segue esta distribuição os pontos no gráfico QQ devem ficar todos sobre a linha contínua. Podemos perceber claramente que ao nos aproximarmos das caudas, os resíduos padronizados,  $z_t$ , se desviam consideravelmente da normalidade, apresentando caudas muito mais longas que  
325 uma distribuição normal. Para o nosso modelo isso era esperado, uma vez que sabemos que as inovações em processos estocásticos de ativos financeiros não seguem uma normal. São os excessos acima de um limiar destes resíduos que, no segundo passo de nossa modelagem, ajustaremos a uma distribuição do tipo GPD.

330 Ainda assim é possível observar que a filtragem dos dados promovida pelo

ajuste AR-GARCH nos serviu para descorrelacionar serialmente os resíduos padronizados, conforme demonstram os dois gráficos seguintes, de forma a ser possível utilizar a técnica de valor extremo a estes resíduos.

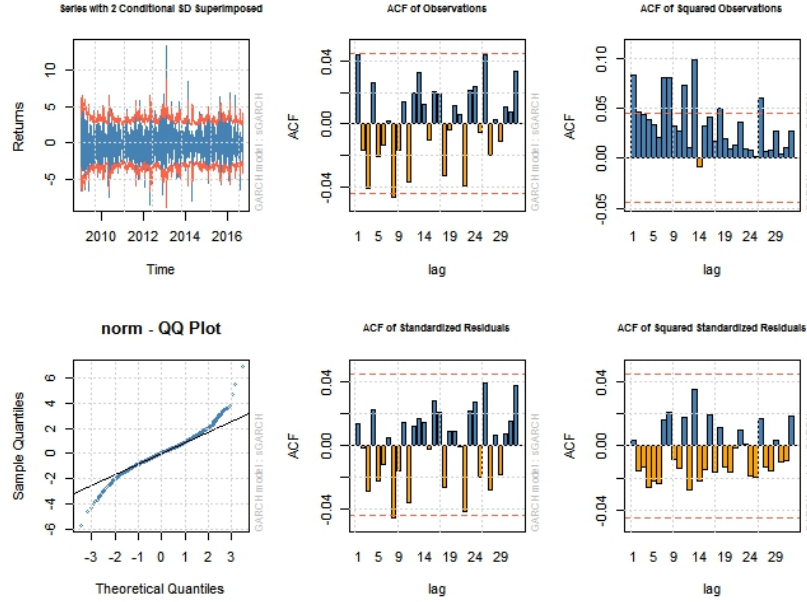


Figure 2: Ajuste de APPL ao modelo AR(1)-GARCH(1,1). Autocorrelação nos quadrados das perdas e distribuição dos resíduos padronizados com caudas longas.

A tabela 1 abaixo traz as informações sobre o modelo ajustado. Os valores estimados dos parâmetros, assim como seus erros padrões, estatística  $t$  e valores  $p$ . Percebe-se que de fato o modelo se ajustou aos dados, considerando que todos os parâmetros estimados são significativos com ao menos 95% de confiança com exceção do parâmetro  $\phi_1$  o qual, pelo gráfico de autocorrelação, já esperávamos fosse pouco significativo.

Ajustado este primeiro modelo, que é o equivalente a filtrar os dados originais de perda com um modelo AR-GARCH de modo que os resíduos padronizados agora sim sejam as realizações de uma variável aleatória *iid*, então podemos aplicar a técnica de modelagem da cauda destes resíduos conforme a teoria de valor extremo. Cabe salientar que é fundamental fazer esta filtragem nos dados,



345 uma vez que a POT supõe dados independentes e identicamente distribuídos, caso contrário não teríamos a convergência dada no teorema 6 de Pickands-Balkema-de Haan.

Table 1: Parâmetros estimados para o modelo AR-GARCH de AAPL

	Estimativa	Erro Padrão	Valor t	$\Pr(> t )$
$\mu$	-0.1683	0.0381	-4.4138	0.0000
$\phi_1$	0.0400	0.0246	1.6282	0.1035
$\omega$	0.2107	0.0514	4.1011	0.0000
$\alpha_1$	0.0876	0.0187	4.6804	0.0000
$\beta_1$	0.8445	0.0301	28.0768	0.0000

Portanto, aplicaremos agora a modelagem POT para os resíduos padronizados, os quais sabemos não serem normalmente distribuídos e que possuem considerável  
350 curtose em excesso (vide gráfico QQ da figura 2) e encontraremos os valores de  $z_\alpha$  e  $E[Z|Z > z_\alpha]$  necessários para computar os valores de  $VarR_\alpha$  e  $ES_\alpha$  das equações (15) e (16).

Os valores de  $z_\alpha$  e  $E[Z|Z > z_\alpha]$  foram calculados para  $\alpha = 0,99$  e são apresentados na tabela 2 abaixo juntamente com os limites inferior e superior de  
355 seu intervalo de confiança a 95%. Note que  $E[Z|Z > z_\alpha]$  na tabela em questão foi chamado de  $s_{.99}$ .

Table 2: Valores de  $z_{.99}$  e  $s_{.99}$  encontrados e seus respectivos intervalos de confiança a 95%

	Inf	Estimativa	Sup
$z_{.99}$	2.40	2.61	2.89
$s_{.99}$	3.11	3.55	4.90

A figura 3 traz algumas informações acerca do ajuste feito a uma GPD dos resíduos padronizados obtidos. Foi escolhido como limiar  $u$  o valor de 1,62 que é o quantil de  $z_t$  a 0,95. Esse valor foi ratificado através da análise qualitativa  
360 do gráfico de vida média residual (MRL plot) conforme encontrado na definição 8. Com a escolha deste limiar, obtivemos 97 observações de perdas em excesso a este valor. Quantidade razoável e que nos permite calcular os parâmetros da distribuição GPD.

Pode-se verificar através da figura 3 que os ajustes feitos foram bastante

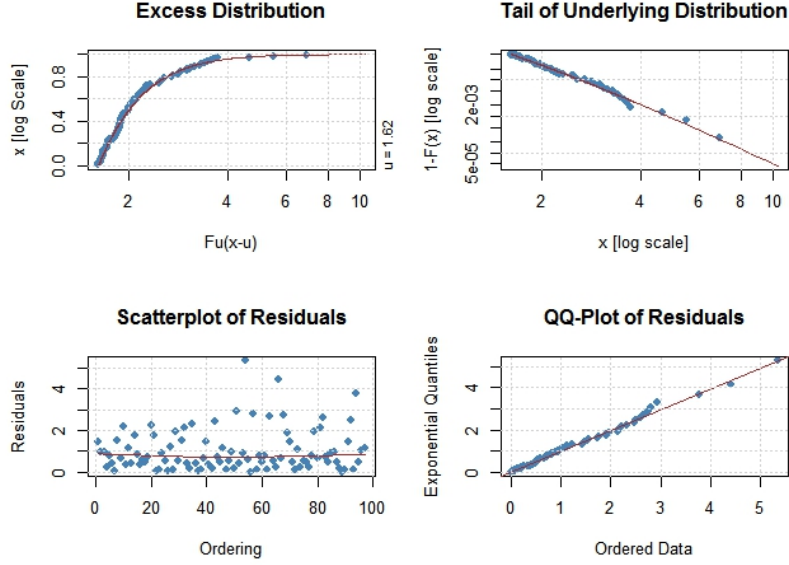


Figure 3: Ajuste dos resíduos padronizados a um modelo POT. Distribuição dos excessos acima do limiar  $u = 1,62$ .

365 satisfatórios, principalmente com a GPD capturando muito bem a curtose em excesso na cauda da distribuição dos resíduos padronizados (*tail of underlying distribution*) e os quantis dos próprios resíduos deste ajuste ficarem juntos dos valores teóricos esperados (*QQ plot of residuals*).

Possuímos todos os parâmetros de nosso modelo híbrido AR-GARCH-POT  
 370 para então calcularmos os valores das medidas de risco condicional dadas pelas equações (15) e (16). Com o modelo AR-GARCH executado no período  $t$  obtemos as estimativas de  $\mu_{t+1}$  e de  $\sigma_{t+1}$  e através do modelo POT obtemos os valores de  $z_\alpha$  e  $E[Z|Z > z_\alpha]$ , no caso do nosso exemplo escolhemos  $\alpha = 0,99$ .

Toda esta metodologia pode ser aplicada período a período, mantendo uma  
 375 memória fixa das últimas  $n$  observações de perdas e que resultará em valores calculados para  $Var_\alpha^{t+1}$  e  $ES_\alpha^{t+1}$ , os quais poderão ser comparados com a realização futura de perda.

#### 4. Processos pontuais

Na seção anterior consideramos apenas a magnitude das perdas em excesso,  
380 sem nos preocuparmos com o momento da ocorrência destas. Implicitamente,  
consideramos que as ocorrências eram independentes entre si, uma condição  
necessária para a modelagem POT.

Entretanto, esta suposição tipicamente é violada pelas séries financeiras, dev-  
ido a autocorrelação entre retornos e principalmente ao *clustering* de volatilidade  
385 nestas séries.

Modelos de processos pontuais, especialmente aqueles conhecidos como pro-  
cessos de Hawkes, ou processos pontuais de auto-excitação, nos permitem modelar  
tanto as magnitudes das perdas em excesso, como a probabilidade de sua ocor-  
rência em período futuro a depender da atual situação. Estes processos tem  
390 natureza dinâmica, assim como aquele desenvolvido na subseção anterior 3.1, e  
nos permitem calcular medidas condicionais de risco com a vantagem de todo o  
processo ser feito em apenas um único passo.

A seguir apresentamos alguns conceitos teóricos sobre processos pontuais de  
Poisson que nos possibilitarão desenvolver o modelo de processos pontuais de  
395 auto-excitação, nos quais realmente estamos interessados. Para uma abordagem  
mais formal e completa sobre processos pontuais, é sugerida a leitura de Daley  
& Vere-Jones (2003, 2008).

Consideremos uma sequência de variáveis aleatórias  $Y_1, \dots, Y_n$  as quais as-  
sumem valores dentro de um determinado espaço de estados  $\mathcal{S}$ . Dado um  
subconjunto  $A \subset \mathcal{S}$ , podemos definir a variável aleatória

$$N(A) = \sum_{i=1}^n I_{\{Y_i \in A\}} \quad (22)$$

que é a contagem do número de observações  $Y_i$  que recaem dentro do conjunto  $A$ .  
A função  $I_{\{Y_i \in A\}}$  é a chamada função indicador, sendo seu valor igual a 1 caso  
400  $Y_i$  esteja contido no subconjunto  $A$  e 0 caso contrário, de forma que a soma dada  
pela equação (22) será justamente a contagem desejada. Neste caso diz-se que

(22) define um processo pontual  $N(\cdot)$ . Um tipo de processo pontual de interesse é aquele de Poisson.

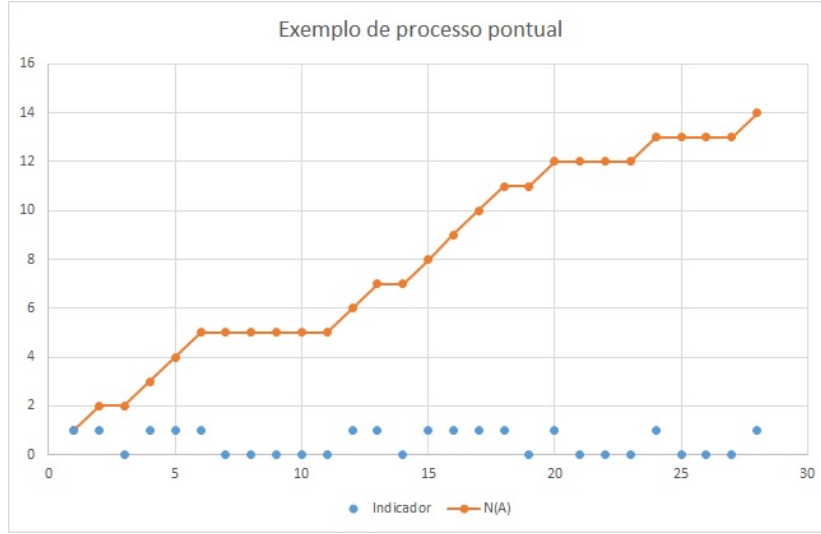


Figure 4: Processo pontual. Cada vez que o indicador alterna para 1 a contagem soma este valor.

**Definição 11** (Processo pontual de Poisson). O processo pontual  $N(\cdot)$  é chamado de Poisson no espaço de estados  $\mathcal{S}$  e com medida de intensidade  $\Lambda$  se as duas condições seguintes são satisfeitas.

(a) Para  $A \subset \mathcal{S}$  e  $k \geq 0$ ,

$$P(N(A) = k) = \begin{cases} e^{-\Lambda(A)} \frac{\Lambda(A)^k}{k!}, & \Lambda(A) < \infty, \\ 0, & \Lambda(A) = \infty. \end{cases}$$

(b) Para qualquer  $m \geq 1$ , se  $A_1, \dots, A_m$  são conjuntos mutuamente exclusivos de  $\mathcal{S}$ , então as variáveis aleatórias  $N(A_1), \dots, N(A_m)$  são independentes.

Também deve ser definida a função intensidade (ou somente intensidade) do processo pontual, que é a derivada da medida de intensidade  $\Lambda(A)$  satisfazendo a equação  $\Lambda(A) = \int_A \lambda(\mathbf{x}) d\mathbf{x}$ . De modo que  $\lambda(\mathbf{x})$  é a função intensidade, que pela sugestão de nossa notação, pode ser uma função multivariável em um vetor  $\mathbf{x}$ .

**Definição 12** (Processo de Poisson homogêneo). Seja  $N$  um processo de Poisson no espaço  $\mathcal{S}$  com intensidade  $\lambda$ , então  $N$  é dito homogêneo se  $\lambda$  não for dependente do espaço  $\mathcal{S}$ , caso contrário é dito não-homogêneo.

Ou seja, em um processo de Poisson homogêneo, sua intensidade depende apenas do tamanho do intervalo onde se está contando o número de ocorrências, mas não da localização deste intervalo.

Um processo de Poisson é dito *marcado*, quando além dos pontos pertencentes ao subconjunto  $A$ , atribuímos a cada um destes pontos uma "marca" com valor dentro do espaço  $\mathcal{M}$ , de forma que a série aleatória passa a ser de pares  $\{(i, L_i) : i \in A \text{ e } L_i \in \mathcal{M}\}$ . Estas marcas podem ser entendidas como a magnitude das perdas em excesso em nosso modelo, enquanto que os pontos do processo serão os períodos em que estas perdas extraordinárias ocorrem.

**Definição 13** (Processos pontuais de Poisson marcados). Suponha um processo pontual de Poisson  $Y(\cdot)$  em  $\mathcal{S}$  com função de intensidade  $\lambda$ . Suponha também que, condicionalmente a  $Y(\cdot)$ , as marcas  $\{m_i : i \in Y\}$  sejam independentes. Então  $N = \{(i, m_i) : i \in Y\}$  é um processo de Poisson marcado. Se as marcas são *iid* com uma distribuição de probabilidades comum  $H$ , então esta é conhecida como a distribuição das marcas.

De acordo com nosso modelo de perdas extremas, podemos entendê-lo como um processo pontual de Poisson marcado ao atribuir aos eventos (i.e. perdas extremas) uma outra informação vinculada, qual seja, o valor desta perda e portanto, os pontos do processo passam a carregar a informação extra da magnitude das perdas. Assim temos um processo que mapeia pares  $(i, l)$  de períodos e perdas a uma probabilidade de uma nova perda extrema.

#### 4.1. Processos pontuais de auto-excitação

São conhecidas as evidências encontradas na literatura que os retornos de séries financeiras não são de fato *iid* e portanto tampouco o são as perdas acima de um limiar. Assim, estas perdas, consideradas no tempo, não ocorrem de acordo com um processo de Poisson homogêneo, tendendo a formarem *clusters*

de perdas (ou retornos) extremos, que correspondem a episódios de concentração de volatilidade. Para contornar este fato, uma dentre várias possibilidades estão os processos de Hawkes, os quais fornecem uma base teórica robusta para este  
445 tipo de processos com autocorrelação.

Passamos nossa atenção aos processos de Hawkes, ou processos pontuais de auto-excitação. Neste tipo de modelagem, a intensidade do processo,  $\lambda(\mathbf{x})$  é dependente dos períodos e magnitudes das marcações anteriores. Traduzindo para nosso modelo de perdas acima de um limiar, a probabilidade de um novo excesso é  
450 dependente do histórico dos períodos e magnitudes dos excessos anteriores. Desta forma temos um modelo dinâmico e condicional para avaliar a probabilidade de eventos extremos, no caso em tela, de perdas excessivas.

Dada a distribuição de perdas realizadas  $L_1, \dots, L_n$  e um valor de limiar  $u$ , assumimos que existam  $N_u$  excessos, cada um ocorrido em um período  $i$ , de  
455 modo que temos a série de pares  $\{(i, L_i) : 1 \leq i \leq n, \quad L_i > u\}$ . Aqui adotamos o indexador de tempo das observações das perdas pertencente aos números naturais, logo  $i \in \mathbb{N}$ . Será também necessária adotar uma notação alternativa apenas para os valores das perdas que foram acima do limiar, e enumerá-las de forma consecutiva. Assim,  $\{(T_j, \tilde{L}_j) : j = 1, \dots, N_u\}$ , com  $T_j \in [1, n]$  e  $\tilde{L}_j$  é o  
460 valor original da perda dado que ela esteja acima do limiar.

Primeiramente consideraremos um modelo de processo pontual para os tempos das perdas em excesso somente. Façamos  $Y_i = iI_{\{L_i > u\}}$ , logo  $Y_i$  retornará o período em que um excesso ocorreu, gerando uma lista de números naturais crescente e não repetidos. O processo pontual dos períodos destes excessos é  
465  $N(\cdot)$  no espaço  $\mathcal{S} = (0, n]$ , e é dado por  $N(A) = \sum_{i=1}^n I_{\{Y_i \in A\}}$  para  $A \subset \mathcal{S}$ . Ou seja, nosso processo é a contagem do número de excessos ocorridos dentro de um intervalo de tempo  $A$  contido em nossa amostra  $\mathcal{S}$ .

Assumimos que este processo pontual seja do tipo Hawkes, com características de auto-excitação, portanto a intensidade em um período  $i$  é condicional aos

470 acontecimentos passados até  $i$ , da seguinte forma

$$\lambda^*(i) = \tau + \psi \sum_{j:0 < T_j < i} h(i - T_j, \tilde{L}_j - u) \quad (23)$$

onde  $\tau > 0$ ,  $\psi \geq 0$  e  $h$  é uma função positiva, decrescente no intervalo de tempo  $s = i - T_j$  e crescente na magnitude do excesso  $l = \tilde{L}_j - u$ . Assim, cada par ordenado de tempo e valor da perda em excesso  $(T_j, \tilde{L}_j)$  anteriores ao período  $t$  contribuem para a intensidade condicional do processo de Poisson. Uma conhecida parametrização para a função  $h$  vem da sismologia através de Ogata (1988). Os processos pontuais de auto-excitação já vêm sendo aplicados a inferência e previsão de tremores sísmicos há décadas, e somente mais recentemente estão sendo aplicados aos "tremores" financeiros e suas ondas de choque. O modelo proposto é chamado de *Epidemic Type Aftershock Sequence* (ETAS) e adotamos a parametrização proposta em Harte (2010), abaixo:

$$h(s, l) = e^{\delta l} \left(1 + \frac{s}{\gamma}\right)^{-\rho}, \text{ onde } \delta, \gamma, \rho > 0.$$

A função intensidade pode ser encarada como a chance instantânea de ocorrência de um ponto do processo no período  $t$ , Chiu et al. (2013), portanto o fato de a transformarmos em uma função condicional resulta que temos agora uma probabilidade condicional de verificarmos no instante  $t$  uma perda em  
 475 excesso e por conseguinte, temos nossas medidas condicionais de risco  $Var_\alpha^t$  e  $ES_\alpha^t$ .

Podemos então tomar nosso modelo POT de perdas acima de um determinado limiar e adequá-lo a um processo de Poisson marcado, com função intensidade auto-excitável. Pode-se separar este modelo em dois tipos, o primeiro com a  
 480 distribuição das marcas do tipo *iid* GPD, de forma que o valor das marcas (i.e. a magnitude das perdas em excesso) é incondicional e distribuído conforme uma distribuição generalizada de Pareto, e um segundo tipo, onde a distribuição das próprias marcas é condicional a informação histórica, mas ainda GPD. Em Chavez-Demoulin et al. (2005) existe uma implementação similar destes

485 modelos e neste artigo alega-se a pouca evidência a favor do segundo modelo,  
onde os parâmetros da GPD seriam também dependentes dos pontos e marcas.  
Desta forma, adotamos aqui o modelo de processo de Hawkes marcado, sendo a  
distribuição das marcas uma distribuição do tipo GPD, dada pela equação (4) e  
com a restrição adicional que  $\xi > 0$  de tal modo que obtemos uma distribuição  
490 com caudas longas para as perdas em excesso.

Vejamos que este modelo não perde a generalização que queremos com relação  
ao fato de os retornos financeiros não serem *iid*. Este fato está bem descrito ao  
adotarmos um processo de Hawkes para os pontos destes, ou seja, a dependência  
dos retornos é dada pela modelagem dos momentos quando estes ocorrem em  
495 excesso. A magnitude destes excessos não interferem na probabilidade de um  
novo evento extremo no futuro imediato.

A primeira vista esta abordagem pode parecer pouco intuitiva, mas ela está  
de acordo com os fatos estilizados das séries financeiras já apresentados, em  
especial, demonstramos que as autocorrelações entre os retornos propriamente  
500 ditos dos ativos é baixa, quando verificável. E este fato está presente em nossa  
abordagem quando afirmamos que a magnitude das perdas em excesso não  
interferem diretamente na probabilidade e na magnitude de uma nova perda  
extrema. As autocorrelações tipicamente encontradas nas séries financeiras são  
entre os retornos ao quadrado, fato que nos remete imediatamente a volatilidade.  
505 Esta sim está modelada através do processo pontual de auto-excitação, quando  
a função intensidade é dependente do histórico do processo até o período atual.

Para encontrarmos a função densidade de probabilidade (p.d.f) conjunta de  
nosso modelo de processo de Poisson marcado (MPP - *Marked Poisson Process*)  
faremos uso da definição de intensidade condicional para um MPP encontrada em,  
510 Daley & Vere-Jones (2003)[p. 249] e também da p.d.f de uma GPD. Começamos  
pela última.

A função densidade de probabilidade de uma distribuição generalizada de



Pareto, parametrizada em  $\xi$  e  $\beta$  é dada por:

$$g_{\xi,\beta}(x) = \frac{1}{\beta} \left(1 + \xi \frac{x}{\beta}\right)^{-1-1/\xi}, \xi \neq 0 \quad (24)$$

e estamos interessados nos valores de  $\xi$  no intervalo  $(0, 1)$ .

**Definição 14** (Intensidade condicional de um processo marcado de Poisson).

Seja  $N$  um processo regular marcado de Poisson em  $\mathbb{N} \times \mathcal{M}$ . A intensidade condicional de  $N$ , em relação ao seu histórico  $\mathcal{H}$ , é a função  $\lambda^*(i, m) = \lambda_g^*(i) f^*(m|i)$ .

Onde  $\lambda_g^*(i)$  é a função intensidade do processo base (i.e. dos tempos dos pontos, no nosso caso a intensidade do processo de Hawkes, dado pela equação (23)) e  $f^*(m|i)$  é a densidade condicional das marcas no período  $i$ . Como estamos trabalhando com uma distribuição GPD independente dos períodos em que ocorrem as perdas extremas,  $f^*(m|i)$  resume-se a nossa p.d.f de uma GPD, dada pela equação (24).

Portanto, nossa p.d.f conjunta para o modelo de processo pontual de auto-excitação das perdas acima de um limiar  $u$ , assumindo uma função núcleo  $h(s, l) = e^{\delta l} \left(1 + \frac{s}{\gamma}\right)^{-\rho}$  pode ser descrita através das seguintes equações:

$$g_{\xi,\beta}(x) = \frac{1}{\beta} \left(1 + \xi \frac{x - u}{\beta}\right)^{-1-1/\xi} \quad (25)$$

$$\lambda_g^*(i) = \tau + \psi \sum_{j: 0 < T_j < i} h(i - T_j, \tilde{L}_j - u) \quad (26)$$

$$h(s, l) = e^{\delta l} \left(1 + \frac{s}{\gamma}\right)^{-\rho} \quad (27)$$

$$s = i - T_j, \quad l = \tilde{L}_j - u \quad (28)$$

$$\lambda^*(i, m) = \lambda_g^*(i) g_{\xi,\beta}(m) \quad (29)$$

Coletando todos os parâmetros a serem estimados no vetor  $\boldsymbol{\theta} = (\tau, \psi, \delta, \gamma, \rho, \xi, \beta)$  e considerando que informamos ao modelo o valor do limiar  $u$ , poderemos estimar todos os parâmetros do modelo através do método de estimação de máxima verossimilhança (MLE - *maximum likelihood estimation*) em um único passo.

A função de verossimilhança logarítmica pode ser deduzida diretamente deste conjunto de equações, sendo a equação (29) a densidade de probabilidades de interesse para a aplicação do método.

**Exemplo 15** (Processo pontual de auto-excitação da Goldman Sachs). Tomaremos como exemplo a aplicação de um processo de Hawkes aos retornos da Goldman Sachs Group Inc. (*ticker* GS) no período entre 05/05/1999 a 04/10/2016, compreendendo 4.384 observações.

Primeiramente é necessário montar a série de pontos do processo, que nada mais são que os pares  $(T_j, \tilde{L}_j - u)$ , onde o valor do limiar  $u$  foi escolhido como sendo o quantil  $q = 0,90$  de nossa amostra. Ou seja, tomamos apenas as 10% maiores perdas de nossa amostra. A partir deste ponto, podemos fazer uma estimativa dos parâmetros do modelo através do método de maximização da verossimilhança, sendo a função de verossimilhança aquela encontrada em (Harte, 2010, eq. (7) p. 6).

Uma vez obtidos os parâmetros do modelo, podemos plotar a função intensidade de base do processo de Hawkes (i. e.  $\lambda_g^*(i)$ ) contra os períodos  $i$  para verificar como esta de fato variou conforme os eventos ocorreram. Na figura 5 é possível verificarmos nossa estimativa da função intensidade de base e compará-la ao eventos do processo de Hawkes, ou seja, as perdas extremas verificadas.

Podemos então calcular a probabilidade condicional de ultrapassarmos um determinado valor de perda  $x$  do nosso ativo, no período  $i$ , dado o conjunto de informações disponíveis. Esta probabilidade nada mais é que a cauda da distribuição das perdas e portanto, o valor  $1 - \alpha$ , onde  $\alpha$  é o intervalo de confiança de nossas medidas de risco. Chamamos de  $\tau^*(i + 1, x)$  esta probabilidade:

$$\begin{aligned} \tau^*(i + 1, x) &= \int_x^\infty \lambda^*(i, y) dy \\ &= \lambda_g^*(i) \bar{G}_{\xi, \beta}(x - u) \end{aligned} \quad (30)$$

O que esta equação nos informa é que a probabilidade de ocorrência de

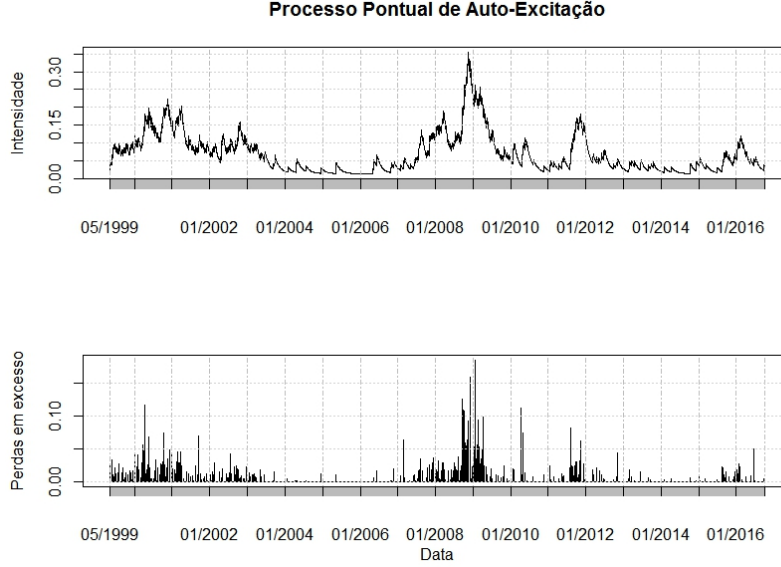


Figure 5: Função intensidade de Hawkes e perdas extremas da Goldman Sachs (GS) no período entre 05/05/1999 a 04/10/2016.

uma perda acima de um valor  $x \geq u$  é dada pelo produto entre nossa função intensidade de base, que não depende de  $x$  e a cauda da distribuição GPD das marcas do processo de Poisson. Ambas já calculadas conforme o exemplo 15. Este resultado não deve vir como uma surpresa, uma vez que a probabilidade de termos uma perda maior que um dado valor  $x \geq u$ , considerando o conjunto de informações até o período  $i$ ,  $\mathcal{G}_i$ , sempre pode ser decomposto da seguinte forma:

$$P(X_{i+1} > x | \mathcal{G}_i) = P(X_{i+1} - u > x - u | X_{i+1} > u, \mathcal{G}_i) P(X_{i+1} > u | \mathcal{G}_i)$$

545 Onde primeiro termo a direita da igualdade é a cauda da distribuição das perdas em excesso, a qual supomos uma GPD independente do período  $i$  e o segundo termo é a probabilidade de ocorrência de um "ponto" no processo pontual, que é justamente a função intensidade de base em um dado período  $i$ .

Assim, a partir de  $\tau^*(i+1, x)$  para encontrarmos digamos,  $Var_{\alpha}^{i+1}$ , o que  
550 precisamos fazer é igualar  $\tau^*(i+1, x)$  a  $1 - \alpha$  e inverter a equação para  $x$ . Este

valor de  $x$  é nossa medida de risco. Para valores de  $\alpha$  suficientemente grandes, de modo que nosso quantil  $VaR_\alpha^{i+1} > u$ , este cálculo pode ser feito e resulta em:

$$VaR_\alpha^{i+1} = u + \frac{\beta}{\xi} \left[ \left( \frac{1-\alpha}{\lambda_g^*(i)} \right)^{-\xi} - 1 \right]. \quad (31)$$

Veja a semelhança desta equação com aquela em (10). Para a abordagem através de processo de Hawkes marcado, substituímos a estimação empírica não-  
555 condicional da probabilidade de acontecer um evento extremo,  $\bar{F}(u) = N_u/N$  por uma estimação que acreditamos mais adequada, o processo de Hawkes na forma de sua função intensidade de base. Da mesma forma, o  $ES_\alpha^{i+1}$  pode ser calculado através da equação (11).

$$ES_\alpha^{i+1} = \frac{VaR_\alpha^{i+1}}{1-\xi} + \frac{\beta - \xi u}{1-\xi} \quad (32)$$

Voltando ao nosso exemplo 15 e aplicando as equações (31) e (32) para a  
560 série de perdas de *Goldman Sachs*, obtemos o seguinte resultado demonstrado na figura 6 abaixo.

Conforme mostra a figura 6, em momentos de baixa volatilidade as medidas condicionais de risco podem cair abaixo de seus valores históricos (i.e. incondicionais) e portanto, esta abordagem pode liberar capital da instituição  
565 financeira de seus requerimentos mínimos. Por outro lado, em momentos de alta volatilidade as medidas condicionais de risco se elevam subitamente e, neste caso, a instituição deveria levantar capital rapidamente.

## References

- Balkema, A. A., & de Haan, L. (1974). Residual Life Time at Great Age. *The*  
570 *Annals of Probability*, 2, 792–804. doi:10.1214/aop/1176996548.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327. doi:10.1016/0304-4076(86)90063-1.
- Chavez-Demoulin, V., Davison, A. C., & McNeil, A. J. (2005). Estimating

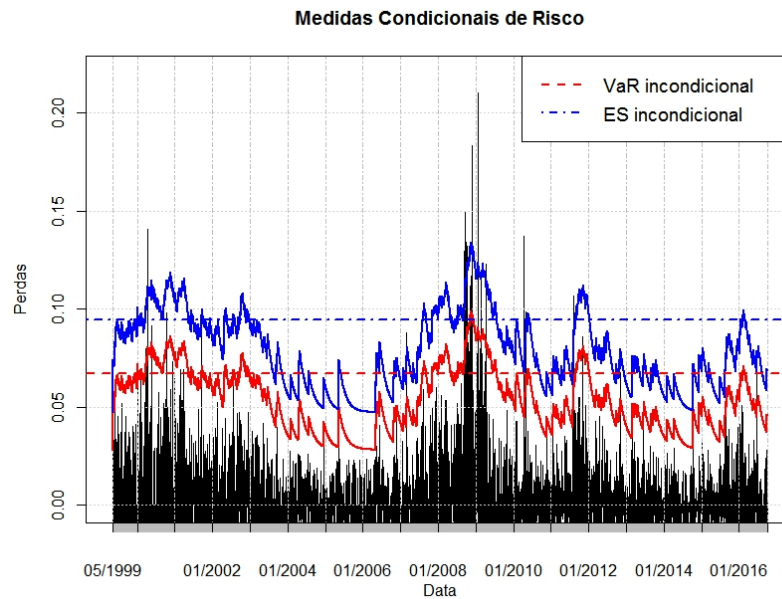


Figure 6: VaR e ES condicionais de *GS*. Linhas contínuas são as medidas condicionais, com ES sempre mais alto que o VaR. As medidas incondicionais são apresentadas para referência.

value-at-risk: a point process approach. *Quantitative Finance*, 5, 227–234.

doi:10.1080/14697680500039613.

Chiu, S. N., Stoyan, D., Kendall, W. S., & Mecke, J. (Eds.) (2013). *Stochastic Geometry and its Applications*. Wiley Series in Probability and Statistics. Chichester, UK: John Wiley and Sons, Ltd. URL: <http://doi.wiley.com/10.1002/9781118658222>. doi:10.1002/9781118658222.

Daley, D. J., & Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Probability and its Applications (2nd ed.). New York: Springer-Verlag. URL: <http://link.springer.com/10.1007/b97277>. doi:10.1007/b97277.

Daley, D. J., & Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure*. Probability and Its Applications (2nd ed.). New York, NY: Springer New

- York. URL: <http://link.springer.com/10.1007/978-0-387-49835-5>.  
doi:10.1007/978-0-387-49835-5.
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling Extremal  
590 Events for Insurance and Finance*. Springer.
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50, 987.  
doi:10.2307/1912773.
- Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency  
595 distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24, 180–190.
- Gnedenko, B. V. (1941). Limit theorems for the maximal term of a variational series. *Comptes Rendus (Doklady) de L'Académie des Sciences de l'URSS*, 32, 7–9.
- 600 Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44, 423–453.
- Gouriéroux, C. (1997). *ARCH Models and Financial Applications*. Springer Series in Statistics. New York, NY: Springer New York.  
URL: <http://link.springer.com/10.1007/978-1-4612-1860-9>. doi:10.  
605 1007/978-1-4612-1860-9.
- Harte, D. (2010). PtProcess : An R Package for Modelling Marked Point Processes Indexed by Time. *Journal of Statistical Software*, 35, 1–32. URL: <http://www.jstatsoft.org/v35/i08/>. doi:10.18637/jss.v035.i08.
- McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for  
610 heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7, 271–300. doi:10.1016/S0927-5398(00)00012-8.
- Ogata, Y. (1988). Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *Journal of the American Statistical Asso-*

*ciation*, 83, 9–27. URL: <http://www.tandfonline.com/doi/abs/10.1080/01621459.1988.10478560>. doi:10.1080/01621459.1988.10478560.

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *Annals of Statistics*, 3, 119–131. doi:10.1214/aos/1176343003.

Tsay, R. S. (2010). *Analysis of financial time series*. (3rd ed.). Wiley.