

Modelagem semiparamétrica de retornos financeiros através da teoria do valor extremo[☆]

Elsevier Inc^{a,b}, Rafael Felipe Bressan^{a,1,*}

^a Avenida Madre Benvenuta, 2007 - Santa Mônica Florianópolis - SC 88035-901

^b 360 Park Avenue South, New York

Abstract

This template helps you to create a properly formatted L^AT_EX manuscript.

Keywords: elsarticle.cls, L^AT_EX, Elsevier, template

2010 MSC: 00-01, 99-00

1. Introdução

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
5 característica, a EVT está sendo utilizada para modelar riscos que possuem 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
10 da cauda da distribuição das perdas de um ativo financeiro maiores que um determinado valor de limiar u . É da modelagem desta cauda, portanto, que faremos as estimativas de risco VaR_α e ES_α .

[☆]Fully documented templates are available in the elsarticle package on CTAN.

*Corresponding author

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

URL: www.elsevier.com (Elsevier Inc)

¹Depto. de Economia/Esag/UDESC

2. O método *peaks over threshold*

Consideremos uma amostra de uma variável aleatória cujas observações sejam
15 independentes e igualmente distribuídas (*iid rv*) $L_i, i \in \mathbb{N}$ que represente as
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
20 é possível demonstrar que a única distribuição para a qual M_n converge com
 $n \rightarrow \infty$ 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-
25 degenerada². Seja $F(x)$ a distribuição original de uma variável aleatória *iid*, é
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
30 dos máximos em bloco de $F(x)$.

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
35 para uma H não-degenerada, então se diz que $F \in MDA(H)$, F pertence ao
domínio de máxima atração de H .

²Distribuição degenerada é aquela cuja densidade de probabilidade está totalmente concen-
trada 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.

40 O teorema 2 foi estabelecido através de três artigos, Fisher & Tippett (1928), 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 ξ é conhecido como o parâmetro de forma da distribuição e
45 dependendo deste valor teremos diferentes tipos de distribuição (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.

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

2.1. 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
55 retornos, ou perdas, dos ativos. Este método é preferido a teoria clássica de 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:

60 **Definição 4** (Distribuição dos excessos). Seja X uma variável aleatória com 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
65 condicional que um valor X retirado dos dados subtraído de u (o excesso) seja
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
70 função de distribuição:

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

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

Os parâmetros ξ e β são conhecidos respectivamente como parâmetros de
forma e escala da distribuição. Na figura 1 abaixo, são mostradas três parametriza-
ções para a função de distribuição acumulada (c.d.f) e para a densidade de
75 probabilidades (p.d.f) de GPD com parâmetro ξ iguais a -0,5, 0 e 0,5 enquanto
que o parâmetro de escala β 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
80 valor extremo em função do teorema de Pickands-Balkema-de Haan (Pickands
(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 $\xi \in \mathbb{R}$.

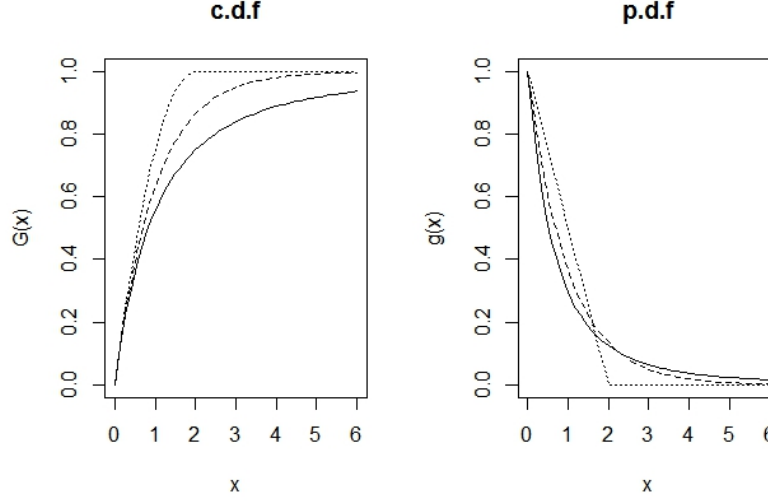


Figure 1: Três parametrizações para uma GPD. A linha sólida corresponde a $\xi = 0,5$, a linha pontilhada a $\xi = -0,5$ e a linha tracejada a $\xi = 0$.

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)),
 85 então a distribuição dos excessos acima de um limiar destas mesmas distribuições convergem para uma GPD, dado um valor de limiar u adequado.

Suposição 7. Seja F a distribuição de perdas com limite direito x_F , assuma 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
 90 u , para $0 \leq x < x_F - u$, $\xi \in \mathbb{R}$ e $\beta > 0$.

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
 95 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
 100 através da integral de sua cauda, $P(X > x) = 1 - P(X \leq x)$. A cauda da GPD
 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 ∞ .

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

Desta forma, o valor esperado de uma GPD $G_{\xi, \beta(u)}(X)$, ou seja, sua média,
 105 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
 110 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,
 115 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
120 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 $F_u(x) \rightarrow G_{\xi,\beta}(x)$
125 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
de u para o qual 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
130 apenas os valores acima deste limiar são utilizados para fazer a estimação dos
parâmetros ξ e β 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
135 $e(u)$ vs u é linear. Por outro lado, um valor limiar muito alto nos impõe o
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 $v \geq u$.

140 Logo, a distribuição dos excessos sobre limiares mais altos que u , também
segue uma GPD com o mesmo parâmetro de forma ξ 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
145 para $u \leq v < \infty$ se $0 \leq \xi < 1$ e, $u \leq v \leq u - \beta/\xi$ se $\xi < 0$.

Esta função é muito útil para calcularmos o ES_α (*expected shortfall*), considerando que Var_α 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 Var_α e ES_α 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 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 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 parâmetros ξ e β dado o limiar u . Por esta modelagem paramétrica podemos conhecer as probabilidades de cauda para valores de x maiores que u .

O quantil α é a inversa da função distribuição e nos retorna o valor para o qual um percentual α de observações da amostra é menor ou igual. Assim sendo,
 170 VaR_α nada mais é que um quantil alto para o qual determinamos que $\alpha\%$ das 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_α e nos basta manipular esta equação até isolarmos VaR_α
 175 como função de $\bar{F}(u)$, α e dos parâmetros da GPD ξ e β . Que nos garante a 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_α pode ser entendida como a média das perdas que excedem o valor dado por VaR_α . Como o próprio VaR_α é um quantil acima do valor de limiar u , ES_α é dado pelo valor do VaR_α somado a função média dos excessos
 180 dada pela equação (8) fazendo $v = VaR_\alpha$. Esta média é convergente para valores 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_α e ES_α , para distribuições de perdas que tiveram suas caudas modeladas através de uma GPD da forma $G_{\xi, \beta(u)}$ com
 185 $\xi < 1$ e $\beta > 0$, podem ser calculadas respectivamente através das equações 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 $\bar{F}(u)$
 190 estamos implicitamente supondo que o número de amostras na série de perdas é significativa, assim como o número de excessos verificados. Daí a importância de se utilizar um valor u adequado, conforme explicitado na seção 2.1.

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

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
195 outras palavras, VaR_α e ES_α 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.

200 De fato, uma das características marcantes das perdas (ou retornos, como 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
205 em um período t é dependente das volatilidades verificadas em períodos anteriores. Um modelo bastante encontrado na literatura que busca modelar estas dependências é o modelo ARCH e suas variantes como GARCH³.

Assim, ao passo que as estimativas de risco desenvolvidas nesta seção são valiosas para prazos mais longos, ainda é necessário desenvolver um modelo
210 que lide com o fato das autocorrelações de volatilidades e portanto, que nossa 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_α e ES_α condicionais que desejamos, ainda dentro da metodologia de *peaks over threshold* - POT.

215 4. Modelo ARMA-GARCH-EVT

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

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

220 anterior, para o momento $t+1$ e posteriores. Desta forma, medidas de risco incondicionais, conforme aquelas derivadas na Seção 3 podem ser adequadas somente para horizontes temporais mais longos, pois implicitamente tomam em consideração os fatos mais recentes com o mesmo valor de predição que fatos mais longínquos.

225 Também já foi bastante estudado e demonstrado no mesmo Capítulo que mod-
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
230 tratam os erros desta modelagem como *iid* e portanto, a metodologia de POT e ajuste de uma GPD pode ser feito.

Revisão

Primeiramente vamos estabelecer um modelo GARCH para as perdas do ativo subjacente. Se denotarmos L_t como sendo a perda observada no período t , μ_t e σ_t são respectivamente a média e o desvio padrão condicionais e mensuráveis
235 através do conjunto de informações disponíveis em $t-1$ e seja Z_t inovações *iid* 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 condi-
240 cionais de risco estamos interessados em um quantil α na cauda de $F_{L_{t+1}|\mathcal{G}_t}(l)$. Este quantil α , que será o nosso VaR_α , é o ínfimo l tal que o valor da distribuição preditiva seja maior ou igual a α . 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:

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

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

245 Considerando que nossa distribuição de perdas é dada pela equação (12) e 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_{α} é o quantil α das inovações Z .

Agora nos falta escolher um processo que modele nossa série temporal dada em
250 (12), ou seja, precisamos especificar o comportamento de μ_t e σ_t . Por suposição 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
255 e o modelo AR(1) para a média condicional. Uma extensão destes modelos 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
260 das medidas condicionais de risco, não faremos maiores assunções acerca da 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.

265 Dadas estas considerações, o modelo adotado segue um formato em dois

Implementar
com modelo
EGARCH

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 μ_{t+1} e σ_{t+1} e portanto, calculamos as inovações implícitas resultantes 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_α e $E[Z|Z > z_\alpha]$, com os quais finalizamos os cálculos de nossas medidas condicionais de risco dadas em (15) e (16).

Nosso modelo completo para as medidas de risco VaR_α e ES_α 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_α e ES_α respectivamente, quando utilizadas em conjunto com aquelas dadas por (10) e (11).

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

Este modelo é ajustado utilizando-se o método de pseudo máxima verossimilhança, 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
290 utilizar tal método para obter nossas estimativas de parâmetros e erros padrão, 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.

295 Seguindo o procedimento em 2 estágios, primeiramente iremos ajustar a 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
300 gráfico as perdas observadas envoltas em uma banda composta por $\pm 2\sigma$. No 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 ϕ_1 em nosso modelo não seja significativo. No terceiro gráfico temos as autocorrelações dos
305 quadrados das perdas, e neste caso sim, de acordo com os fatos estilizados de 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
310 QQ (quantil-quantil) muito utilizado para verificar a distribuição de variáveis 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
315 consideravelmente da normalidade, apresentando caudas muito mais longas que 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.

320

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.

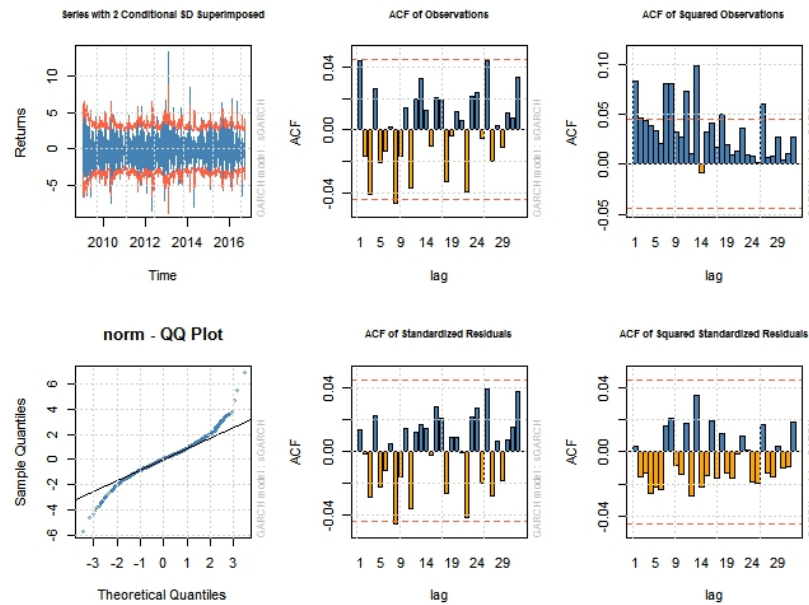


Figure 2: Ajuste de APPL ao modelo AR(1)-GARCH(1,1). Autocorrelação nos quadrados das perdas e distribuição dos resíduos padronizados com caudas longas.

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 ϕ_1 o qual, pelo gráfico de autocorrelação, já esperávamos fosse pouco significativo.

330

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, 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	$\Pr(> t)$
μ	-0.1683	0.0381	-4.4138	0.0000
ϕ_1	0.0400	0.0246	1.6282	0.1035
ω	0.2107	0.0514	4.1011	0.0000
α_1	0.0876	0.0187	4.6804	0.0000
β_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 curtose em excesso (vide gráfico QQ da figura 2) e encontraremos os valores de z_α e $E[Z|Z > z_\alpha]$ necessários para computar os valores de $VarR_\alpha$ e ES_α das equações (15) e (16).

Os valores de z_α 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 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

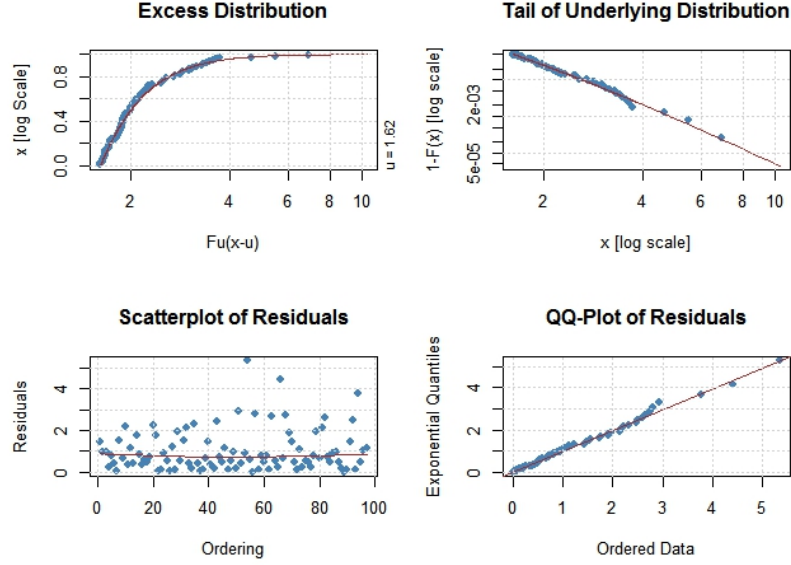


Figure 3: Ajuste dos resíduos padronizados a um modelo POT. Distribuição dos excessos acima do limiar $u = 1,62$.

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.

355 Pode-se verificar através da figura 3 que os ajustes feitos foram bastante 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*).

360 Possuímos todos os parâmetros de nosso modelo híbrido AR-GARCH-POT 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 μ_{t+1} e de σ_{t+1} e através do modelo POT obtemos os valores de z_α e $E[Z|Z > z_\alpha]$, no caso do nosso exemplo escolhemos $\alpha = 0,99$.

365 Toda esta metodologia pode ser aplicada período a período, mantendo uma

memória fixa das últimas n observações de perdas e que resultará em valores calculados para Var_{α}^{t+1} e ES_{α}^{t+1} , os quais poderão ser comparados com a realização futura de perda.

5. Conclusão

370 6. Função de distribuição de valores extremos generalizada

References

- Balkema, A. A., & de Haan, L. (1974). Residual Life Time at Great Age. *The 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.
- 375 Chavez-Demoulin, V., Davison, A. C., & McNeil, A. J. (2005). Estimating 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.
- 380 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.
- 385 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.
- 390

- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling Extremal 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 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.
- 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:10.1007/978-1-4612-1860-9.
- 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 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 Association*, 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.