Measuring tail risk at high-frequency: An L_1-regularized extreme value regression approach with unit-root predictors

Li Sun

Coauthors: Luca Trapin,
Julien Hambuckers

ESEM at Barcelona, 28th August 2023

Contents

Table of Contents

Introduction: research background

- **High-frequency (HF) financial data + its tail risk**:

- **Extreme value regression (EVR)**:

- **Nonstationary predictors**:

Introduction: research background

- **High-frequency (HF) financial data + its tail risk**:
  frequency and severity of extreme events, data abundance, stylized facts

- **Extreme value regression (EVR)**:
  generalized Pareto distribution (GPD) models, see
  Chavez-Demoulin and Embrechts (2004); Chavez-Demoulin et al. (2016); Massacci (2017), Bee et al. (2019), Schwaab et al. (2021), etc.

- **Nonstationary predictors**: direct evidence + indirect evidence

Introduction: research objective

- to model the extreme loss severity distribution of a financial asset in HF markets
- to find a clear association of market indicators with the potential loss severity
- to predict financial risk measures (e.g. VaR) for financial risk management

Introduction: research objective

- to model the extreme loss severity distribution of a financial asset in HF markets
  $\rightarrow$ an appropriate model specification

- to find a clear association of market indicators with the potential loss severity
  $\rightarrow$ L_1-regularized model regression: automatic variable selection

- to predict financial risk measures (e.g. VaR) for financial risk management
  $\rightarrow$ construction of predictors & predictive performance

Table of Contents

Model specification

- **random variables**:

p_t	$r_t := \log(p_t) - \log(p_{t-1})$	$l_t := -r_t$	y_t	$\boldsymbol{z}_{t-1}$
stock price	return	loss	excess loss	candidate predictors

- **dynamic peaks-over-threshold (POT)**:

assume	$F_t \in \mathcal{D}(GP(k_t, \sigma_t))$,	where F_t is the cdf of l_t conditional on $\mathcal{F}_{t-1}$;
obtain	$y_t \overset{\mathcal{D}}{\sim} GP(k_t, \sigma_t)$,	where $\mathrm{GP}(y_t; k_t, \sigma_t) = 1 - \left(1 + k_t \frac{y_t}{\sigma_t}\right)^{-1/k_t}$.

- **We propose a model specification on $\{k_t\}, \{\sigma_t\}$:**

$$\begin{cases} \log\left(\dfrac{k_t}{0.5 - k_t}\right) = \beta_{1,0} + \displaystyle\sum_{j=1}^{p} \beta_{1,j}\, z_{j,t-1}\,, \\ \log(\sigma_t) = \beta_{2,0} + \displaystyle\sum_{j=1}^{p} \beta_{2,j}\, z_{j,t-1} + \beta_{2,p+1} \log(\sigma_{t-1})\,. \end{cases}$$

Model specification: pros

Advantages of $\begin{cases} \log\left(\frac{k_t}{0.5 - k_t}\right) = \beta_{1,0} + \sum_{j=1}^{p} \beta_{1,j}\, z_{j,t-1}\,, \\ \log(\sigma_t) = \beta_{2,0} + \sum_{j=1}^{p} \beta_{2,j}\, z_{j,t-1} + \beta_{2,p+1}\, \log(\sigma_{t-1})\,. \end{cases}$

- $0 \leq k_t \leq 0.5, \sigma_t > 0$
- k_t, σ_t time-varying over information set $\mathcal{F}_{t-1}$
- allow autoregressive structure in $\{\log(\sigma_t)\}$
- allow local-to-zero unit-root covariates:

  $\exists j, \{z_{j,t}\} \in I(1)$ but with $\beta_{1,j}, \beta_{2,j} = O(\frac{1}{\sqrt{T}})$

Table of Contents

Maximum likelihood estimation (MLE)

ML estimator:

$$\widehat{\beta}^{mle} = \arg\max_{\beta\in\Theta} \ \mathcal{L}(\beta; \{y_t\}, \{\boldsymbol{z}^*_{t-1}\})$$
$$= \arg\max_{\beta\in\Theta} \ \sum_{t=1}^{T} \mathbb{1}\{y_t > 0\} \left(-\log(\sigma_t) - \left(\frac{1}{k_t} + 1\right) \log(1 + k_t \frac{y_t}{\sigma_t}) \right)$$

where $\boldsymbol{z}^*_t := [\ z_{1,t}, \ldots, z_{p_0,t}, \frac{z_{p_0+1,t}}{\sqrt{T}}, \ldots, \frac{z_{p,t}}{\sqrt{T}}\ ]'$.

Properties of MLE — Theorems 1&2:

- Consistency: $\left\|\widehat{\beta}^{mle} - \beta^{o*}\right\| = o_p(1)$
- Asymptotic non-normality: $\sqrt{T}(\widehat{\beta}^{mle} - \beta^{o*}) = O_p(1)$

Adaptive L_1-regularized maximum likelihood estimation (ALMLE)

General form of ALMLE:

$$\widehat{\boldsymbol{\beta}}^{\text{al}} = \arg\min_{\boldsymbol{\beta}\in\Theta} \; -\mathcal{L}(\boldsymbol{\beta}; \{y_t\}, \{\boldsymbol{z}_t^*\}) + \lambda_{k,T} \sum_{i=1}^{p} w_{k,i}|\beta_{1,i}| + \lambda_{\sigma,T} \sum_{j=1}^{p+1} w_{\sigma,j}|\beta_{2,j}|,$$

Tuning parameter selection $(\widehat{\lambda}_{k,T}, \widehat{\lambda}_{\sigma,T})$ using BIC

$$\arg\min_{\lambda_{k,T}\in S_{\lambda_{k,T}},\, \lambda_{\sigma,T}\in S_{\lambda_{\sigma,T}}} \; -2\log(\mathcal{L}(\widehat{\boldsymbol{\beta}}^{\text{al}}(\lambda_{k,T}, \lambda_{\sigma,T}); \{y_t\}, \{\boldsymbol{z}_t^*\})) + \log(T)\left(\sum_{i=1,\dots,p} \mathbb{1}\{\widehat{\beta}^{\text{al}}_{1,i} \neq 0\} + \sum_{j=1,\dots,(p+1)} \mathbb{1}\{\widehat{\beta}^{\text{al}}_{2,j} \neq 0\}\right),$$

where $S_{\lambda_{k,T}}$ and $S_{\lambda_{\sigma,T}}$ are logarithmic tuning parameter grids of $\lambda_{k,T}$ and $\lambda_{k,T}$ respectively.

ALMLE: issue with GPD when $I(1)s$ truly active

- **Issue**: findings — variable underselection
- **Root reason**: Theorem 3 — the role of local optimizers at $k_t = 0.5$
- **Goal**: fairly preserve the sensitivity to variable selection along tuning parameter grids
- **Solution**: propose a 2-step ALMLE

2-step ALMLE (ALMLE2S)

Step 1 with $\widehat{\beta}^{\mathbf{k,al}}$ — determine the active predictor set $\mathcal{A}_T^{k,al}$ for $k_t(\cdot)$

Step 2 with $\widehat{\beta}^{\mathbf{tal}}$ — determine the active set $\mathcal{A}_T^{\sigma,al}$ and all model coefficients

2-step ALMLE (ALMLE2S)

Step 1 with $\widehat{\boldsymbol{\beta}}^{\text{k,al}}$ — determine the active predictor set $\mathcal{A}_T^{k,al}$ for $k_t(\cdot)$

$$\widehat{\boldsymbol{\beta}}^{\text{k,al}} = \operatorname*{arg\,min}_{\{\boldsymbol{\beta}\in\Theta|\beta_{2,j}=0,i=1,\dots,p+1\}} -\mathcal{L}(\boldsymbol{\beta};\{y_t\},\{\boldsymbol{z}_t^*\}) + \lambda_{k,T}\sum_{i=1}^{p}\widetilde{w}_{k,i}|\beta_{1,i}|,$$

Step 2 with $\widehat{\boldsymbol{\beta}}^{\text{tal}}$ — determine the active set $\mathcal{A}_T^{\sigma,al}$ and all model coefficients

$$\widehat{\boldsymbol{\beta}}^{tal} = \operatorname*{arg\,min}_{\{\boldsymbol{\beta}\in\Theta|\beta_{1,i}=0,(1,i)\notin\mathcal{A}_T^{k,al}\}} -\mathcal{L}(\boldsymbol{\beta};\{y_t\},\{\boldsymbol{z}_t^*\}) + \widehat{\lambda}_{k,T}\sum_{i=1}^{p}\widetilde{w}_{k,i}|\beta_{1,i}| + \lambda_{\sigma,T}\sum_{j=1}^{p+1}\widetilde{w}_{\sigma,j}|\beta_{2,j}|,$$

where $\mathcal{A}_T^{k,al} := \left\{(1,j) : j \geq 1, \widehat{\beta}_{1,j}^{k,al} \neq 0\right\}$, and

$$\begin{cases} \widetilde{w}_{k,i} = \dfrac{1}{\widehat{\beta}_{1,i}^{\text{k,mle}}}\,\dfrac{1}{\widehat{\beta}_{1,i}^{\text{mle}}}, & i = 1,\dots,p, \\ \widetilde{w}_{\sigma,j} = \dfrac{1}{\widehat{\beta}_{2,j}^{\text{mle}}}, & j = 1,\dots,p+1, \end{cases}$$

with $\widehat{\boldsymbol{\beta}}^{\text{k,mle}} := \operatorname*{arg\,min}_{\{\boldsymbol{\beta}\in\Theta|\beta_{2,j}=0,j=1,\dots,p+1\}} -\mathcal{L}(\boldsymbol{\beta};\{y_t\},\{\boldsymbol{z}_t^*\})\,.$

ALMLE2S: the oracle property

Oracle Property of $\widehat{\beta}^{tal}$ — Theorem 4:

- Model selection consistency:

$$\lim_{T\to\infty} P\left\{\mathcal{A}_T^{tal} = \mathcal{A}\right\} = 1\,,$$

where $\mathcal{A}_T^{tal} := \mathcal{A}_{k,T}^{tal} \cup \mathcal{A}_{\sigma,T}^{tal}$.

- Limiting distribution of $\widehat{\beta}^{tal}$:

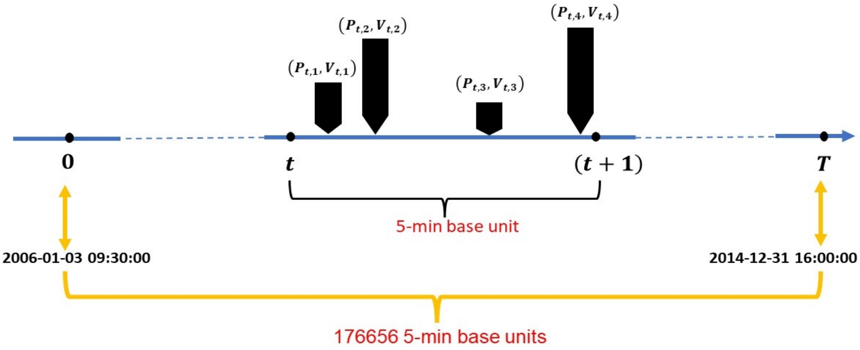
$$\begin{cases} \sqrt{T}\left(\widehat{\beta}_{\mathcal{A}}^{tal} - \beta_{\mathcal{A}}^{o*}\right) = O_p(1) \\ \sqrt{T}\left(\widehat{\beta}_{\mathcal{A}^c}^{tal} - \beta_{\mathcal{A}^c}^{o*}\right) = o_p(1) \end{cases}$$

Table of Contents

Empirical application: study the high-frequency excess loss distributions of nine large liquid U.S. stocks

Stocks	American Express	Boeing	General Electric	Home Depot	IBM	Johnson and Johnson	JPMorgan Chase	Coca-Cola	ExxonMobil
Symbols	AXP	BA	GE	HD	IBM	JNJ	JPM	KO	XOM

- **Raw data:** transaction data (in milliseconds)



Empirical application: model set-up

List of variables in use

in preparation: $P_t = \text{median}\{P_{t,1}, \ldots, P_{t,n_t}\}$, $l_t = -(\log(P_t) - \log(P_{t-1}))$

dependent variable: $y_t = (l_t - u_t)\,\mathbb{1}\{l_t - u_t > 0\}$, with $u_t = \text{Quantile}_{90\%}(\{P_{t-1}, \ldots, P_{t-h}\})$

42 candidate predictors using info in 3 different frequencies: within the 5-min base unit(**W**), across (**A**) 5-min variables, or as a ratio (**R**) between the two frequencies	**liquidity measures**	price impact (**PI**): e.g. Amihud Illiquidity Measure spread (**S**): e.g. Roll Measure volatility of liquidity (**Vol**): e.g. Transaction Volume Volatility
	volatility measures	i.e., Micro Noise Return Volatility, Micro Realized Volatility, Realized Volatility, MNRV2RV, MRV2RV

Empirical application: model selection result

k_t model selection

feature summary	clean selection – mainly on W and R frequencies, PI liquidity proxies, relative volatility measure
selected predictors (selection frequency > 50%, estimated coefficient sign)	$\text{TV}_t = \sum_{i=1}^{n_t} P_{t,i} V_{t,i}$, $(89\%, +)$ $\text{TQ}_t = \sum_{i=1}^{n_t} V_{t,i}$, $(89\%, +)$ $\text{AM}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} \frac{\lvert R_{t,i} \rvert}{P_{t,i} V_{t,i}}$, $(59\%, -)$ $\text{EAM}_t = \frac{\max(P_{t,BU}) - \min(P_{t,BU})}{\text{TV}_t}$, $(89\%, -)$ $\text{RTVV}_t = \frac{\text{TVV}_t}{\frac{1}{T_w} \sum_{j=0}^{T_w - 1} \text{TV}_{t-j}}$ with $T_w = 6$, $(56\%, +)$ $\text{RTQV}_t = \frac{\text{TQV}_t}{\text{TQ}_{t,T_w}}$ with $T_w = 6$, $(56\%, +)$ $\text{MNRV2RV} = \frac{\text{MNRV}_t}{\text{RV}_t}$ with $T_w = 12$, $(89\%, +)$

σ_t model selection

feature summary	no clear frequency pattern, more volatility measures are selected, MNRV and $\log(\sigma_{t-1})$ being selected 100%
selected predictors (selection frequency > 50%, estimated coefficient sign)	TV_t, $(67\%, +/-)$ TQ_t, $(89\%, +/-)$ $\text{MTQV}_t = \sqrt{\frac{1}{n_t} \sum_{j=1}^{n_t} \left(V_{t,j} - \frac{1}{n_t} \sum_{i=1}^{n_t} V_{t,i} \right)^2}$, $(67\%, +/-)$ $\text{MNRV}_t = \sqrt{\frac{1}{n_t} \sum_{i=1}^{n_t} (\log(P_{t,i}) - \log(P_t))^2}$, $(100\%, +)$ $\text{RollMod}_t^- = \text{RollMod}_t \mathbb{1}\{\text{RollMod}_t < 0\}$ with $\text{RollMod}_t = \frac{\text{cov}\left(\triangle P_{t,T_w}, \triangle P_{t-1.T_w}\right)}{P_t^m}$ and $T_w = 6$, $(56\%, +)$ $\text{MNRV2RV} = \frac{\text{MNRV}_t}{\text{RV}_t}$ with $T_w = 12$, $(78\%, +)$ $\text{MRV2RV} = \frac{\text{MRV}_t}{\text{RV}_t}$ with $T_w = 12$, $(67\%, +/-)$ $\log(\sigma_t)$ $(100\%, +)$

Empirical application: model predictive performance

- VaR backtesting
- Kolmogorov–Smirnov (K-S) test: on the goodness of fit of the predicted GPD over the out-of-sample period[1]

Stock NamesVaR risk level	0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	0.999	0.9999	K-S Test p-values
AXP	0.8999	0.9111	0.9224	0.9314	0.9414	0.9520	0.9631	0.9729	0.9827	0.9913	0.9986	0.9996	0.208
BA	0.9000	0.9098	0.9210	0.9309	0.9417	0.9522	0.9630	0.9738	0.9836	0.9930	0.9985	0.9996	0.0431
GE	0.8999	0.9171	0.9254	0.9376	0.9475	0.9622	0.9709	0.9811	0.9889	0.9943	0.9989	0.9999	2.68×10^{-12}
HD	0.9001	0.9097	0.9217	0.9339	0.9453	0.9551	0.9652	0.9756	0.9840	0.9934	0.9990	0.9998	0.0023
IBM	0.9003	0.9094	0.9204	0.9312	0.9414	0.9509	0.9614	0.9702	0.9809	0.9905	0.9980	0.9993	0.8954
JNJ	0.9000	0.9083	0.9195	0.9284	0.9382	0.9480	0.9571	0.9682	0.9786	0.9888	0.9977	0.9995	0.3322
JPM	0.9000	0.9094	0.9176	0.9294	0.9385	0.9500	0.9612	0.9711	0.9815	0.9904	0.9982	0.9997	0.3993
KO	0.8996	0.9109	0.9180	0.9273	0.9380	0.9489	0.9594	0.9709	0.9813	0.9909	0.9980	0.9994	0.3786
XOM	0.8988	0.9056	0.9136	0.9218	0.9303	0.9396	0.9502	0.9630	0.9742	0.9863	0.9972	0.9992	6.32×10^{-8}

Table: Out-of-sample VaR Coverage Rates and p-values for the K-S Test.

[1](90%/10% in-sample/out-of-sample, i.e. 158991/17665 base unit ratio)

Table of Contents

Conclusion – in response to the research objective

- to model the extreme loss severity distribution of a financial asset in HF markets
  ✓ a suitable model specification allowing for both stationary and local unit-root predictors
- to find a clear association of market indicators with the potential loss severity
  ✓ the two-step adaptive L_1-regularized maximum likelihood estimator (ALMLE2S) with the oracle property to perform automatic variable selection
- to predict financial risk measures (e.g. VaR) for financial risk management
  ✓ using 42 liquidity and volatility predictors, we find the extreme severity distribution to be well predicted by low levels of price impact in period of high volatility of liquidity and volatility