

R program to implement the DPOT model

P. Araújo Santos
Instituto Politécnico de Santarém and CEAUL
paulo.santos@esg.ipsantarem.pt

M.I. Fraga Alves
Universidade de Lisboa and CEAUL
isabel.alves@fc.ul.pt

September 17, 2012

Abstract: Threshold methods, based on fitting a stochastic model to the excesses over a threshold, were developed under the acronym POT (peaks over threshold). To eliminate the tendency to clustering of violations, a model-based approach within the POT framework that uses the durations between excesses as covariates was proposed by Araújo Santos and Fraga Alves (2012) and denoted by DPOT. In this work, to implement the DPOT model, a program written in the R language is presented and validated with a simulation study.

Keywords: quantitative risk management; statistics of extremes; financial time series.

1 Introduction

Value-at-Risk (VaR) aggregates several components of risk into a single number and has emerged as the standard measure in quantitative risk management (for a detailed discussion of VaR, see Jorion, 2000). In this work we present a program written in the R language to implement a VaR model denoted by DPOT VaR.

The POT method is based on the excesses over a high threshold u and on the Pickands-Balkema-de Haan Theorem (see Balkema and de Haan (1974) and Pickands(1975)). For distributions in the maximum domain of attraction of an extreme value distribution, this theorem states that when u converges to the right end point of the distribution, the excess distribution converges to the Generalized Pareto Distribution (GPD):

$$G_{\gamma,\sigma}(y) = \begin{cases} 1 - (1 + \gamma y/\sigma)^{-1/\gamma}, & \gamma \neq 0 \\ 1 - \exp(-y/\sigma), & \gamma = 0, \end{cases} \quad (1.1)$$

where $\sigma > 0$, and the support is $y \geq 0$ when $\gamma \geq 0$ and $0 \leq y \leq -\sigma/\gamma$ when $\gamma < 0$. Smith (1987) proposed a tail estimator based on a GPD approximation to the excess distribution. Inverting this estimator, an equation is obtained, which allows to calculate the VaR forecast. With financial time series, a relation between the excesses and the durations between excesses is usually observed. Araújo Santos and Fraga Alves (2012) propose using this dependence to improve the risk forecasts with duration-based POT (DPOT) models. For estimation, these models use the

durations, at time of excess i , as the preceding v excesses $(d_{i,v})$. At time t , $d_{t,v}$ denotes the duration until t as the preceding v excesses.

The DPOT model assumes the GPD for the excess above u , such that

$$Y_t \sim GPD\left(\gamma, \sigma_t = \alpha/(d_{t,v})^c\right), \quad (1.2)$$

where γ and α are parameters to be estimated, whereas c plays the role of a tuning parameter. Inverting the tail estimator based on the conditional GPD, the following expression is obtained to calculate the DPOT VaR forecast:

$$\widehat{\text{VaR}}_{t+1|t}^{DPOT(v,c)}(p) = u + \frac{\hat{\alpha}}{\hat{\gamma}(d_{t,v})^c} \left(\left(\frac{n}{n_x p} \right)^{\hat{\gamma}} - 1 \right), \quad (1.3)$$

where n_x denotes the sample size, n the number of excesses, $\hat{\gamma}$ and $\hat{\alpha}$ are estimators of γ and α . Applying the maximum likelihood theory to estimate the parameters, the log likelihood obtained is

$$\begin{aligned} \log L(\gamma, \alpha) &= \log \prod_{i=v}^n f_{Y_i}(y_i) \\ &= - \sum_{i=v}^n \log \left(\frac{\alpha}{(d_{i,v})^c} \right) - \left(\frac{1}{\gamma} + 1 \right) \sum_{i=v}^n \log \left(1 + \frac{\gamma}{\alpha} y_i (d_{i,v})^c \right). \end{aligned} \quad (1.4)$$

Empirical findings in Araújo Santos and Fraga Alves (2012) suggest $v = 3$ and that the method is robust for different values of c in the interval between 0.7 and 0.8. In section 2 we present a R program to implement the VaR model presented in equation (1.3) and in section 3 the program is validated with a simulation study.

2 R program

After running the R code presented below the following message appears: `In log(1 + gamma * y/(alpha1 * (1/x)^c)) : NaNs produced`. The reason for this message is that the optimizer choose values based on a deterministic search algorithm and the `Warning` messages occur when the values do not obey to $0 \leq y_i \leq -\sigma_t/\gamma$ when $\gamma < 0$ (the support when gamma is negative). When this happens, we have logarithm of a negative number and then the message `NaN`, but the optimizer continue to other iterations choosing other values until reach convergence.

```
#### For running this example we suggest to download the daily
#### prices of SP 500 index with at least the first 1001 days,
#### compute the returns and save them in the file
#### with the name SP_500.txt
#### Choose coverage <- 0.01 to forecast VaR(0.01)
```

```

#### Choose c <- 0.75 to implement the DPOT(c=0.75)
#### Choose th <- 0.1 to implement a threshold such that
#### 10% of the values are larger than the threshold

coverage <- 0.01
c <- 0.75
th <- 0.1

#### log-likelihood function which takes three arguments: theta is
#### the vector of parameters, y the excesses and x the durations:

gpdlik <- function(theta,y,x){
  alpha1 <- theta[1]
  gamma <- theta[2]
  n<-length(y)

  logl<- -sum(log(alpha1*(1/x)^c))-(1/gamma+1)*sum(log(1+gamma*y/(alpha1*(1/x)^c)))
  return(-logl)
}

#### We read the log returns from the text file SP_500.txt.
#### Then we compute the symmetric of log returns and choose the
#### returns from day 1 until day 1000, to illustrate the
#### calculation of one-day VaR forecast for day 1001

xx <- read.table("SP_500.txt")

a <- xx*-1 a <- a[,1]
b <- a[1:1000] #from day 1 until day 1000
len <- length(b)

#### Calculation of excesses and durations
#### since the preceding 3 excesses

b_sort <-sort(b)
u <-b_sort[floor(th*len)]
bb <- b[b>u]
bb <- bb-u

duration <- 1
j <- 1

xexc <- rep(0,times=length(bb))

for(ii in 1:len){
  if (b[ii]>u){

    xexc[j] <- duration
    duration <- 1
    j <- j+1
  }

  else {
    duration <-duration+1
  }
}

lag1_xexc <-rep(0,times=length(bb))
d2 <-rep(0,times=length(xexc))
limit <- length(xexc)-1
xxxx <- xexc[1:limit]
lag1_xexc <- c(0, xxxx)
limit2 <- length(xexc)-2
xxxx <- xexc[1:limit2]
lag2_xexc <- c(0, 0, xxxx)

```

```

limit3 <- length(bb)
bb <- bb[3:limit3]
xexc <- xexc[3:limit3]
lag1_xexc <- lag1_xexc[3:limit3]
lag2_xexc <- lag2_xexc[3:limit3]
d3 <- xexc+lag1_xexc+lag2_xexc

#### durations since the preceding 3 excesses (v=3)
#### We use the optim with Nelder and Mead algorithm to
#### maximize the log likelihood

model <- optim(c(0.5,0.5), gpdlik, y=bb, x=d3)
mle1 <- model$par[1]
mle2 <- model$par[2]

#### With the VaR DPOT estimator we compute the forecast

delta <- mle1*(1/(duration+xexc[length(xexc)]+xexc[length(xexc)-1]))^c
var_forecast <- u + ((th/coverage)^mle2-1)*(delta/mle2)

#### One-day-ahead VaR forecast:
var_forecast

```

3 Simulation study for validation of the R program

To validation of the R program we perform a simulation study. First, a geometric process is considered for the durations such that $E[D_i] = 10$. Additionally, a discrete Weibull process is considered for the durations with the shape parameter equal to 0.75, implying some tendency to clustering, and such that $E[D_i] \simeq 10$. For the excesses process we assume the DPOT with the specification proposed in (1.3), with $v = 3$, $c = 0.75$ and the parametrization $\alpha = 5.28$ and $\gamma = 0.208$. This parametrization corresponds to the parameters estimates obtained with the estimation procedure applied to the returns of S&P 500 index from January 4, 1950 to May 18, 2010. See Section 3.3 in n Araújo Santos and Fraga Alves (2012).

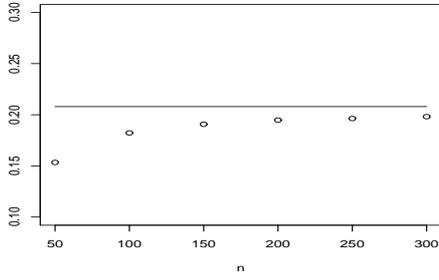
- Geometric process

$$\begin{aligned}
 D_i &\sim \text{Geometric}(q = 0.9) \\
 Y_i &\sim \text{GPD}(\gamma = 0.208, \sigma_i = 5.28/(d_i + d_{i-1} + d_{i-2})^{0.75})
 \end{aligned} \tag{3.1}$$

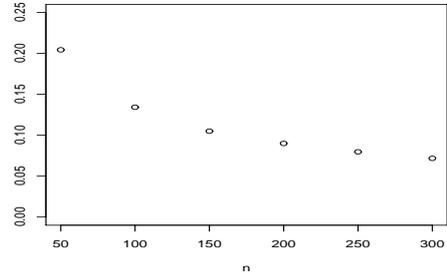
- Discrete Weibull process

$$\begin{aligned}
 D_i &\sim \text{discrete Weibull}(\theta = 0.75, q = 0.81) \\
 Y_i &\sim \text{GPD}(\gamma = 0.208, \sigma_i = 5.28/(d_i + d_{i-1} + d_{i-2})^{0.75})
 \end{aligned} \tag{3.2}$$

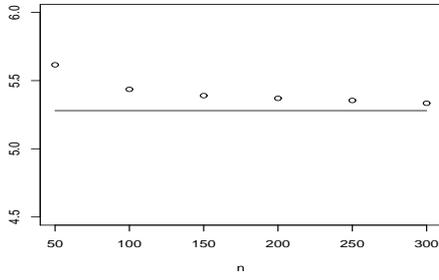
The simulated mean values and root of mean square error (RMSE) for the processes (3.1) and (3.2) are presented in Figure 5, using 5000 simulations in each sample size ($n = 50, 100, 150, 200, 250, 300$). The results suggest that the estimation procedure works well, although some bias is found in the smaller sample sizes. This kind of bias for smaller samples is usual with ML estimators. We only show results for the parametrization $\alpha = 5.28$ and $\gamma = 0.208$, but with further simulations similar results hold for other values of α and γ .



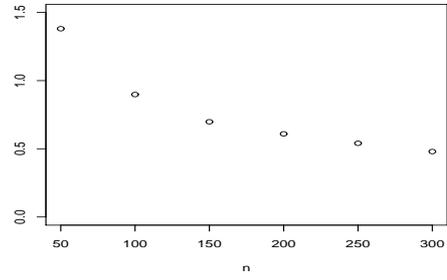
(a) Simulated $E(\hat{\gamma})$. Process (3.1).



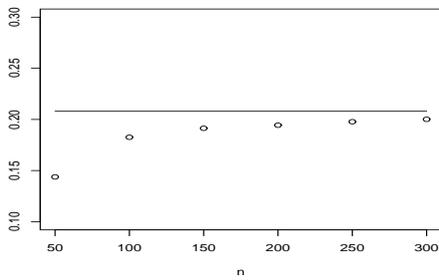
(b) Simulated $RMSE(\hat{\gamma})$. Process (3.1).



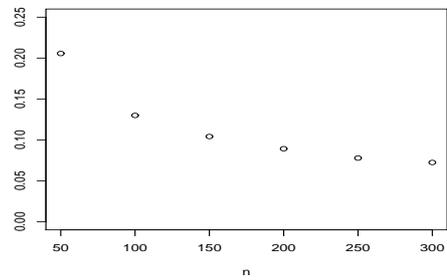
(c) Simulated $E(\hat{\alpha})$. Process (3.1).



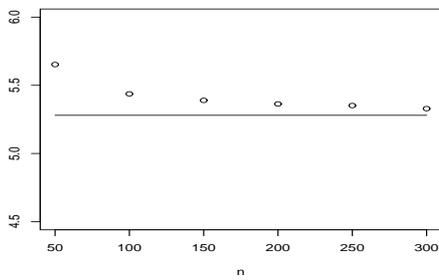
(d) Simulated $RMSE(\hat{\alpha})$. Process (3.1).



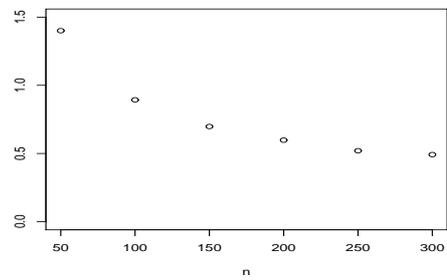
(e) Simulated $E(\hat{\gamma})$. Process (3.2).



(f) Simulated $RMSE(\hat{\gamma})$. Process (3.2).



(g) Simulated $E(\hat{\alpha})$. Process (3.2).



(h) Simulated $RMSE(\hat{\alpha})$. Process (3.2).

Figure 3.1: Simulated mean values and RMSE of $\hat{\gamma}$ and $\hat{\alpha}$ with $n = 50, 100, 150, 200, 250, 300$.

References

- [1] Araújo Santos, P. and Fraga Alves, M.I. (2012). Forecasting Value-at-Risk with a Duration-based POT method, *Mathematics and Computers in Simulation*. In press. <http://dx.doi.org/10.1016/j.matcom.2012.07.016>.
- [2] Balkema, A.A. and de Haan, L. (1974). Residual Life Time at Great Age, *Ann. Probab.*, **2**, 792-804.
- [3] Jorion, P. (2000). *Value at Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill, New York.
- [4] Pickands III, J. (1975). Statistical Inference using Extreme value Order Statistics, *Ann. Statist.*, **3**, 119-131.
- [5] R Development Core Team: R (2012). A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [6] Smith, R. (1987). Estimating tails of probability distributions, *Ann. Statist.*, **15**, 1174-1207.