Supplement

-

Efficient estimation in extreme value regression models of hedge funds tail risks

J. Hambuckers[1,†], M. Kratz[2] and A. Usseglio-Carleve[3]

[1] University of Liège - HEC Liège, Belgium.

[2] ESSEC Business School, France

[3] Université d'Avignon, France

[†] Corresponding author. Email: jhambuckers@uliege.be.

Introduction

This supplement provides additional simulation, computational and empirical results for the paper "Efficient estimation in extreme value regression models of hedge fund tail risks". In Section 1, we display additional graphs related to our simulation study, and present the results of a second simulation study, using increasing contamination rates in DGP II from the main manuscript, as well as two novel DGPs. In Section 2, we detail the mean estimation and filtering procedure used to process the data before conducting the tail risk analysis. In Section 3, we first conduct a study of the tail-heaviness of our data, and of the correctness of the Gaussian and exponential assumption for the body of the data. We then report the estimated regression coefficient for σ and s, and provide additional estimates of the tail risk model splitting the funds according to their use of financial leverage or not. We also detail the bootstrap procedure used to construct robust confidence interval for the regression parameters. Finally, in Section 4, we provide additional theoretical details related to our proposed censored likelihood estimation procedure.

1 Additional simulation results

In this section, we present additional simulation results to complement Section 3 of the main manuscript.

1.1 Starting values

As starting values, we use the following quantities throughout the paper for the constant parameters:

- μ_0: the mean of the sample from which we removed the top 20% of observations.
- β_0^s: the log of the mean absolute deviation of the sample, from which we removed the top 20% of observations.
- β_0^ξ: the log of the unconditional MLE computed on the top 5% of the data.
- β_0^σ: the log of the unconditional MLE computed on the top 5% of the data.

All initial estimates of the regression parameters are set to small values close to 0, i.e. 0.001.

In addition, for each DGP, we choose the following values for σ and s:

- $\beta^\sigma = [\log(0.08),\ 0.2,\ -0.2,\ 0,\ 0.2,\ 0]$,
- $\beta^s = [\log(0.045),\ -0.5,\ -0.3,\ 0,\ 0.3,\ 0]$.

1.2 Bias

In Section 3 of the main manuscript, we discuss the bias of the different estimators. In Tables 1 to 3, we report the average bias for each parameter and each DGP. The bias is computed as $\frac{1}{B}\sum_{b=1}^{B}\left(\hat{\beta}_j^\xi - \beta_j^\xi\right)$, $j = 1, \ldots, 5$. Although all methods exhibit a certain degree of bias in DGP III, estimates obtained from `WMLE` and `CWMLE` tend to be closer to the true coefficients than the MLE and the POT-based approaches.

Table 1: Estimated bias from $B = 200$ simulated samples for DGP I.

DGP	Crit.	CPOT99	CPOT95	CPOT90	POT99	POT95	POT90	MLE	WMLE	CWMLE
DGP I	β_0^ξ	-3.947	-1.343	-1.468	-3.431	-1.394	-1.532	-0.010	-0.011	-0.013
	β_1^ξ	0.480	1.261	1.231	0.438	1.138	1.197	-0.008	-0.004	-0.010
	β_2^ξ	0.944	-1.294	-1.487	-0.527	-1.374	-1.604	-0.126	-0.123	-0.120
	β_3^ξ	1.513	1.306	1.528	1.170	1.506	1.758	0.05	0.05	0.054
	β_4^ξ	-1.189	1.167	0.992	-0.679	1.125	1.018	0.045	0.044	0.061
	β_5^ξ	-0.174	0.026	0.026	-0.849	0.047	0.040	-0.012	-0.014	-0.011

1.3 Coverage, power and length of confidence intervals

In Tables 4 to 6, we report median lengths of 95% confidence intervals for the different regression parameters, obtained with the various methods on DPG I to III.

1.4 Additional simulations: extensions of DGP II

We use here, as a basis, the simulation setting of DGP II outlined in Section 3.1. of the main document, and consider two different changes:

Table 2: Estimated bias from $B = 200$ simulated samples for DGP II.

DGP	Crit.	CPOT99	CPOT95	CPOT90	POT99	POT95	POT90	MLE	WMLE	CWMLE
DGP II	β_0^ξ	-2.236	-1.528	-1.673	-2.539	-1.572	-1.636	0.842	0.060	0.059
	β_1^ξ	0.986	1.280	1.453	1.508	1.164	1.362	-0.366	-0.012	-0.041
	β_2^ξ	-1.821	-1.226	-1.329	-1.909	-1.505	-1.632	0.039	-0.228	-0.252
	β_3^ξ	0.342	1.355	1.368	0.785	1.603	1.665	0.160	0.102	0.143
	β_4^ξ	0.487	0.845	1.002	0.413	0.793	0.953	-0.080	0.147	0.157
	β_5^ξ	-0.534	0.419	0.240	-1.048	0.377	0.099	-0.046	0.044	0.054

Table 3: Estimated bias from $B = 200$ simulated samples for DGP III.

DGP	Crit.	CPOT99	CPOT95	CPOT90	POT99	POT95	POT90	MLE	WMLE	CWMLE
DGP III	β_0^ξ	-2.745	-0.875	-0.982	-3.871	-0.756	-0.931	0.950	0.283	0.287
	β_1^ξ	-0.428	0.719	0.789	-0.379	0.599	0.846	-0.225	0.127	0.118
	β_2^ξ	-0.474	-0.594	-0.726	-0.660	-0.463	-0.519	0.836	-0.049	-0.050
	β_3^ξ	-0.054	0.720	0.926	1.979	0.385	0.665	-0.682	0.077	0.078
	β_4^ξ	0.738	0.316	0.574	1.356	0.118	0.332	-0.646	0.013	0.020
	β_5^ξ	0.750	0.107	0.190	0.970	0.100	0.100	0.008	0	-0.002

Table 4: Median length of the 95% confidence intervals for the different estimators, obtained from $B = 200$ samples following DGP I.

DGP	Crit.	CPOT99	CPOT95	CPOT90	POT99	POT95	POT90	MLE	WMLE	CWMLE
DGP I	β_0^ξ	14.62	6.12	4.89	15.50	6.28	4.95	1.92	1.82	1.84
	β_1^ξ	17.02	7.43	6.03	16.92	7.55	6.07	2.36	2.23	2.24
	β_2^ξ	17.45	7.21	5.74	18.12	7.50	5.91	2.33	2.32	2.25
	β_3^ξ	16.52	7.16	5.77	18.76	7.38	5.86	2.36	2.23	2.20
	β_4^ξ	23.06	9.58	7.64	24.01	9.77	7.89	2.99	2.99	2.97
	β_5^ξ	18.02	7.38	5.83	18.89	7.37	5.79	2.27	2.22	2.23

Table 5: Median length of the 95% confidence intervals for the different estimators, obtained from $B = 200$ samples following DGP II.

DGP	Crit.	CPOT99	CPOT95	CPOT90	POT99	POT95	POT90	MLE	WMLE	CWMLE
DGP II	β_0^ξ	12.81	5.96	4.77	14.44	6.27	4.90	3.92	1.92	1.92
	β_1^ξ	15.71	6.95	5.75	16.74	7.41	5.96	4.65	2.38	2.37
	β_2^ξ	15.17	6.97	5.75	17.44	7.39	6.00	4.46	2.27	2.31
	β_3^ξ	115.37	7.11	5.58	15.95	7.56	5.84	4.40	2.34	2.41
	β_4^ξ	20.70	9.57	7.47	22.08	9.72	7.49	6.30	2.97	3.04
	β_5^ξ	15.60	7.16	5.74	16.67	7.52	5.87	5.05	2.23	2.25

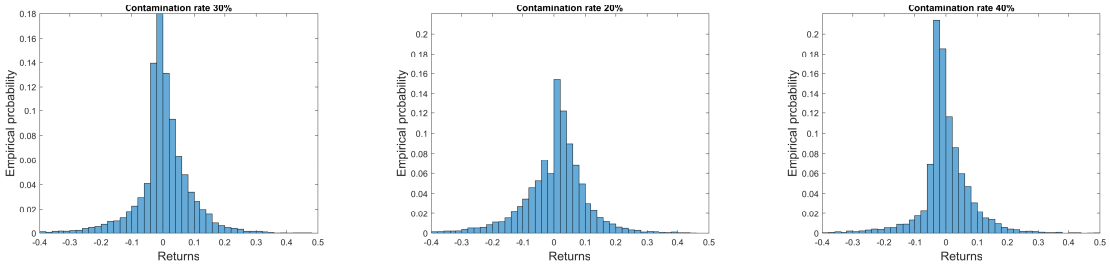
Table 6: Median length of the 95% confidence intervals for the different estimators, obtained from $B = 200$ samples following DGP III.

DGP	Crit.	CPOT99	CPOT95	CPOT90	POT99	POT95	POT90	MLE	WMLE	CWMLE
DGP III	β_0^ξ	14.20	5.47	4.06	14.09	5.28	4.00	3.17	1.46	1.50
	β_1^ξ	17.78	6.65	4.74	17.71	6.42	4.78	3.84	1.73	1.73
	β_2^ξ	16.90	6.46	4.82	18.16	6.37	4.74	3.92	1.74	1.73
	β_3^ξ	16.62	6.59	4.80	17.93	6.31	4.74	3.82	1.74	1.74
	β_4^ξ	21.62	8.72	6.29	22.94	8.34	6.14	5.20	2.23	2.22
	β_5^ξ	16.91	6.78	4.86	18.21	6.68	4.81	3.86	1.75	1.74

(1) On the one hand, we remove the effects of the covariates in s. As such, we reduce the heteroscedasticity in the data, which could generate additional estimation uncertainty in the conditional POT based on quantile regression (notice that, since we assume σ and ξ to be functions of the covariates, we still have some heteroscedasticity, but only driven by tail observations). We keep the contamination rate at 30%, as in DGP II.

(2) On the other hand, instead of contaminating the 30% smallest observations, we use two other contamination rates c, namely 20% and 40%. All the other parameters of the G-E-GPD and GPD distributions remain unchanged. With a larger contamination rate, we expect to favor the POT more compared to DGP II, while with a smaller contamination rate, we should favor the MLE.

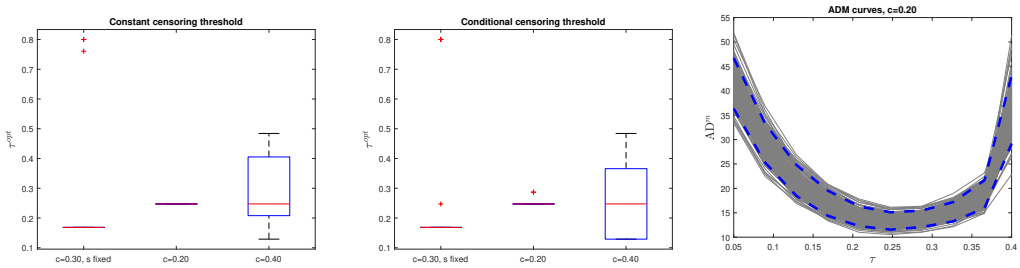
In Figure 1, we display representative histograms of the simulated data. In Figure 2, we portray the selected τ^{opt} for the different DGPs, and an example of AD^m curves for $c = 0.40$ (see Equation (12) and following in the main document, for exact definitions of these two quantities).



(i) $c = 0.30$, fixed s (ii) $c = 0.20$ (iii) $c = 0.40$

Figure 1: Simulated data obtained with different contamination thresholds c. For $c = 0.30$, we also impose s to be constant.

In Figures 3 and 4, we report the estimated regression coefficient for ξ and the corresponding RMSE, scaled by the RMSE of the MLE. Estimates obtained from the G-E-GPD model with a censoring estimation procedure (referred as WMLE) are virtually unbiased and much less dispersed



(i) (ii) (iii)

Figure 2: Selected censoring thresholds using either an unconditional quantile (panel (i)) or a conditional quantile (panel (ii)) in the weight function. Panel (iii): AD^m curves for $c = 20\%$. Blue: 5% and 95% empirical quantile of AD^m for fixed values of τ.

than the POT-based estimates or the classical MLE. In terms of RMSE, WMLE and CWMLE are much better than MLE in all settings. Gains in RMSE with respect to the MLE varies between 18% and 58%, and are the strongest for β_0^ξ.

In Figure 5, we displays the full power curves for the null hypotheses $H_0 : \beta_j^\xi = \beta_j^{H0}$, $j = 0, 1, 5$, for different values of β_j^{H0} increasingly far from the true parameter. WMLE systematically provides a good trade-off between size and powers, with its coverage rate being close to the nominal 95% (dashed red curve) and its power being equivalent to or dominating the power of the other methods. One may observe a stable performance across DGPs for the POT-based approaches and WMLE, however with a clear superiority of WMLE. On the contrary, since the MLE is not a robust method, its performance is bad for any degree of contamination rates.

In conclusion, these results confirm the superiority of the G-E-GPD approach with the censoring estimation method over the POT-based methods, for these new variants of DGP II.

1.5 Additional simulations: increased misspecifications (DGP IV)

In this last simulation section, we investigate the performance of a fourth DGP. We first assume that y_{it} follows a Fréchet distribution with parameters ξ_{it}, σ_{it} and μ_0. Then, we introduce an additional misspecification in the left tail, similar to DGP II and III, by resampling the smallest 30% of the observations from a GPD. Due to the highly skewed nature of the Fréchet distribution, we expect to favour more the POT, as the underlying true model will be further away from the G-E-GPD auxilliary model. Covariates are simulated following the same equations as in the other DGPs.

In Figure 6, we display the histogram, selected τ^{opt} and AD^m curves for the DGP. The data have a skewed distribution, with heavy tails on both sides. We always select a censoring threshold around 0.15, with the AD^m curve exhibiting a clear minimum. In terms of RMSE (Figure 6, bottom left), although we see a large improvement for the POT methods, our approach is still vastly superior: We observe a reduction in relative RMSE between 25% and 40% for β_0^ξ to β_4^ξ. Only for β_5^ξ (coefficient related to the fund-specific covariate) we do observe a negligible improvement of 2%. Then, in Figure

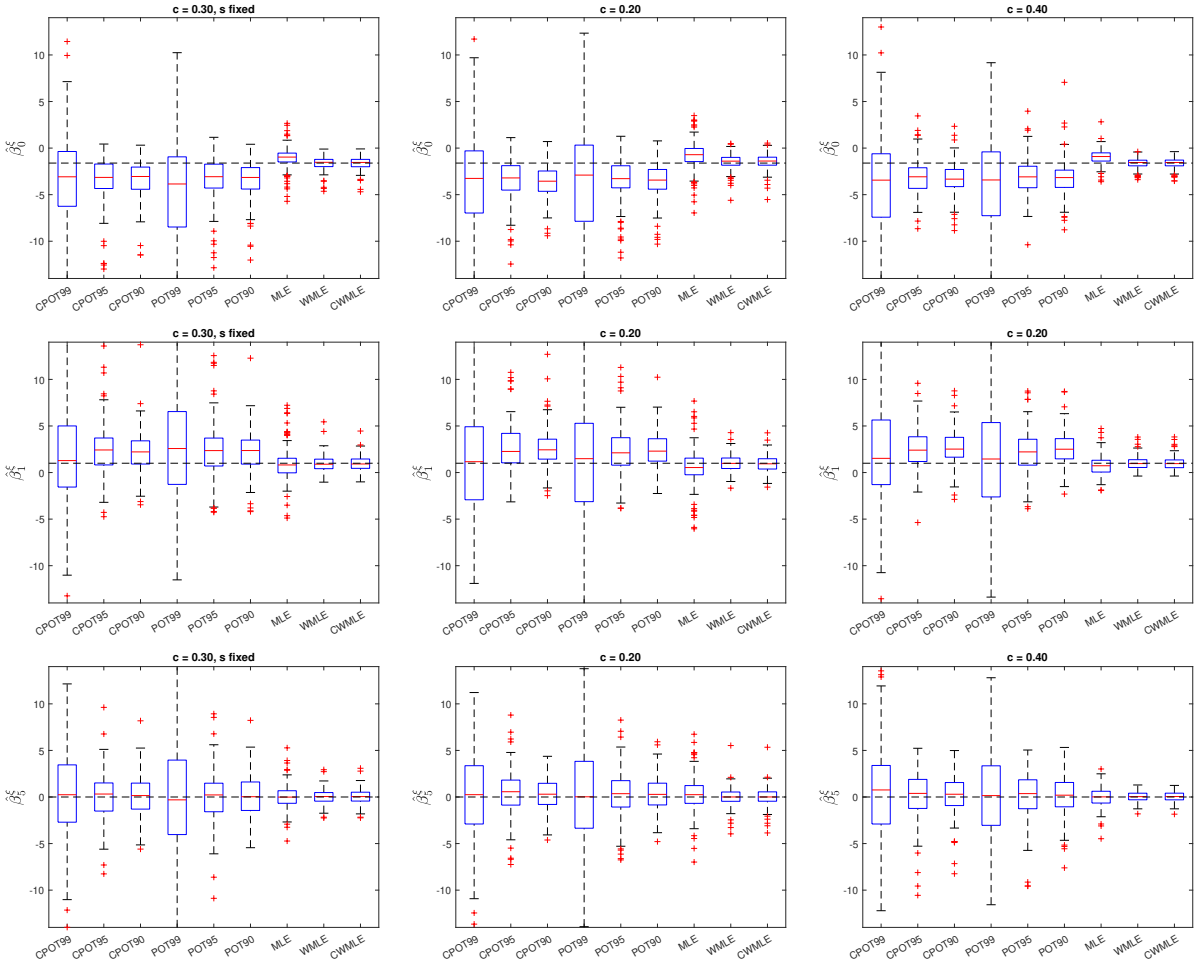
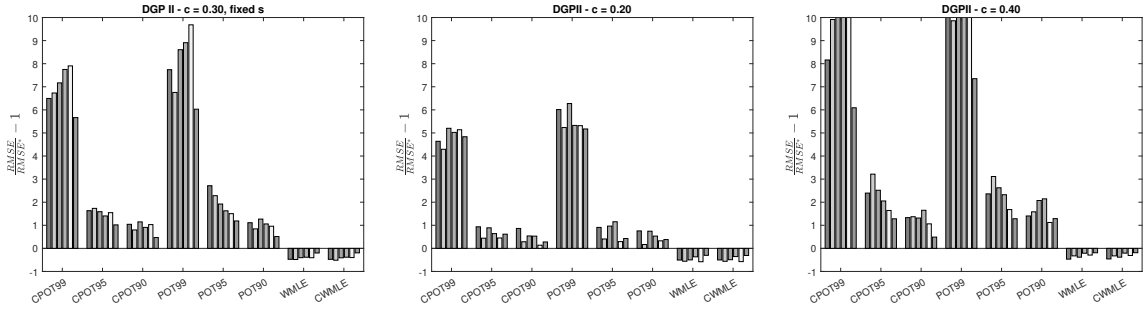


Figure 3: Estimated parameters $\hat{\beta}_0^{\xi}$ (top), $\hat{\beta}_1^{\xi}$, and $\hat{\beta}_5^{\xi}$ (bottom) obtained with the proposed method and various POT approaches. Dashed: values of the true parameters. Columns `CPOT99`, `CPOT95` and `CPOT90` display the results obtained with the (conditional) POT approach, using quantile regression at the levels 99%, 95%, and 90% for the threshold. Columns `POT99`, `POT95` and `POT90` refer to the results of POT obtained with an unconditional empirical quantile at the same levels. Columns `WMLE` and `CWMLE` denote the results obtained with the censored likelihood procedure, using either a global censoring threshold, or a censoring threshold based on quantile regression. `MLE` relates to classical MLE.



(i) (ii) (iii)

Figure 4: $RMSE/RMSE^* - 1$ for $\hat{\beta}_j^{\xi}$, $j = 1, \ldots, 5$ (from left to right), for the DGP without heteroscedasticity (left) and the different contamination rates (middle and right).

7, we display the estimated regression coefficients for the tail parameter ξ_{it}. Although we observe a certain degree of bias with the proposed method, the estimates are much closer to the true values than

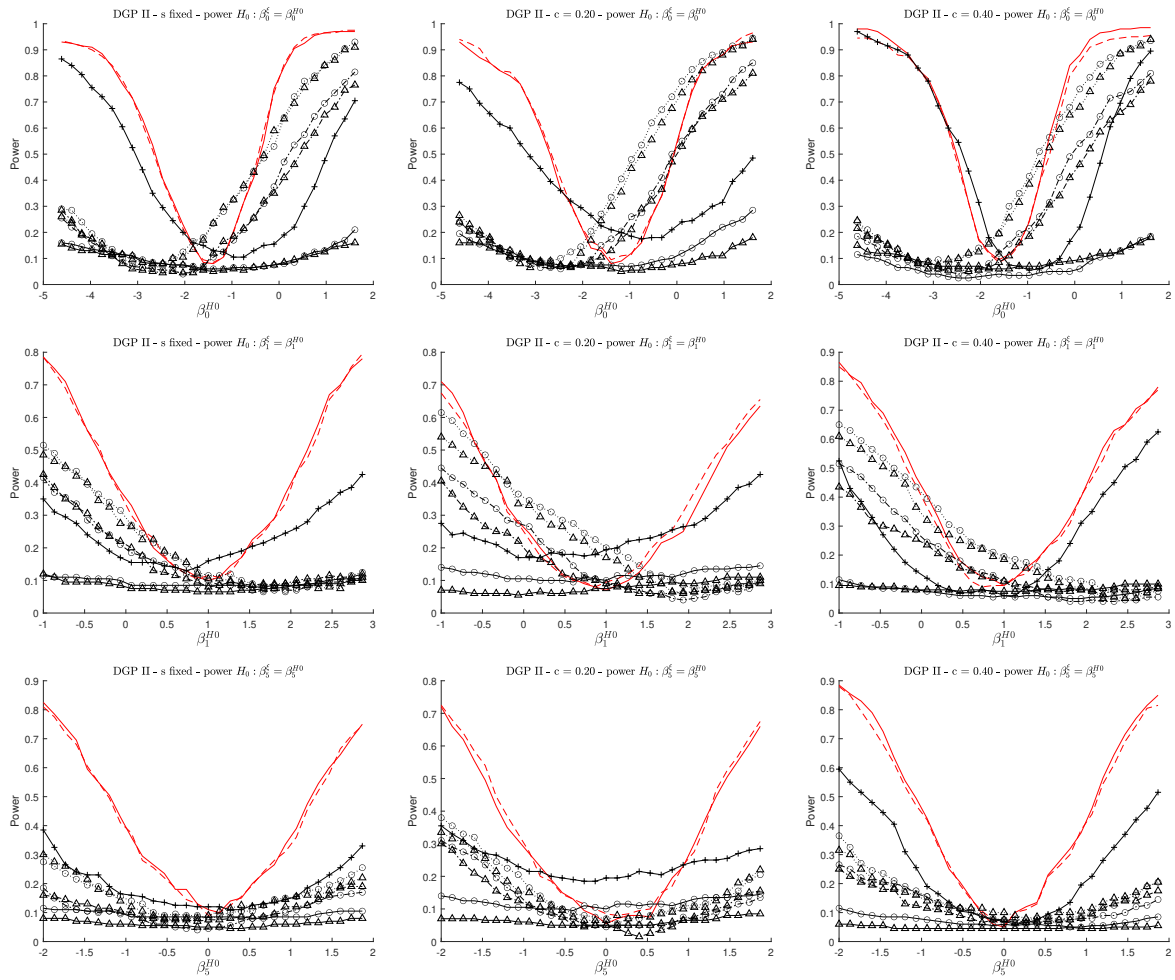


Figure 5: Power curves for the null hypotheses $H_0 : \beta_j^\xi = \beta_j^{H0}$, $j = 0, 1, 5$, for different values of β_j^{H0}. From left to right: DGP I to DGP III. For β_j^{H0} equal to the true parameter, we expect a rejection rate equal to the 5% test level. Red solid and dashed: `WMLE` and `CWMLE`. $+$: `MLE`. $\triangle$: POT estimates based on unconditional thresholds. o: POT estimates based on conditional thresholds. Thresholds are taken as quantiles at the 99% (black solid), 95% (dashed dotted) and 90% (dotted) levels in the POT approaches.

the POT-based approach. Finally, in Figure 8, we display the power curves for the various regression coefficients. Our conclusions are qualitatively similar to those of DGP I to III, and to the ones made for the extended DGP II in Section 1.4.

1.6 Additional simulations: DGP I to III for $T = 1,000$

In this subsection, we display several graphs summarizing the results obtained with $T = 1,000$ instead of $T = 250$. The idea here is to be closer to the asymptotic behavior and especially to eliminate strong finite-sample bias of the POT approach. We only report the relative RMSEs and point estimates for $\hat{\boldsymbol{\beta}}^\xi$, the confidence intervals being particularly time-consuming to compute for such large samples.

Overall, regarding point estimates, our conclusions stay alike. We observe a reduction of the dispersion for the POT-based methods, but they all stay biased compared to the WMLE and CWMLE (Figures 9 to 11). Interestingly, from an RMSE standpoint, some of the POT-based methods become

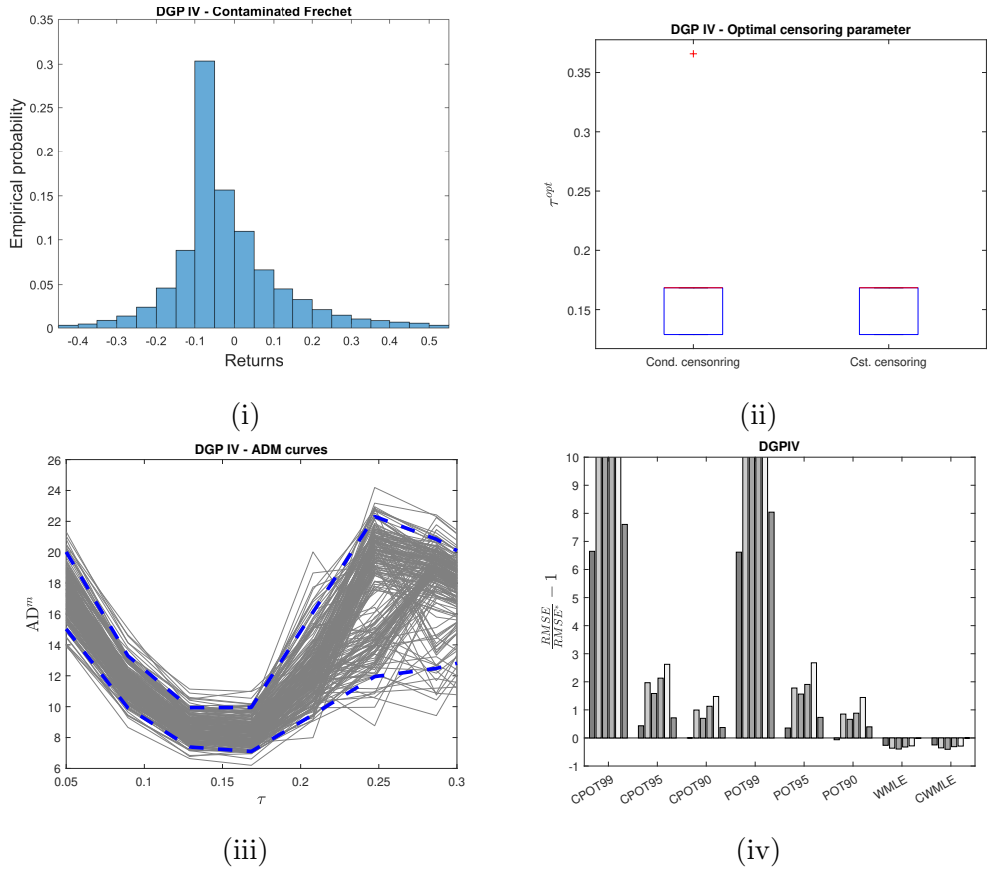


Figure 6: (i) Simulated data obtained with DGP IV. (ii) Selected τ^{opt}. (iii) AD^m curves (iv) RMSE for the various parameters.

more attractive for DGP III (misspecified case); see Figure 12. This result illustrates that, in practice, one needs very large samples to truly benefit from the asymptotic properties of the EVT in combination with the POT.

2 Pre-processing of the data

Data have been extracted from the EurekaHedge data, for funds reporting a *Long/Short Equity* trading strategy over the period 01/1995-05/2021. To address backfill and survivorship biases, we removed the first 12 months reported by each fund, and solely kept funds with at least 60 months of uninterrupted reported history. In addition, we included both dead and live funds in our analysis. Our final sample consists of roughly 189,000 monthly returns spanning 1,484 funds.

In Section 4 of the main manuscript, we conduct our analysis on hedge fund returns from which we removed the conditional mean, following the idea of Kelly and Jiang [2014]. To estimate this conditional mean component, we use the high-frequency asset pricing model of Patton and Ramadorai [2013]. Patton and Ramadorai [2013] propose refining the classical seven-factor asset pricing model of Fung and Hsieh [2004] to account for intra-month variations in exposure to risk factors. Starting

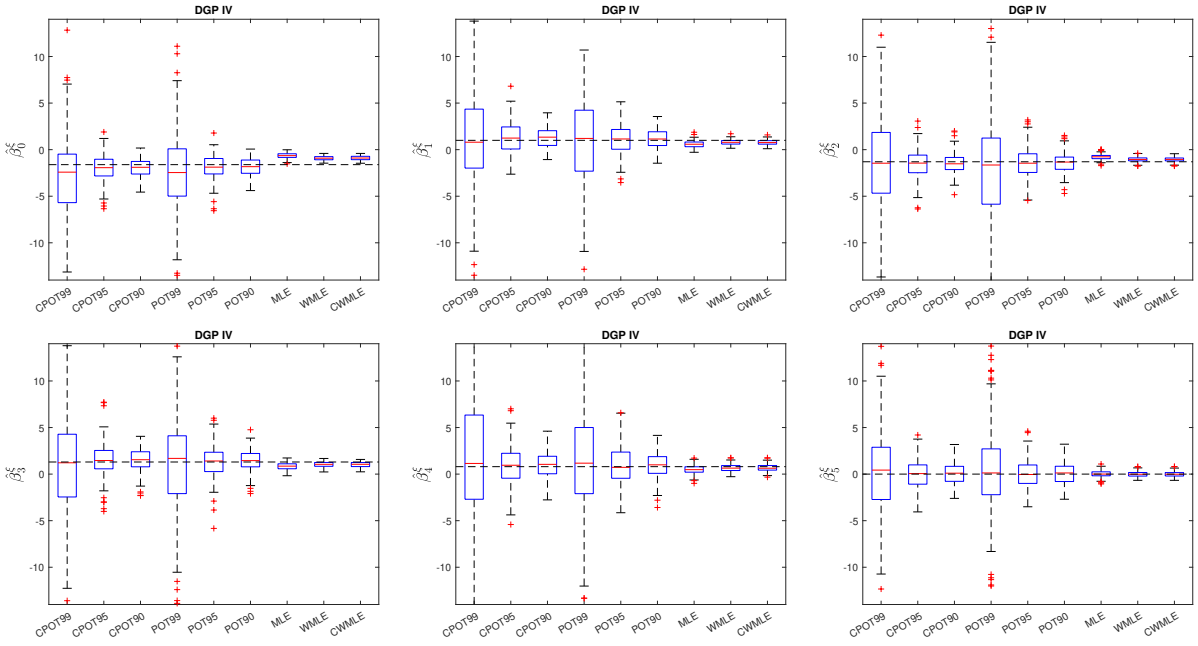


Figure 7: Estimated parameters $\hat{\beta}_0^\xi$ to $\hat{\beta}_5^\xi$ obtained with the proposed method and various POT approaches for DGP IV. Dashed: values of the true parameters. Columns `CPOT99`, `CPOT95` and `CPOT90` display the results obtained with the (conditional) POT approach, using quantile regression at the levels 99%, 95%, and 90% for the threshold. Columns `POT99`, `POT95` and `POT90` refer to the results of POT obtained with an unconditional empirical quantile at these levels. Columns `WMLE` and `CWMLE` denote the results obtained with the censored likelihood procedure, using either a global censoring threshold, or a censoring threshold based on quantile regression. `MLE` relates to classical MLE.

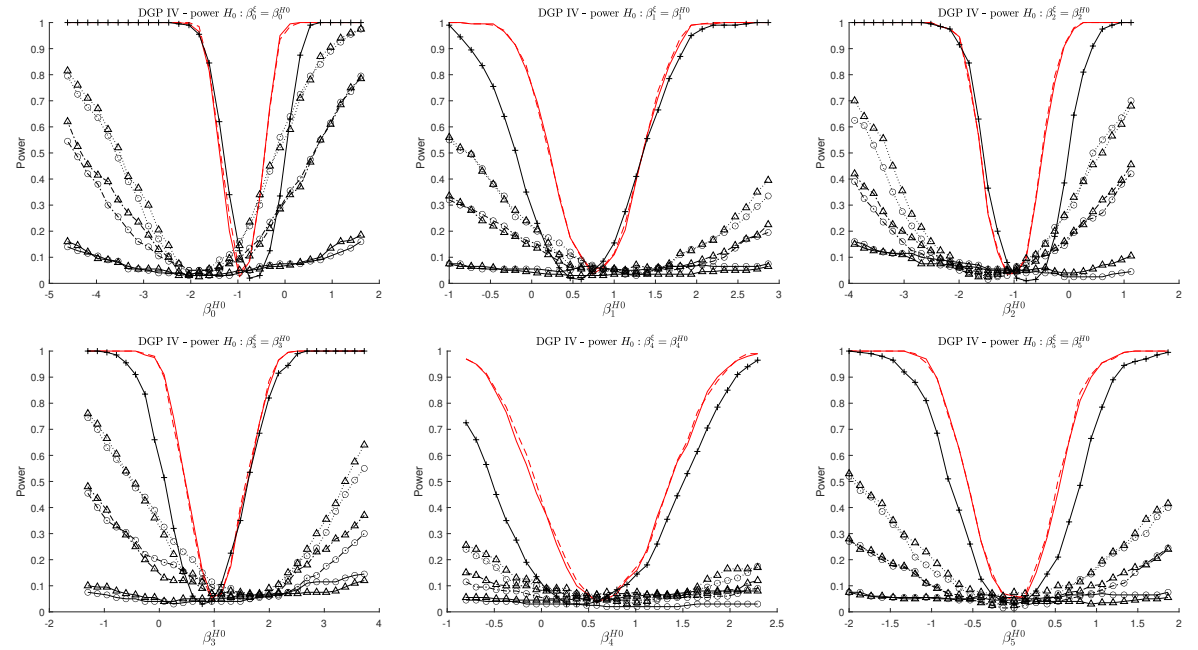


Figure 8: For DGP IV, power curves for the null hypotheses $H_0 : \beta_j^\xi = \beta_j^{H0}$, $j = 0, \ldots, 5$, for different values of β_j^{H0}. For β_j^{H0} equal to the true parameter, we expect a rejection rate equal to the 5% test level. Red solid and dashed: `WMLE` and `CWMLE`. $+$: `MLE`. $\triangle$: POT estimates based on unconditional thresholds. o: POT estimates based on conditional thresholds. Thresholds are taken as quantiles at the 99% (black solid), 95% (dashed dotted) and 90% (dotted) levels in the POT approaches.

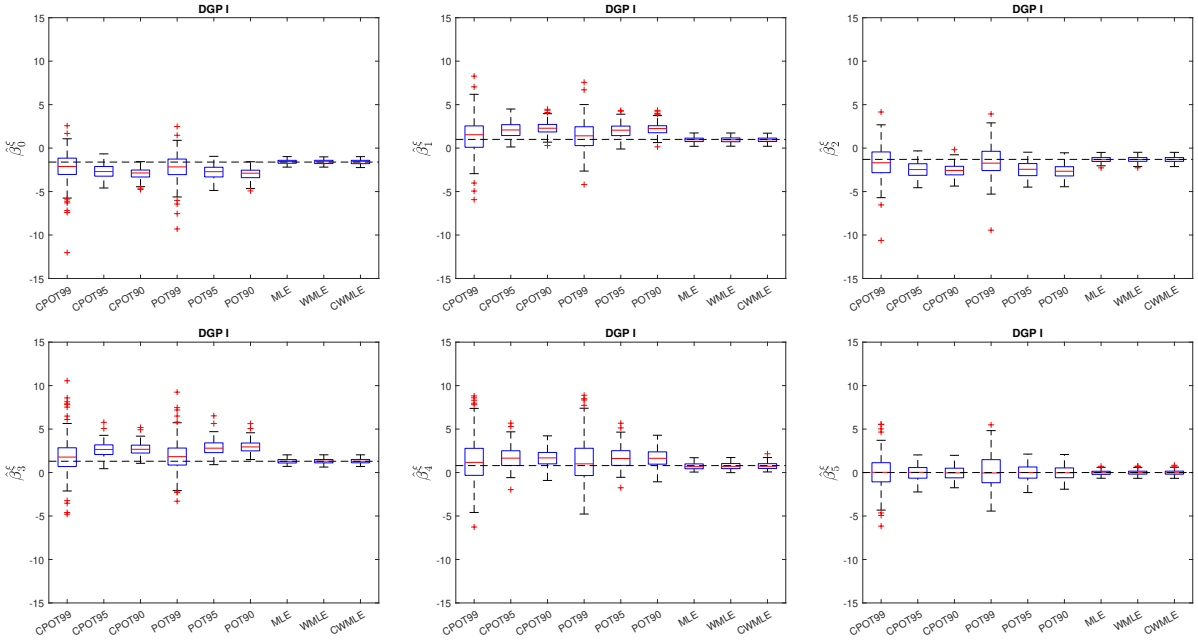


Figure 9: Estimated parameters $\hat{\beta}_0^{\xi}$ to $\hat{\beta}_5^{\xi}$ for DGP I, for $T = 1,000$. See previous figures for a detailed legend.

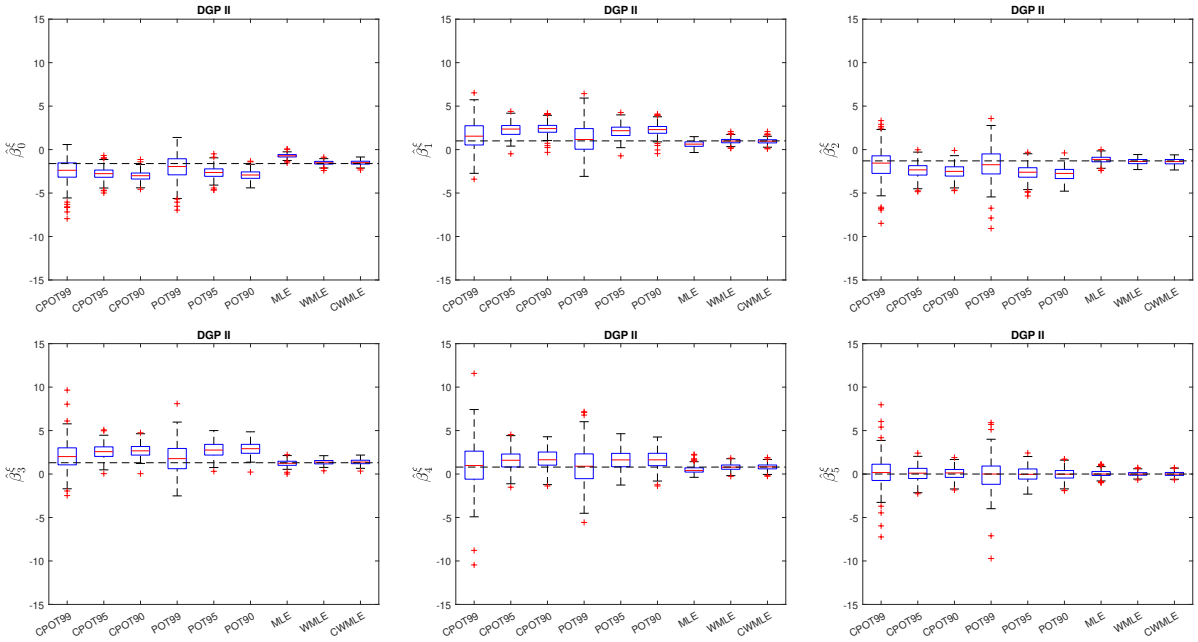


Figure 10: Estimated parameters $\hat{\beta}_0^{\xi}$ to $\hat{\beta}_5^{\xi}$ for DGP II, for $T = 1,000$. See previous figures for a detailed legend.

from the model of Fung and Hsieh [2004], we have

$$r_{it} = \alpha_i + \sum_{j}^{p} \beta_{ijt} f_{jt} + y_{it}, \tag{1}$$

where r_{it} denotes monthly returns of fund i at time t. Patton and Ramadorai [2013] assume in addition that intra-month returns (e.g. daily returns) are given by:

$$r_{id}^{\star} = \alpha_i + \sum_{j=1}^{p} \beta_{ijd} f_{jd}^{\star} + \varepsilon_{id}^{\star}, \tag{2}$$

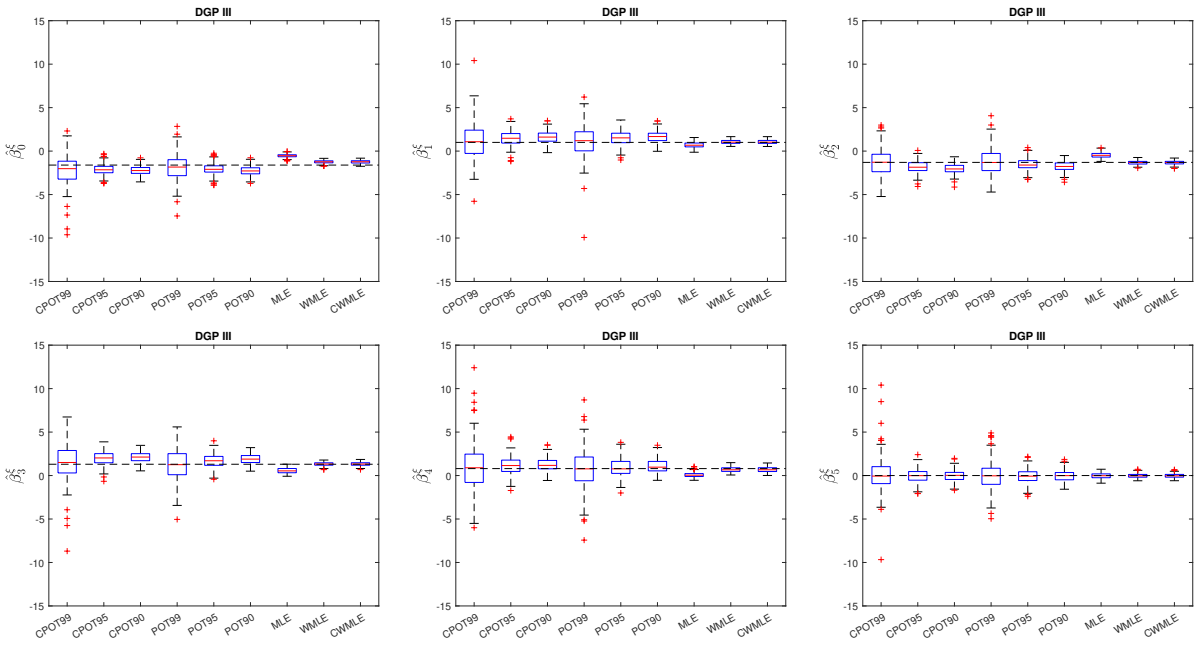
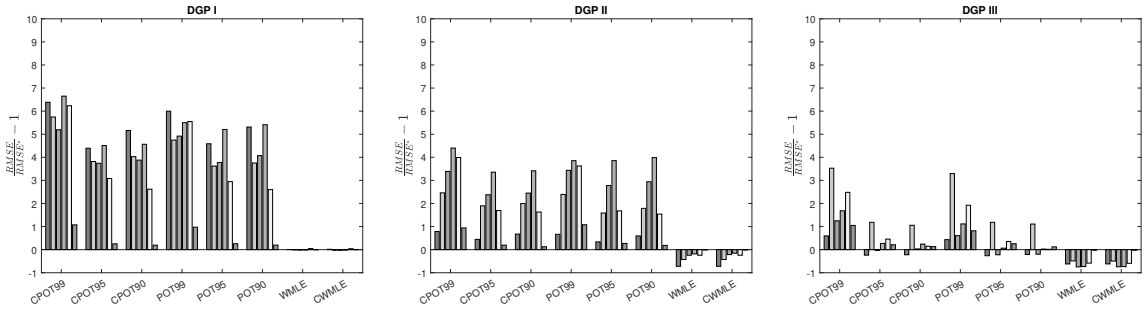


Figure 11: Estimated parameters $\hat{\beta}_0^{\xi}$ to $\hat{\beta}_5^{\xi}$ for DGP III, for $T = 1,000$. See previous figures for a detailed legend.



(i) (ii) (iii)

Figure 12: $RMSE/RMSE^* - 1$ for $\hat{\beta}_j^{\xi}$, $j = 1, \ldots, 5$ (from left to right), for DGP I to III and $T = 1,000$.

where r_{id} is the daily return of fund i on day d, for $d \in \mathcal{M}(t)$ with $\mathcal{M}(t)$ the set of days belonging to month t, $f_{jd}^{\star}$ is the j-th risk factor measured on day t and β_{ijd} describes the time-variant exposure to this factor for fund i. The parameter β_{ijd} is made time-variant by conditioning it to a covariate Z measured at the monthly and daily frequencies, and common to all funds. We assume the following dynamics for β_{ijd}:

$$\beta_{ijd} = \beta_{0ij} + \gamma_{ij} Z_{t-1} + \delta_{ij} Z_{d-1}^{\star}. \tag{3}$$

Here, Z_{t-1} is a monthly covariate and $Z_{d-1}^{\star}$ is measured at a daily frequency[1]. Substituting (3) in (2), we obtain

$$r_{id}^{\star} = \alpha_i + \sum_{j=1}^{p} \beta_{0ij} f_{jd}^{\star} + \sum_{j=1}^{p} \gamma_{ij} Z_{t-1} f_{jd}^{\star} + \sum_{j=1}^{p} \delta_{ij} Z_{d-1}^{\star} f_{jd}^{\star} + \varepsilon_{id}^{\star}. \tag{4}$$

[1]This equation can be extended to the case of K covariates driving β_{ijd}, but for simplicity of exposition, we consider here that only one variable is involved in changes of intra-month risk exposure.

Returns on individual hedge funds being only available at a monthly frequency, we need to express monthly returns from the daily returns. Denoting by n_t the number of observations in the set $\mathcal{M}(t)$, we assume that:

$$r_{it} = \sum_{d \in \mathcal{M}(t)} r^{\star}_{id}, \tag{5}$$

such that

$$r_{it} = \alpha_i n_t + \sum_{j}^{p} \beta_{0ij} \sum_{d \in \mathcal{M}(t)} f^{\star}_{jd} + \sum_{j}^{p} \gamma_{ij} \bar{f}_{jt} Z_{t-1} + \sum_{j}^{p} \delta_{ij} \sum_{d \in \mathcal{M}(t)} f^{\star}_{jd} Z^{\star}_{d-1} + \sum_{d \in \mathcal{M}(t)} \epsilon^{\star}_{id}. \tag{6}$$

Following Patton and Ramadorai [2013], we consider up to two variables as conditioning covariates $\mathbf{Z}_t = (Z_{t1}, Z_{t2})$ and $\mathbf{Z}^*_d = (Z^*_{d1}, Z^*_{d2})$:

1. Market liquidity, measured by the TED spread (3-month LIBOR rate minus the 3-month T-Bill rate)

2. Market volatility, proxied with the VIX.

Regarding the risk factors (f_{jt} and f^*_{jd}), we use the four factors, out of the seven provided by Fung and Hsieh [2004], that are available on a daily basis:

1. The equity risk factor (SP500), measured by the excess return of the S&P500;

2. The small-minus-big factor (SMB), measured by the difference between the Russell 2000 index and the S&P500 index returns;

3. The bond market factor (TCM10Y) measured by the change in the 10-year treasury constant maturity yield;

4. The credit spread factor (BAAMTSY) measured by the change in the Moody's Baa yield minus the 10-year treasury constant maturity yield.

To obtain a parsimonious model, we perform both model estimation and selection simultaneously using an adaptive LASSO approach [Zou, 2006]. The penalization parameter is chosen using the Bayesian information criterion (BIC). Final estimates of the coefficients are obtained from the post-LASSO estimator (i.e. the OLS estimates using only the active set identified by the LASSO at the model selection step). Our tail risk analysis is then conducted on the estimated residuals of this model, i.e. on

$$\hat{y}_{it} = r_{it} - \hat{\alpha}_i + \sum_{j=1}^{p} \hat{\beta}_{ijt} f_{jt}, \tag{7}$$

that we multiply by -1 to focus on the upper tail.

3 Additional empirical results

In this section, we present additional results to the main analysis, as well as the results of two additional regression models, complementing Section 4 of the main manuscript.

3.1 Heavy-tailed behavior of the data

In this subsection, we propose to justify the heavy-tail assumption of the hedge fund returns used in our empirical analysis. Then, for each month t (from September 2002 to September 2021), we estimate the Extreme Value Index (EVI) of all the n_t available negative residuals $y_{t,1} \leq \ldots \leq y_{t,n_t}$. For that purpose, we propose to use the moment estimator (see Dekkers et al. [1989]), defined by

$$\widehat{\xi}_{t,k} = M^{(1)} + 1 - \frac{1}{2}\left[1 - \frac{(M^{(1)})^2}{M^{(2)}}\right]^{-1}, \; M^{(j)} = \frac{1}{k}\sum_{i=1}^{k} \log\left(\frac{r_{t,n_t-i+1}}{r_{t,n_t-k}}\right)^j,$$

where $1 \leq k \leq n_t - 1$ has to be carefully chosen. Note that this estimator does not assume the heavy-tail assumption $\xi_t > 0$. The asymptotic normality of $\widehat{\xi}_{t,k}$ also allows us to provide, for each t, a 95% confidence interval for ξ_t. A lower bound greater than 0 would thus make the heavy-tail assumption reasonable. Due to a lack of data at some months (less than 70 observations before September 2003), we merged the residuals at time t with those at time $t-1$ and $t+1$ and computed $\widehat{\xi}_{t,k}$ for all t with $k_t = \lfloor n_t/4 \rfloor$. The estimates (with their associated confidence intervals) are reported in Figure 13.

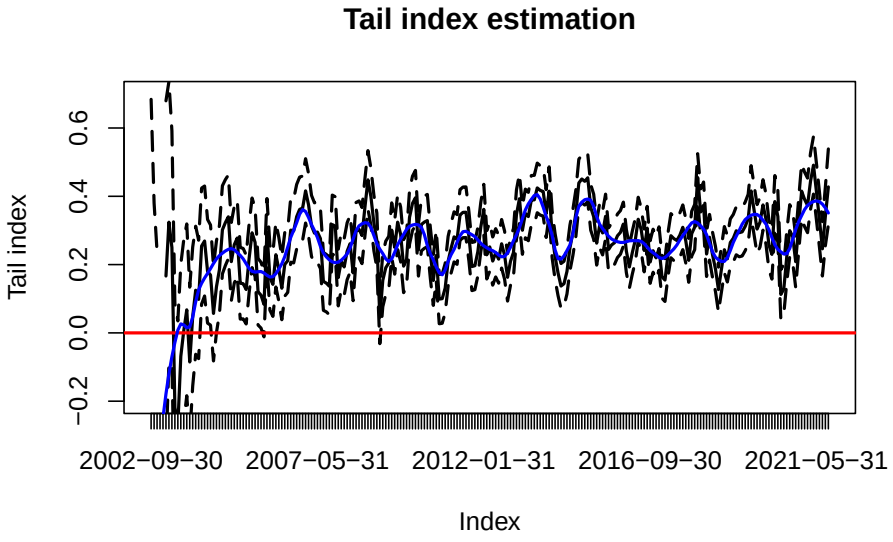


Figure 13: EVI estimation of the negative residuals (solid black curve, and its smoothed version in blue, obtained using the `R` function `loess` with `span` parameter 0.1) at each month, and its associated confidence interval (dashed black curves). The red curve has equation $y = 0$.

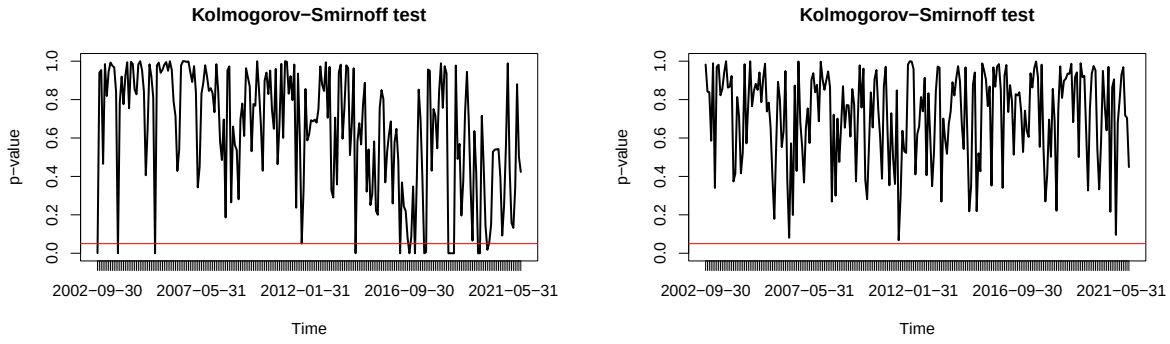
It is interesting to notice that 96.9 % of the EVI estimates and 92.1 % of the lower bounds are greater than 0. With a few exceptions (mainly corresponding to months with a few data), the heavy tail assumption thus seems to be reasonable.

3.2 Normality and exponential behavior of the data

Similarly to the heavy-tail behaviour of the data, we now investigate briefly the correctness, on the one hand, of the normality assumption for observations in the body of the distribution, and on the other hand, the exponential assumption for observations in the upper part of the distribution (but not extreme).

To conduct our tests, we define that an observation belongs to the body of the distribution if it takes a value between the empirical quantiles at levels 40% and 80%, in a given month. Similarly, we assume that observations between the empirical quantiles at levels 80% and 95% of a given month belong to the exponential bridge. We then fit, on a month-by-month basis, truncated Gaussian and truncated exponential distributions on the subsamples. We conduct Kolmogorov-Smirnov goodness-of-fit tests for a uniform distribution after probability integral transformation of our data.

Figure 14 reports the p-values of the tests for each month (left panel for normality, right panel for exponential). We only reject the normality assumption 16 months (7%, slightly higher than the type I error of 5%). The exponential assumption is never rejected. Inspecting manually the histogram of the pseudo-residuals, the fit appears satisfactory. In Figure 15, we display illustrative histograms for April 2016 (a month where the normality of the body is rejected) and February 2015. For the normality assumption on April 2016, one sees departures from uniformity, while the adjustment appears reasonably good for the other histograms.

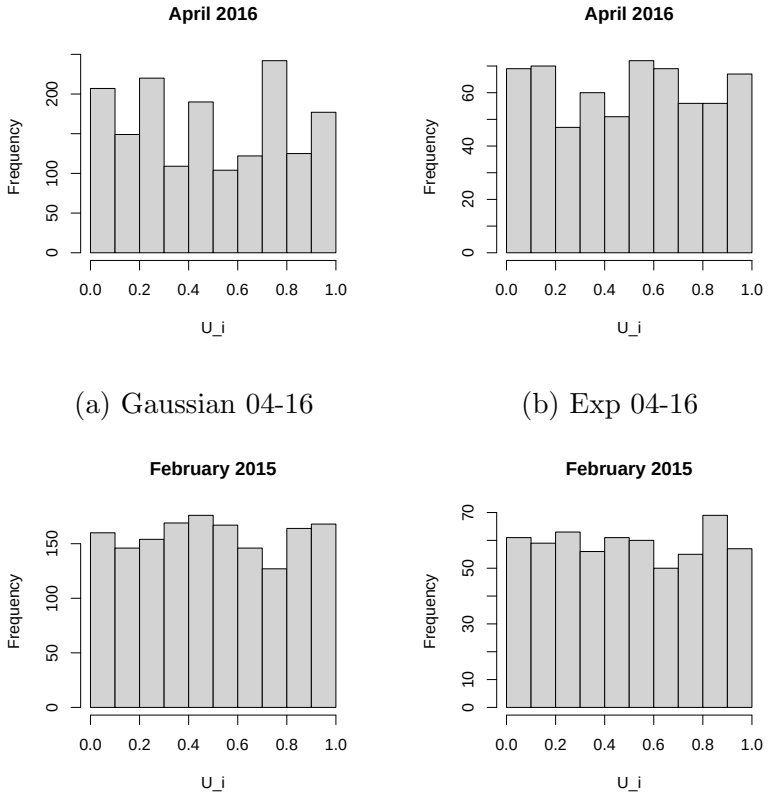


(a) Gaussian (b) Exponential

Figure 14: P-values of Kolmogorov-Smirnov test for (a) Normality or (b) exponential distribution, conducted on a month-by-month basis.

3.3 Additional results to the main analysis

First, in Figure 16, we report the estimated regression coefficients for σ_t. These coefficients are hardly interpretable, since they are not invariant to the true (unknown) full density, contrary to ξ_t. Nevertheless, allowing covariates to enter the scale parameters gives more flexibility to the overall estimation procedure. We observe that several coefficients are found to be significantly different from



(a) Gaussian 04-16 (b) Exp 04-16

(c) Gaussian 02-15 (d) Exp 02-15

Figure 15: Pseudo-residuals obtained from probability integral transformations for April 2020 (top) and December 2011 (bottom), and either a normal (left) or an exponential (right) distribution.

zero, justifying the use of these covariates in our model. For clarity, we only report `WMLE`, `MLE` and `POT97.5` estimates.

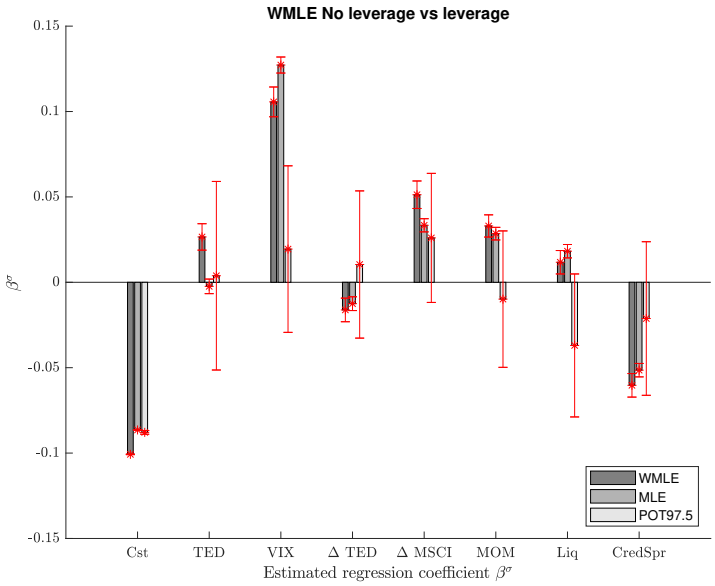


Figure 16: Estimated regression effects for the scale parameter (σ), obtained with `WMLE`, `MLE` or `POT97.5`. Red: 95% confidence intervals. Notice that, for clarity of the graph, the constant has been divided by a factor of 40.

Then, in Figure 17, we reproduce our main regression analysis from Section 4, however splitting the sample of funds between non-leveraged (dark gray) and leveraged ones light gray). Results are similar across groups and with respect to the main analysis. One interesting difference is related to the effect for TED. Tail risks of leveraged funds appear to be more sensitive to funding liquidity levels (TED). For clarity, we only report the results obtained with WMLE.

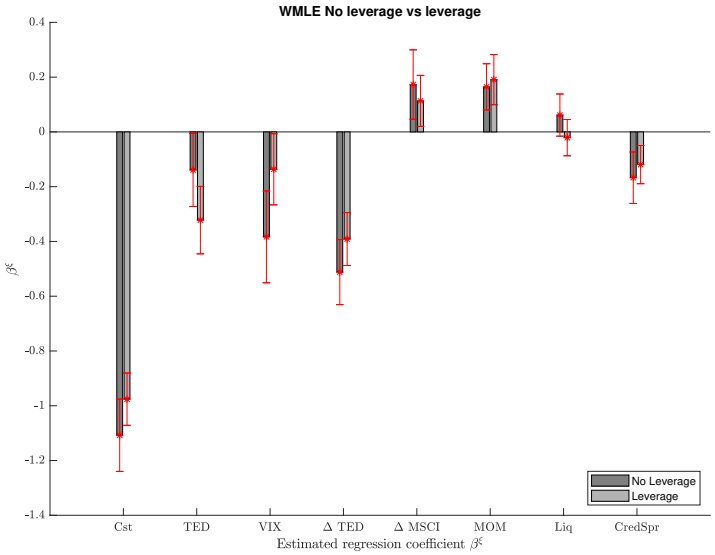


Figure 17: Estimated regression effects obtained with WMLE, for the funds recorded as not leveraged (first column) or leveraged (second columns) obtained with WMLE. Red: 95% confidence intervals.

3.4 Descriptive statistics of the alpha's distribution

In Tables 7 and 8 , we report descriptive statistics of the estimated annualized $\hat{\alpha}_i^*$ obtained with the classical asset pricing models. For the CAPM and the FH model, we find average $\hat{\alpha}_i^*$ values close to one another, as reported in Ardia et al. [2023], at 2.11% and 2.14% respectively. Consistent with the findings of Ardia et al. [2023], the JKKT model delivers a smaller average $\hat{\alpha}_i^*$ (0.92%) and a higher average adjusted R^2. At the strategy level, similar observations can be made for the average $\hat{\alpha}_i^*$, while the average adjusted R^2 is particularly good for *L/S Equity* funds. In Figure 18, we display the empirical distribution of the obtained alpha, according to the asset pricing model and the strategy of the funds. While the distribution of $\hat{\alpha}_i^*$ obtained with the CAPM displays a pronounced skewness in favor of high-alpha funds, this effect disappears almost entirely with the JKKT model. Finally, we compute π^+ (see Figure 19 for each asset pricing model, either for the entire cross-section of funds or strategy-by-strategy.

3.5 Contemporaneous regressions

In this subsection, we briefly comment the results obtained when using contemporaneous $\hat{\xi}(\mathbf{x}_t)$ instead of lagged $\hat{\xi}(\mathbf{x}_{t-1})$ in our predictive regressions. Notice that, in this setting, our contemporaneous tail

Table 7: Descriptive statistics for the cross-sectional distribution of $\hat{\alpha}_i^*$ (in %, annualized) for the full sample.

All funds			
Model	CAPM	FH	JKKT
$\bar{\hat{\alpha}}_i^*$	2.11	2.14	0.92
std($\hat{\alpha}_i^*$)	5.55	5.71	6.77
Adj. R^2	0.19	0.24	0.26

Table 8: Descriptive statistics for the cross-sectional distribution of $\hat{\alpha}_i^*$ (in %, annualized) sorted along their investment strategies.

Strat.	L/S Equity funds			CTA/Managed futures		
Model	CAPM	FH	JKKT	CAPM	FH	JKKT
$\bar{\hat{\alpha}}_i^*$	2.13	2.22	1.00	2.35	2.13	0.68
std($\hat{\alpha}_i^*$)	5.72	5.85	6.57	5.57	6.07	8.16
Adj. R^2	0.24	0.27	0.32	0.04	0.13	0.11
Strat.	Fixed Income			Macro		
Model	CAPM	FH	JKKT	CAPM	FH	JKKT
$\bar{\hat{\alpha}}_i^*$	1.86	2.14	0.69	1.79	1.68	1.42
std($\hat{\alpha}_i^*$)	4.51	4.55	5.16	6.03	5.76	6.91
Adj. R^2	0.23	0.32	0.31	0.14	0.20	0.19

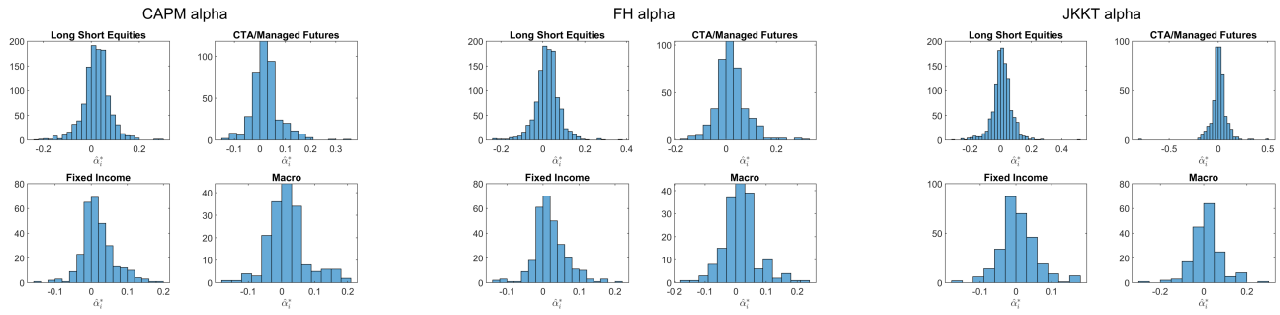


Figure 18: Empirical distribution of $\hat{\alpha}_i^*$ obtained with the different asset pricing models and for the different strategies.

risk measures may be endogenous to the returns since they are built from residuals of the returns themselves. In Figure 20 we display the proportions π^{+-} obtained with the different asset pricing models and tail risk measures. For the CAPM, HKU explains clearly a larger number of funds' returns than the two other measures. The difference is less noticeable for FH and JKKT, in particular with KJ, which displays similar proportions. Proportions π^{+-} are larger than for the regressions with lagged values, taking values ranging between 25% and 45%. This result showcases the potential, for our risk measure, to explain the cross-section of hedge funds returns.

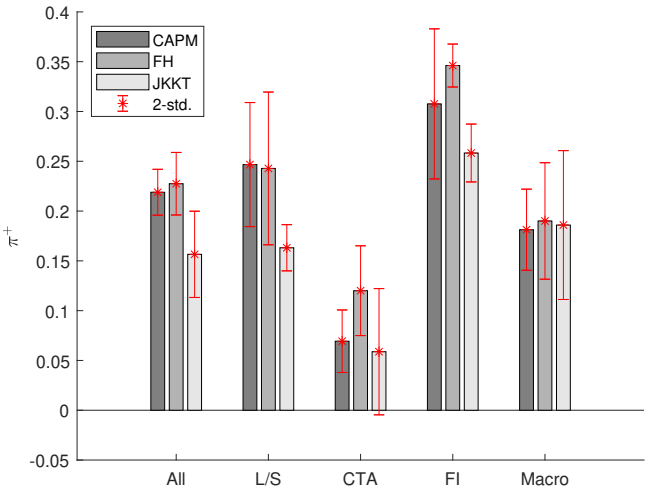
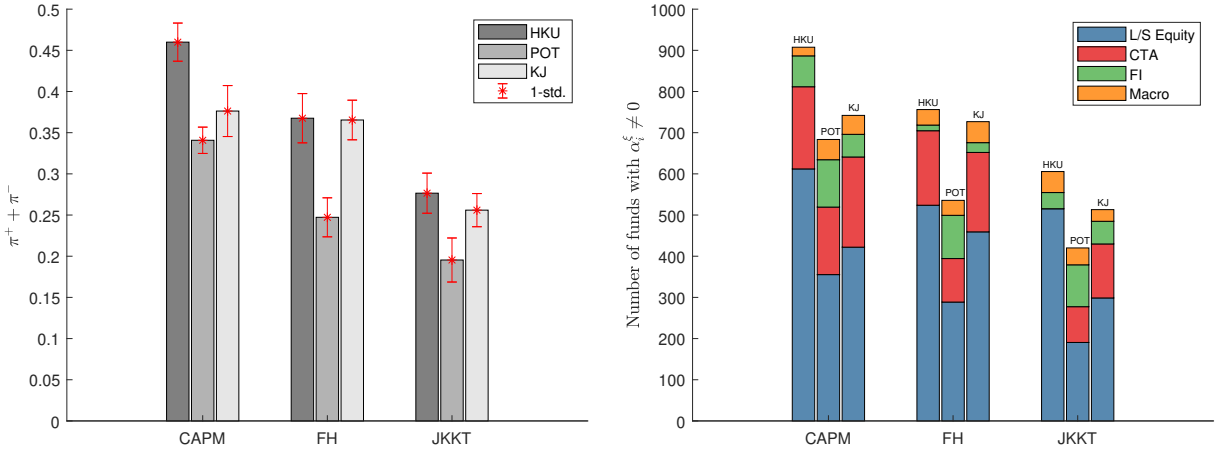


Figure 19: Estimated proportion of funds with $\alpha_i^* > 0$, obtained with the different asset pricing models. Red lines denote 2-standard errors intervals.



(a) (b)

Figure 20: (a) Proportions π^{+-} of funds with $\alpha_i^\xi \neq 0$, obtained from eq. (12) with either contemporaneous HKU (dark grey), POT (grey) or KJ (light grey) estimates of $\xi(\mathbf{x}_{it})$. (b) Number of funds, split between investment strategies, found to exhibit $\alpha_i^\xi \neq 0$. These quantities have been obtained by applying first the procedure of Barras et al. [2010] to obtain the proportion π^{+-} at the strategy level, then multiplying the result by the number of funds with a specific strategy.

3.6 Bootstrap confidence intervals for differences in average alpha

We describe in this section the bootstrap procedure used to compute pointwise confidence intervals for differences in average alphas between deciles of funds ranked on their tail risk exposures, accounting for the estimation uncertainty stemming from a first-stage estimation of the alphas.

Following Kosowski et al. [2006] and Kosowski et al. [2007], we use a fund-by-fund residuals-based bootstrap with fixed regressors. That is, starting from Eq. (12) in the main manuscript, we compute, for a given asset pricing model (e.g. CAPM, FH or JKKT) and for each fund $i = 1, \cdots, I$, the time series of the residuals:

$$\hat{\eta}_{it}^* = r_{it} - r_t^f - \hat{\alpha}_i^* - \boldsymbol{\beta}_i^* \tilde{f}_t.$$

We create $b = 1, \ldots, B$ pseudo-time series of resampled residuals $\{\hat{\eta}_{it}^{*b}\}_{T_i}$ of length T_i (i.e. the reporting duration of fund i) and construct similarly pseudo-time series of funds' excess returns:

$$r_{it}^{b} - r_t^f = \hat{\alpha}_i^* + \boldsymbol{\beta}_i^* \tilde{f}_t + \hat{\eta}_{it}^{*b}.$$

We use OLS to estimate the bootstrap alphas $\hat{\alpha}_i^{*b}$, for $b = 1, \ldots, B$ and each fund. Then, for each b and each decile D_d, $d = 1, \ldots, 10$, based on our full-sample tail risk estimates, we compute

$$\bar{\hat{\alpha}}_i^{*b} = \frac{1}{I/10} \sum_{i \in D_d} \hat{\alpha}_i^{*b},$$

and conversely obtain the bootstrap distribution of the differences in average alpha between bottom deciles and the top (last) decile. Final confidence intervals at a 95% confidence level are obtained with the percentile bootstrap method.

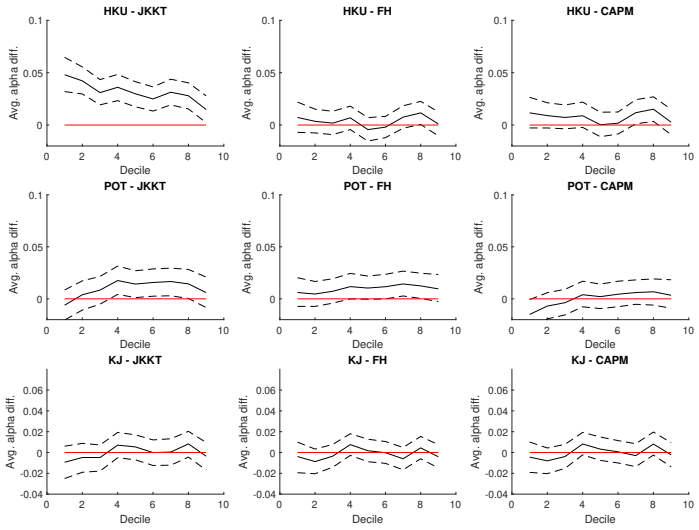


Figure 21: Solid: differences in average alpha computed between deciles 1 to 9, and the 10^{th} decile. Deciles are based on tail risk exposures. Dashed: 95% bootstrap confidence intervals with $b = 500$. Left to right: results obtained with JKKT, FH and CAPM alphas, respectively. Top to bottom: results for HKU, POT and KJ tail risk measures, respectively.

In Figure 21, we display the empirical differences between the bottom deciles and the top decile, and their associated pointwise bootstrap confidence intervals obtaiend with the method described above (we set $B = 500$). For HKU, zero is not included in the intervals for the difference between the first and 10^{th} deciles, for all asset pricing models. With the JKKT asset pricing model, we find also a significant difference between the 10^{th} decile and the second to the 8^{th} deciles. For the other measures, no clear pattern emerged.

4 A discussion on WML estimation

In Section 2.3 of the main paper, we introduced a set of conditions for the weighted maximum likelihood (WML) estimator to be consistent and asymptotically normal. These conditions being tedious to check with our hybrid regression model (and the asymptotic covariance matrix not being explicit), we discuss them here on a simpler (unconditional) example that exhibits a heavy tail. For illustration, we consider the case of a Pareto distribution (with survival distribution y^{-1/ξ_0} for all $y > 1$), as it is a typical example of heavy-tailed distributions. The (explicit) ML estimator (or WML estimator with $\tau = 0$) is known to be consistent, unbiased, and asymptotically normal with variance ξ_0^2.

If $\tau \neq 0$, the WML estimator $\hat{\xi}^w(\tau)$ becomes the solution of

$$\sum_{i=1}^{n} \psi_\xi(y_i) = 0, \quad \text{where } \psi_\xi(y) := -\frac{1}{\xi}\mathbb{1}_{\{y>q(\tau)\}} + \frac{1}{\xi^2}\log(y)\mathbb{1}_{\{y>q(\tau)\}} - \frac{1}{\xi^2}\log(q(\tau))\frac{q(\tau)^{-\frac{1}{\xi}}}{1-q(\tau)^{-\frac{1}{\xi}}}\mathbb{1}_{\{y\leq q(\tau)\}},$$

leading to the fixed-point relationship:

$$\hat{\xi}^w(\tau) = \frac{\frac{1}{n}\sum_{i=1}^{n}\log(y_i)\mathbb{1}_{\{y_i>q(\tau)\}} - \frac{\log(q(\tau))q(\tau)^{-1/\hat{\xi}^w(\tau)}}{1-q(\tau)^{-1/\hat{\xi}^w(\tau)}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{y_i\leq q(\tau)\}}}{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{y_i>q(\tau)\}}}.$$

Assuming, for convenience, $q(\tau)$ to be known (in practice, it would have to be estimated, making the theoretical study more involved) and equal to $(1-\tau)^{-\xi_0}$ under a correct specification, then straightforward calculations lead to

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n}\log(y_i)\mathbb{1}_{\{y_i>q(\tau)\}} &= \mathbb{E}\left[\log(y_i)\mathbb{1}_{\{y_i>q(\tau)\}}\right](1+o_{\mathbb{P}}(1)) = \xi_0(1-\tau)(1-\log(1-\tau))(1+o_{\mathbb{P}}(1)), \\ \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{y_i>q(\tau)\}} &= \mathbb{P}[y_i > q(\tau)](1+o_{\mathbb{P}}(1)) = (1-\tau)(1+o_{\mathbb{P}}(1)), \\ \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{y_i\leq q(\tau)\}} &= \mathbb{P}[y_i \leq q(\tau)](1+o_{\mathbb{P}}(1)) = \tau(1+o_{\mathbb{P}}(1)). \end{cases}$$

Notice that $|\mathbb{E}[\psi_\xi(y)]| > 0$ if $\xi \neq \xi_0$, and condition (i) for the consistency (see the end of Section 2.3) may be obtained, at least on a large interval of $\mathbb{R}_+$, by the law of large numbers. The consistency of $\hat{\xi}^w(\tau)$ may thus be deduced from the following equation:

$$\frac{\hat{\xi}^w(\tau)}{\xi_0} = \frac{(1-\tau)(1-\log(1-\tau)) + \frac{\log(1-\tau)(1-\tau)^{\xi_0/\hat{\xi}^w(\tau)}}{1-(1-\tau)^{\xi_0/\hat{\xi}^w(\tau)}}\tau}{1-\tau}(1+o_{\mathbb{P}}(1)).$$

Moreover, for $0 < \varepsilon < \xi_0$, the derivative $\frac{\partial}{\partial\xi}\psi_\xi(y_i)$ is clearly bounded on $\xi \in (\xi_0-\varepsilon, \xi_0+\varepsilon)$, then Condition (I) holds with

$$\overline{\psi}(y) = \frac{1}{(\xi_0-\varepsilon)^2}\mathbb{1}_{\{y>q(\tau)\}} + \frac{2}{(\xi_0-\varepsilon)^3}\log(y)\mathbb{1}_{\{y>q(\tau)\}} + \frac{2}{(\xi_0-\varepsilon)^3}\log(q(\tau)) + \frac{1}{(\xi_0-\varepsilon)^4}\frac{\log(q(\tau))^2 q(\tau)^{-1/(\xi_0+\varepsilon)}}{\left(1-q(\tau)^{-1/(\xi_0+\varepsilon)}\right)^2} - \frac{2}{(\xi_0-\varepsilon)^3}\frac{\log(1-\tau)}{\tau}.$$

Note also that $\mathbb{E}[\overline{\psi}(y)^2] < \infty$ since $\mathbb{E}[\log(y)^2 \mathbb{1}_{\{y>q(\tau)\}}] = \frac{\log(q(\tau))^2 + 2\xi_0 \log(q(\tau)) + 2\xi_0^2}{q(\tau)^{1/\xi_0}} < \infty$.
In addition, it is straightforward to check that Condition (II) is satisfied with

$$V_{\xi_0} = (\tau - 1) \frac{1 + \frac{\log(1-\tau)^2}{\tau}}{\xi_0^2}.$$

$\hat{\xi}^w(\tau)$ being computed as the solution of $\sum_{i=1}^{n} \psi_\xi(y_i) = 0$, Condition (III) is fulfilled, hence the asymptotic distribution

$$\sqrt{n}\left(\hat{\xi}^w(\tau) - \xi_0\right) \to \mathcal{N}\left(0, \frac{\xi_0^2}{1 - \tau + \frac{\log(1-\tau)^2(1-\tau)}{\tau}}\right), \text{ as } n \to \infty.$$

Note that the case $\tau = 0$ gives back the classical asymptotic distribution of the ML estimator.

To conclude, consistency and rate of convergence were obtained for the WML estimator of the Pareto parameter in this simplified setting. Considering our hybrid (and conditional) model G-E-GPD would make the theoretical study much more challenging, where we have to deal with many parameters and constraints. However, the asymptotic variance may be computed and estimated accurately (numerically).

References

D. Ardia, L. Barras, P. Gagliardini, and O. Scaillet. Is it Alpha or Beta? A Formal Evaluation of Hedge Fund Models. *Journal of Financial Economics*, Forthcoming, 2023.

L. Barras, O. Scaillet, and R. Wermers. False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The Journal of Finance*, 65(1):179–216, 2010.

A.L.M. Dekkers, J.H.J. Einmahl, and L. De Haan. A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, pages 1833–1855, 1989.

W. Fung and D.A. Hsieh. Hedge Fund Benchmarks: A Risk-Based Approach. *Financial Analysts Journal*, 60(5):65–80, September 2004.

B. Kelly and H. Jiang. Tail Risk and Asset Prices. *The Review of Financial Studies*, 27(10):2841–2871, 06 2014.

R. Kosowski, A. Timmermann, R. Wermers, and H. White. Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis. *The Journal of Finance*, 61(6):2551–2595, 2006.

R. Kosowski, N.Y. Naik, and M. Teo. Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. *Journal of Financial Economics*, 84(1):229–264, 2007.

A.J. Patton and T. Ramadorai. On the High-Frequency Dynamics of Hedge Fund Risk Exposures. *The Journal of Finance*, 68(2): 597–635, 2013.

H. Zou. The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.