Efficient estimation for EV regression models of tail risks

Marie KRATZ
ESSEC Business School Paris

Joint work with **Julien HAMBUCKERS** (HEC Liège, Belgium)
and **Antoine USSEGLIO-CARLEVE** (Avignon Univ., France)

Les 30 ans du LMM - Journées 'Risque'

Le Mans, May 23-24, 2024

Extreme Value (EV) regression context

EV regression (Davison and Smith, 90): analysis of factors affecting the likelihood of extreme events

In EVR models: tail distribution of the variable of interest approximated by a generalized Pareto distribution (GPD) (EVT - Balkema-Pickands theorem), which scale and shape parameters = functions of covariates (that influence the distribution of extreme events)

Applications in various fields, as environmental and climate science, insurance, economics, finance.

In finance, EVR models useful to study the link between the state of the financial markets and the likelihood of extreme losses of economic entities.

Application on Hedge Funds (HF) data: estimation of the conditional tail distributions of a large cross-section of HF, to assess the effect of market variables on extreme loss distributions of investment vehicles such as hedge funds (heterogeneous data)

Measuring tail risks of hedge funds (HF)

Hedge funds: rely on sophisticated trading strategies and complex financial products to generate an economic profit.

Identifying financial conditions that influence their extreme downside risks should help:

- to anticipate threats to financial stability
- to evaluate their future performance

Difficult as HF data with (i) short history and (ii) unbalanced nature of available data
$\hookrightarrow$ prevents from applying time series methods or standard EVT, as used for stocks

To overcome this issue, pooling all of the funds' returns and using an EVR model relying on financial factors to control for heterogeneity across funds and time; see Kelly and Jiang (2014; pooling principle, static way), Mhalla et al. (2022); Dupuis et al. (2022)

Extreme Value Regression (EVR)

Let Y_{it}: loss at time $t \in \{1, \dots, T\}$ for a given fund i. Assume:

- Heavy-tailed (Fréchet) loss - Upper tail of the conditional loss distribution above a threshold u_{it} well approximated by a GPD (with $\xi_{it} > 0$)

$$\mathbb{P}\left(Y_{it} > y_{it} \,|\, Y_{it} > u_{it}\right) \underset{u_{it}\text{ large}}{\sim} \left(1 + \frac{\xi_{it}(y_{it} - u_{it})}{\sigma_{it}}\right)^{-1-1/\xi_{it}}$$

- $\xi_{it} = \xi(\mathbf{x}_{it})$ and $\sigma_{it} = \sigma(\mathbf{x}_{it})$ are the conditional shape and scale parameters of the tail distribution, with ξ_{it} d-dimensional vector of covariates observed over time (heterogeneity of the data).

  We connect them with a log-link function, as in classical GLM:

$$\log(\xi_{it}) = \mathbf{x}_{it}^T \boldsymbol{\beta}^{\xi} \qquad \left(\text{and } \log(\sigma_{it}) = \mathbf{x}_{it}^T \boldsymbol{\beta}^{\sigma}\right)$$

  with $\boldsymbol{\beta}^{\xi}$ (and $\boldsymbol{\beta}^{\sigma}$) vectors of regression coefficients, including the constants

$\hookrightarrow$ Interest in $\xi(x_{it})$, our tail risk measure, and $\boldsymbol{\beta}^{\xi}$ that captures the marginal effects of changes in covariates on the tail risk

(See e.g. Davison and Smith (1990); Coles (2001); Chavez-Demoulin et al. (2016).)

The devil is in the (de)tails...

Choosing tail-threshold u, a non-trivial task in EVR models...

We need $u(\mathbf{x})$, e.g., obtained from quantile regression.

$\hookrightarrow$ Several issues:

- Loss of efficiency and weak power of tests for large thresholds
- Selection bias: the larger the α-quantile, the less excesses for given values of the covariates
- Threshold selected in a preliminary step $\Rightarrow$ additional estimation uncertainty in the other parameters, difficult to account for in inferential procedures (He et al., 2022)
- Lack of threshold stability (Eastoe and Tawn, 2009; de Carvalho et al., 2022)

Methodology

1. Extend the EVR model below the tail-threshold via a splicing distributional regression model based on Debbabi et al.'s model/method (efficiency+threshold)
2. Use an artificial censoring procedure to give less importance to non-tail observations (bias-variance tradeoff).
3. Estimate the model with a (conditional) weighted maximum likelihood approach to avoid numerical issues in case of misspecification.

In the literature:

- *Regression context* - Automatic threshold selection procedure for EVR models in a Bayesian context; de Carvalho et al. (2022)
- *Non regression* - Automatic threshold; e.g. Scarrot and MacDonald (2012), Naveau et al. (2016), Bader et al. (2018), Debbabi et al. (2017), Dacorogna et al. (2023)
- *Weighted ML* (not in regression setting): Vandewalle et al. (2007), Diks et al. (2011), Hüser and Davison (2014), Einmahl et al. (2016)

G-E-GPD (splicing) regression model

We build on the automatic threshold method with a hybrid model by Debbabi et al. (2017); automatic tail threshold

Assume the conditional density of $y_{it} \in \mathbb{R}$ w.r.t. a vector of predictors $\mathbf{x}_{it}$,

$$f(y_{it};\boldsymbol{\theta},\mathbf{x}_{it}) = f(y_{it};\, \mu_0, s(\mathbf{x}_{it}), \xi(\mathbf{x}_{it}), \sigma(\mathbf{x}_{it})),$$

satisfies

$$f(y_{it};\boldsymbol{\theta},\mathbf{x}_{it}) = \begin{cases} \gamma_{1,it}\, \varphi(y_{it};\mu_0, s_{it}^2) \text{ if } y_{it} \leq u_{it} \\ \gamma_{2,it}\, e(y_{it};\lambda_{it}) \text{ if } u_{it} \leq y_i \leq u_{it}^* \\ \gamma_{3,it}\, g(y_{it} - u_{it}^*;\, \xi_{it}, \sigma_{it}) \text{ if } y_{it} \geq u_{it}^* \end{cases}$$

with g GPD pdf, φ Gaussian pdf, and e exponential pdf (with parameter λ_{it}), and where f assumed $\mathcal{C}^1$.
$\hookrightarrow$ Model distribution denoted G-E-GPD

G-E-GPD (splicing) regression model

Smoothness ($\mathcal{C}^1$) condition imply the relations

$$\begin{cases} u^*_{it} = \mu_0 + \lambda_{it}\, s^2_{it}\,; \quad \lambda_{it} = (1+\xi_{it})/\sigma_{it}; \;\; \gamma_{1,it} = \gamma_{2,it}\, \dfrac{e(u^*_{it}\,; \lambda_{it})}{\varphi(u_{1,i}\;; \mu_0, s^2_{it})}; \\ \gamma_{2,it} = \left[\xi_{it}\, e^{-\lambda_{it}\, u_{it}} + \left(1 + \lambda_{it}\, \dfrac{\Phi(u^*_{it}\,; \mu_0, s^2_{it})}{\varphi(u^*_{it}\,; \mu_0, s^2_{it})}\right) e^{-\lambda_{it}\, u^*_{it}}\right]^{-1}; \\ \gamma_{3,it} = \sigma_{it}\, \gamma_{2,it}\, e(u_{it}; \lambda_{it}) \end{cases}$$

with $\quad u_{it} = u^*_{it} + \sigma_{it}/\xi_{it}, \quad (\xi_{it} > 0),$ (see e.g. Embrechts et al.'s book)

$$\text{for} \quad \begin{aligned} \log(s_{it}) &= \mathbf{x}^T_{it}\, \boldsymbol{\beta}^s \\ \log(\sigma_{it}) &= \mathbf{x}^T_{it}\, \boldsymbol{\beta}^\sigma \\ \log(\xi_{it}) &= \mathbf{x}^T_{it}\, \boldsymbol{\beta}^\xi \end{aligned}$$

and $\boldsymbol{\theta} = \left[\mu_0, \boldsymbol{\beta}^s, \boldsymbol{\beta}^\sigma, \boldsymbol{\beta}^\xi\right] \in \mathbb{R}^{3d+4}$: the free parameter to estimate.

Estimation methods

- Maximum Likelihood Estimator (MLE)

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{I} \sum_{t=t_{i,1}}^{t_{i,n_i}} \log(f(y_{it};\boldsymbol{\theta},\mathbf{x}_{it})),$$

- **Weighted Maximum Likelihood Estimator (WMLE)**, for lowering the importance of body/left tail (below $q(\tau)$) observations in the likelihood function; based on *artificial censoring procedure* (Cuesta-Albertos et al. (2008); Diks et al. (2011)), to decrease specification issues

$$\hat{\boldsymbol{\theta}}^{w}(\tau) = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{I} \sum_{t=t_{i,1}}^{t_{i,n_i}} \mathbb{1}_{(y_{it} \geq \mathbf{q}(\tau))} \log(f(y_{it};\boldsymbol{\theta},\mathbf{x}_{it})) + \mathbb{1}_{(y_{it} < \mathbf{q}(\tau))} \log(F(q(\tau);\boldsymbol{\theta},\mathbf{x}_{it})),$$

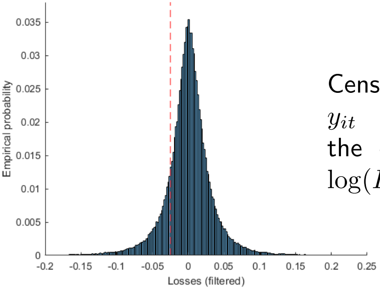
with $F(\cdot;\boldsymbol{\theta},\mathbf{x}_{it})$ the cdf of y_{it}.

Estimation principle known as *minimum scoring rule inference* (Dawid et al., 2016); it guarantees unbiased estimation under limited assumptions

Conditional Weighted ML Estimator (CWMLE)

For the censoring threshold $q(\tau)$, choose an observation-specific threshold, as conditional quantile $q(\tau) = q_{it}(\tau)$ s.t. $\mathbb{P}(y_{it} \leq q_{it}(\tau) | \mathbf{X} = \mathbf{x}_{it}) = \tau$ (estimated via quantile regression on $\mathbf{y}$)

$$\hat{\boldsymbol{\theta}}^{cw}(\tau) = \arg\max_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^{I} \sum_{t=t_{i,1}}^{t_{i,n_i}} \mathbb{1}_{(y_{it} \geq \mathbf{q_{it}}(\tau))} \log(f(y_{it}; \boldsymbol{\theta}, \mathbf{x}_{it})) + \mathbb{1}_{(y_{it} < \mathbf{q_{it}}(\tau))} \log(F(q_{it}(\tau); \boldsymbol{\theta}, \mathbf{x}_{it}))$$



Censoring threshold $q_{it}(\tau)$ s.t., for $y_{it} < q_{it}(\tau)$, the contribution to the censored likelihood function is $\log(F(q_{it}(\tau); \boldsymbol{\theta}, \mathbf{x}_{it}))$

$\hookrightarrow$ For $q(\tau)$ known (in practice, empirical quantile of $(y_{it})_{i,t}$), $\hat{\boldsymbol{\theta}}^{cw}(\tau) =$ **M-estimator**, as defined in Van der Vaart (2000) .

Asymptotic normality

Data-driven selection of the censoring threshold τ

using the modified Anderson-Darling (AD^m) statistic to assess the goodness-of-fit of a parametric distribution, with a *special weight to extreme values in the upper tail* of the distribution (Babu and Toreti, 2016): Select

$$\tau^{opt} = \arg \min_{\tau \in [0.05, .5]} \mathsf{AD}^m(\tau),$$

where:

$$AD^m(\tau) = n \int_{-\infty}^{+\infty} \frac{\left(\Phi(x) - \tilde{F}_n(x, \tau)\right)^2}{1 - \Phi(x)} d\Phi(x) \qquad \text{with}$$

- Standard normal cdf Φ; Φ^{-1}: its quantile function
- $\tilde{F}_n(\cdot, \tau)$: empirical cdf of the re-indexed pseudo-residuals $(\hat{\varepsilon}_k(\tau))_{k=1,\cdots,n}$, $k = 1, \ldots, n$, $n = \sum_{i=1}^{I} n_i$;
- pseudo-residuals computed from the PIT:

$$\hat{\varepsilon}_{it}(\tau) = \Phi^{-1}(F(y_{it}; \hat{\boldsymbol{\theta}}^{w \text{ or } cw}(\tau), \mathbf{x}_{it})),$$

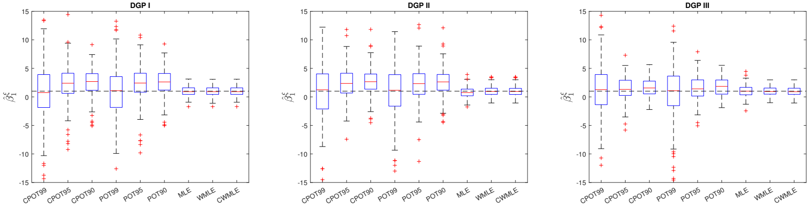
Under *correct specification and good estimation* of the splicing model (in part. in the tail), residuals $\overset{d}{\sim}$ Gaussian (Dunn and Smyth, 1996)

Simulation study - estimates and RMSE

Objective: to quantify the gains in terms of bias and variance of our splicing approach over standard POT approaches. Consider 3 settings DGPsimulation

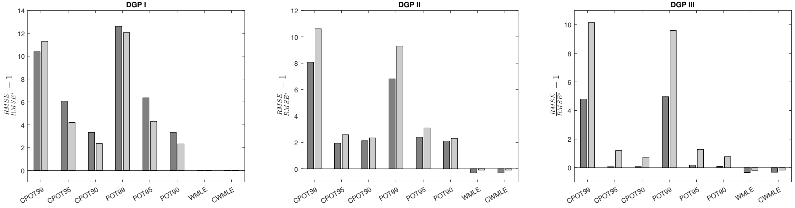
(DGPI) correct specification; (DGPII) contaminated body; (DGPIII) misspecified model

- *Estimation of β^{ξ}:*



Estimated parameters $\hat{\beta}_1^{\xi}$ obtained with the various methods. Dashed: values of the true parameters.

- *RMSE*



$RMSE/RMSE^* - 1$ for $\hat{\beta}_0^{\xi}$ (dark grey) and $\hat{\beta}_1^{\xi}$ (light grey), for DGP I to III, with $RMSE^*$ for the MLE.

Application: Hedge funds (HF) data

Well acknowledged that HF exhibit tail risks but no clear characterization of the link between economic risk factors and the conditional tail distribution of HF

Recall: Tail risk measured with conditional tail index of the cross-sectional distribution of the returns ($\xi(x_t)$)

- EurakHedge database: US HF monthly returns 01/1995 - 12/2021 (Long/short equity funds reporting ≥ 60 months (uninterrupted history)
  Sample: 189,000 pooled observations spanning 1,484 funds, average reporting of 130 months)
- Filtering with *Asset Pricing (AP) model* of Patton & Ramadorai (2013) on a fund-by-fund basis (to remove time variations in the mean of the returns (r_{it})).
  $\hookrightarrow$ Consider the negative residuals $\hat{y}_{it}$ associated to the observed returns r_{it} (at time t for fund i) (i.e. the residuals of the AP model multiplied by -1)
- Tail risk analysis on $\hat{y}_{it}$, pooled (negative) residuals, assumed, at a given point in time, to have their *tail distribution driven by the same statistical model* (Mhalla et al.,2022; Dupuis et al.,2022; Appendix)

HF data

$\hookrightarrow$ From $\{\hat{y}_{1t}, \ldots, \hat{y}_{1T}, \ldots\}$, estimate $\xi(\boldsymbol{x}_t)$ and $\boldsymbol{\beta}^{\xi}$, using EVR model with our splicing regression approach
Note that $\xi(\boldsymbol{x}_{i,t}) = \xi(\boldsymbol{x}_t)\ \forall i = 1, \cdots, I$. The predictor matrix of covariates has been standardized (to have mean and variance of each column equal to 0 and 1, respectively).

- For the censoring threshold: $\tau^{opt} = \arg \min_{\tau \in [0.05, .5]} \text{AD}^m(\tau) \sim 0.25$, i.e. misspecification of the G-E-GPD model in the left tail
- Tail-thresholds u_{it} obtained with WMLE and CWMLE
  $\hookrightarrow$ For HF data: For most combinations of the covariates, suitable thresholds $=$ conditional quantiles at levels $\in (0.985, 0.995)$ (average ~ 0.99). For some combinations, lower threshold levels as small as 0.96.

$\hookrightarrow$ Overall, around 1% of observations exceed these thresholds

HF data: Regression effects (e.g. for β^ξ)

Covariate	WMLE	CWMLE	MLE	POT95	POT97.5
$\beta^\xi(\texttt{TED})$	-0.40***	-0.40***	-0.28***	-0.18	-0.12
	$[-0.52, -0.28]$	$[-0.52, -0.28]$	$[-0.42, -0.15]$	$[-0.38, 0.02]$	$[-0.41, 0.17]$
$\beta^\xi(\texttt{VIX})$	-0.51***	-0.50***	-0.04	-0.07	-0.12
	$[-0.65, -0.36]$	$[-0.66, -0.35]$	$[-0.15, 0.07]$	$[-0.25, 0.11]$	$[-0.40, 0.16]$
$\beta^\xi(\Delta\texttt{TED})$	0.12	0.12	0.31***	0.16**	0.05
	$[-0.09, 0.33]$	$[-0.08, 0.31]$	$[0.19, 0.42]$	$[0.01, 0.32]$	$[-0.19, 0.29]$
$\beta^\xi(\Delta\texttt{MSCI})$	0.24***	0.23***	0.09**	0.05	-0.04
	$[0.17, 0.30]$	$[0.16, 0.30]$	$[0.01, 0.17]$	$[-0.09, 0.19]$	$[-0.24, 0.15]$
$\beta^\xi(\texttt{MOM})$	0.15***	0.16***	-0.01	-0.12	-0.16
	$[0.09, 0.20]$	$[0.11, 0.21]$	$[-0.08, 0.05]$	$[-0.26, 0.02]$	$[-0.36, 0.04]$
$\beta^\xi(\texttt{Liq})$	-0.16***	-0.15***	-0.01	-0.08	0.04
	$[-0.24, -0.09]$	$[-0.23, -0.08]$	$[-0.09, 0.06]$	$[-0.26, 0.09]$	$[-0.23, 0.31]$
$\beta^\xi(\texttt{CredSpr})$	-0.58***	-0.60***	-0.29***	0.15	0.23
	$[-0.66, -0.51]$	$[-0.67, -0.52]$	$[-0.39, -0.19]$	$[-0.03, 0.33]$	$[-0.02, 0.47]$

Estimated regression effects for the different estimation methods. Confidence intervals at the 95% level are in brackets below the estimates. CWMLE has been obtained using all variables as conditioning variables.*** and ** indicate coefficients significant at the 1% and 5% test levels, respectively.

Conclusion

- On the method(s)
  - ↪ POT is asymptotically ok but inefficient.
  - ↪ We introduced a simple approach (via G-E-GPD) to estimate EVR models more efficiently without having to select an arbitrary threshold
  - ↪ To challenge the method, various statistical tests performed on simulated data
- On the application to HF: Tail risk of hedge funds, over time, are well explained by `TED`, `VIX`, `MSCI` and `CredSpr`
- How to deal with large numbers of covariates/model selection?

For more details and references: see the arXiv preprint (2023-24)

Thank you for your attention!

Debabbi et al.'s method

A self-calibrating method for modelling heavy tailed probability distributions (*N. Debbabi, M. Kratz, M. Mboup* (2017))

Frame: (right) **heavy-tailed continuous** data $\rightarrow$ fit the tail using a **GPD** with a positive tail index (Fréchet domain of attraction)

Main motivation: to suggest a **unsupervised method to determine the threshold** above which we fit the GPD, and to have a good fit for the **entire distribution**

Way: introduce a **hybrid model** with 3 components (G-E-GPD):

- a Gaussian distribution to model the mean behavior (CLT)
- a **GPD** for the **tail** (Pickands theorem)
- an exponential distribution to bridge the mean and tail behaviors

Assumption: the distribution (which belongs to the Fréchet domain of attraction) has a density that is C^1.

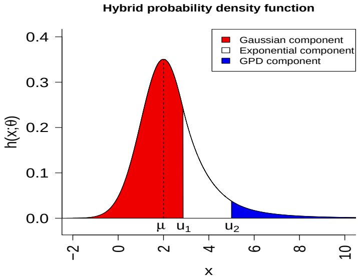
Remark. The main component in this hybrid model is the GPD one (for heavy tail), the mean behavior having to be adapted to the context. For instance, for *positive asymmetric distribution*, we have replaced the Gaussian component with a *Lognormal* one (LOGN-E-GPD hybrid model)(see Dacorogna et al.2023).

$$h(x;\theta) = \begin{cases} \gamma_1 \ f(x;\mu,s), & \text{if} \quad x \leq u_1, \\ \gamma_2 \ e(x;\lambda), & \text{if} \quad u_1 \leq x \leq u_2, \\ \gamma_3 \ g(x-u_2;\xi,\sigma), & \text{if} \quad x \geq u_2, \end{cases}$$

f: Gaussian pdf (μ, s^2).
e: Exponential pdf with intensity λ.
g: GPD pdf with tail index ξ and scale parameter β.

$\sigma = [\mu, \sigma, u_2, \xi]$: the parameters vector.
γ_1, γ_2 and γ_3: the weights (evaluated from the assumptions (in part. C^1))
$\beta = u_2\xi > 0$; $\lambda = \frac{1+\xi}{\sigma}$; $u_1 = \mu + \lambda s^2$

Back to splicing Regression Model



Confidence interval

From the theory of M-estimators (Van der Vaart, 2000), we have:

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{W}(\tau) - \boldsymbol{\theta}_0\right) \to \mathcal{N}(0, V(\boldsymbol{\theta}_0))$$

with:

$$\begin{aligned}
\mathcal{I}(\boldsymbol{\theta}_0) &= \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\Psi(y,\boldsymbol{\theta}_0)\right] \\
\mathcal{S}(\boldsymbol{\theta}_0) &= \mathbb{E}\left[\left\{\frac{\partial}{\partial\boldsymbol{\theta}_0}\Psi(y,\boldsymbol{\theta}_0)\right\}\left\{\frac{\partial}{\partial\boldsymbol{\theta}_0}\Psi(y,\boldsymbol{\theta}_0)\right\}^{\top}\right] \\
V(\boldsymbol{\theta}_0) &= \mathcal{I}(\boldsymbol{\theta})^{-1}\mathcal{S}(\boldsymbol{\theta}_0)\left\{\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\right\}^{\top},
\end{aligned}$$

$$\begin{aligned}
\Psi(y,\boldsymbol{\theta}_0) = \sum_{i=1}^{n} & \mathbb{1}(y_i \geq q_i(\tau))\log(f(y_i;\boldsymbol{\theta}_0,\mathbf{x}_i)) \\
& + \mathbb{1}(y_i < q_i(\tau))\log(F(q_i(\tau);\boldsymbol{\theta}_0,\mathbf{x}_i)).
\end{aligned}$$

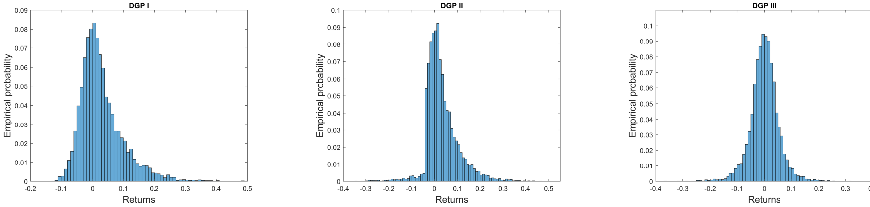
Back to estimation

Simulated data - DGP

DGP I Assume the response variable y_{it} simulated from the splicing regression model, making the model perfectly specified. Data exhibit a heavy right tail and positive skewness. True conditional threshold (above which $y \sim$ EVR model) is not linear in the covariate (quadratic polynomial)

DGP II Misspecification in the left tail, now heavier compared to DGP I (but models identical in the right tail)

DGP III Both tails heavier than the normal as well, and the right tail is not exactly GP-distributed (e.g. data simulated from a t-location scale)



Simulated data from the different DGPs.