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

J. Hambuckers[1,4,†], M. Kratz[2] and A. Usseglio-Carleve[3]

Abstract

Extreme value regression offers a convenient framework to assess the effect of market variables on hedge funds tail risks, proxied by the tail index of the cross-section of hedge funds returns. However, its major limitation lies in the need to select a threshold below which data are discarded, leading to significant estimation inefficiencies. In this paper, our main contribution consists in introducing a method to estimate simultaneously the tail index and the threshold parameter from the entire sample at hand, improving estimation efficiency. To do so, we extend the tail regression model to non-tail observations with an auxiliary splicing density, enabling the threshold to be internally determined without truncating the data. We then apply an artificial censoring mechanism to decrease specification issues at the estimation stage. Empirically, we investigate the determinants of hedge funds tail risks over time, and find a significant link with funding liquidity indicators. We also find that our tail risk measure has a significant predictive ability for the returns of around 25% of the funds. In addition, sorting funds along a tail risk sensitivity measure, we are able to discriminate between high- and low-alpha funds under some asset pricing models.

Keywords: Extreme value theory, generalized Pareto regression, censored maximum likelihood.

JEL: C13, C46, G23

[1] University of Liège - HEC Liège, Department of Finance, 4000 Liège, Belgium.

[2] ESSEC Business School, IDO department & CREAR, 95000 Cergy, France.

[3] Avignon Université, LMA UPR 2151, 84000 Avignon, France.

[4] Georg-August Universität Göttingen, Chair of Statistics, 37073 Göttingen, Germany

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

Acknowledgements: The authors thank the editor (B. Kelly), an associate editor, two anonymous referees, seminar participants of the Econometric Institute at the Erasmus University Rotterdam, HEC Lausanne Operation department, KU Leuven econometrics and statistics department, Humboldt-Universität zu Berlin statistics and econometrics groups, and EDHEC Financial Econometrics and Market Risks Workshop for insightful comments, as well as P. Hübner for assistance in preparing the data. JH acknowledges the financial support of the National Bank of Belgium and the FNRS.

1 Introduction

In finance, extreme value regression (EVR), pioneered by Davison and Smith (1990), finds its use in the analysis of factors affecting the likelihood of extreme losses of economic entities (Chavez-Demoulin et al., 2016; Massacci, 2017; Bee et al., 2019)[1]. In EVR models, one assumes that the tail distribution of a variable of interest is well approximated by a generalized Pareto distribution (GPD), a direct result of extreme value theory (Balkema and de Haan, 1974; Pickands, 1975). In addition, owing to heterogeneity in the data (e.g., related to time or contextual factors), the parameters of the GPD, and in particular its tail index, are assumed to be functions of covariates supposed to influence the distribution of extreme events. From a practical standpoint, the estimation of this model relies on the peaks-over-threshold (POT) approach: The analyst first selects high threshold values, possibly varying across time or covariates. Then, they discard all data smaller than these thresholds, and the parameters are estimated using solely the set of pooled observations larger than the thresholds, termed extreme values, with maximum likelihood procedures. The underlying idea is to select thresholds high enough to ensure that the GPD approximation is good, but also small enough to maintain a reasonable sample size. However, both the truncation procedure of the POT and the regression structure in the model severely impact our ability to obtain reasonably accurate estimates. Our main methodological contribution is therefore to propose a more efficient estimation method that also removes most of the subjectivity routinely applied to the choice of these thresholds.

Empirically, we consider the problem of estimating over time the tail index of a cross-section of hedge funds returns, and to relate its variations to time-varying financial conditions. Hedge funds are investment vehicles relying on sophisticated trading strategies and complex financial products to generate an economic profit. Estimating this tail in-

[1]See also Einmahl and He (2023) for an excellent review on applications of extreme value analysis of heterogeneous data in finance and economics.

dex and identifying financial conditions that influence its dynamics is important to assess extreme downside risks, anticipate threats to financial stability (Billio et al., 2012) but also to evaluate hedge funds' performance and take asset allocation decisions, since funds often pursue strategies akin to selling "earthquake insurance". i.e. deep out-of-the-money options (Stulz, 2007). However, although EVR models are natural candidates to conduct this estimation, this task is particularly complicated by the short reporting history and the limited cross-section of available hedge funds data. Indeed, considering that the POT approach, e.g., as suggested in Kelly and Jiang (2014), discards 95% of the data each month, this procedure leads to a significant loss of information. Moreover, since hedge funds databases are relatively recent and that preliminary filtering practices impose to discard early returns to guard against a backfill bias (Jorion and Schwarz, 2019), historical data rarely start before the end of the 90's, shortening the time series to 250 monthly observations at most. Figure 1 illustrates these data limitations: We display the number of funds' returns retained by the POT over time in our database, for four major investment strategies. The median number of monthly reporting entities is around 1,100 for the entire panel, but the effective median monthly sample size after POT thresholding is around 50 observations or less for specific strategies. In addition, because of pre-filtering for backfill bias (Jorion and Schwarz, 2019), our sample starts in 2002, which limits the length of our panel to 220 months. Thus, although the usual maximum likelihood estimator (MLE) of EVR models exploits both the cross-sectional and the time series dimensions by pooling all exceedances (Mhalla et al., 2022), the overall variances of the estimated tail indices and regression effects stay large while the selection of suitable thresholds remains a challenging step. In Section 3, we further illustrate this issue in an extensive and realistic simulation study, where we document a major loss of precision at the estimation stage even for fairly large sample sizes. There is therefore a strong need to find statistical alternatives that exploit the scarce hedge funds data more effectively in estimating extreme value regression

effects and tail indices. This is the goal pursued in this paper.

Notice that an alternative modeling strategy to obtain tail regression effects could be devised from the proposal of Kelly and Jiang (2014) for stock data: First, estimate T tail indices with the Hill (1975) estimator applied independently on exceedances from given months. Then, regress the obtained estimates on the p variables of interest. Aside from conceptually neglecting the underlying regression structure, this approach is hindered by the limited monthly cross-section of hedge funds data because the variance of the Hill (1975) estimator is solely a function of the number of monthly exceedances (denoted $n_t^{exc.}$). Therefore, each monthly tail index estimate will exhibit a large estimation uncertainty, and the two-step approach would inevitably suffer from a significant first-stage estimation uncertainty increasing in the length T of the time series, with no clear solution to propagate it for inference in the second stage. Differently, the EVR approach reduces this problem to estimating directly p regression coefficients (and indirectly T tail indices via the regression model) from $n^{exc.} = \sum_{t=1}^{T} n_t^{exc.}$ pooled exceedances. Thus, estimation uncertainty decreases with T and $n_t^{exc.}$, and EVR estimates are more efficient than the dynamic pooling approach of Kelly and Jiang (2014) as long as $p < T$, with both approaches being roughly equivalent if one includes time fixed-effects and no other covariates in the analysis ($p \approx T$). These considerations lead us to focus on improving the estimation of EVR models in the rest of the paper, since, as shown in Figure 1, both T and $n_t^{exc.}$, for $t = 1, \ldots, T$, are small[2].

So far, the literature on this problem is limited with regard to extreme value models with regression effects. To the best of our knowledge, only de Carvalho et al. (2022) and

[2]Moreover, notice that POT-based estimators work on a double asymptotic framework: The asymptotic properties are established under the assumption that the sample size (say n_t for the monthly Hill (1975) estimator) grows to infinity, while the ratio between the number of exceedances and the sample size ($n_t^{exc.}/n_t$) tends to zero (Mason, 1982). As such, even if one has access to more hedge funds data by combining several databases, one should normally select, with growing monthly cross-sections, a higher percentile threshold in the POT to avoid an increase in the bias, which slows down the variance reduction in practice. This issue is of course less severe for stocks, for which we have a stable cross-section of around 7,000 entities, such that month-by-month estimates may be obtained from a few hundred observations. However, the problem persists if one is interested in specific industries with a limited cross-sectional dimension.

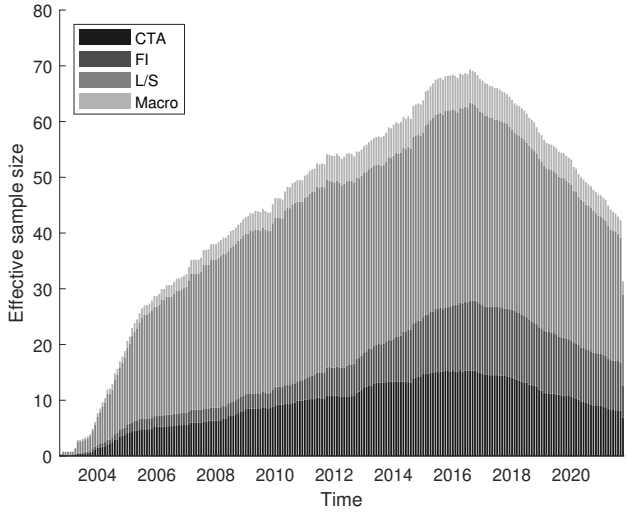


Figure 1: Monthly number of hedge funds' returns with *Long/Short Equity*, *CTA*, *Macro* or *Fixed Income* investment styles in the Eurekahedge database, after applying the POT of Kelly and Jiang (2014). As preliminary filtering steps, we applied the procedure of Jorion and Schwarz (2019) to limit backfill bias, and imposed a minimum reporting duration of 60 months at the fund level. For the POT, we consider a 95% monthly quantile as threshold.

Castro-Camilo et al. (2019) explicitly discuss automatic threshold estimation procedures for EVR models, although in a Bayesian context and not with a focus on improved inference. On the other hand, inference with automatized threshold estimation has been discussed extensively in the non-regression setting (see, e.g., Tancredi et al., 2006, Dacorogna et al., 2023 and references therein) and inspires the present work. Furthermore, the present work is complementary to the contributions of de Carvalho et al. (2022) and Naveau et al. (2016) on unconditional threshold selection with extended distributions, of Bader et al. (2018) and MacDonald et al. (2011) on flexible tail modeling, and extends the framework of Carreau and Bengio (2009) and Debbabi et al. (2017) with a regression structure.

Our modeling strategy is as follows. To avoid discarding too much data and improve on both the precision and the threshold selection problem in EVR, we formulate an auxiliary splicing distributional regression model (Kneib et al., 2021) that extends the EVR model below the supposed extreme value threshold. Splicing distributions are common tools in insurance to fit claim-size data (see Reynkens et al., 2017, for a review), combining a distribution for the body of the data (e.g. Gaussian) with a distribution for the upper

tail (e.g. GPD). To ensure that the resulting distribution is proper, constraints on the derivatives of the distributions are imposed at their junction points. Inspired by this setting, we develop a computationally tractable version of a splicing distributional regression model, in which the tail parameters are expressed as functions of candidate covariates. As a corollary of the definition of the parameters in our model, the junction point between the body and the right tail can be interpreted as a conditional threshold in an EVR model. We then use this formulation of the problem to estimate the regression parameters in the upper tail using the entire dataset, without resorting to a preliminary truncation of the data. Consequently, we achieve significant variance reduction in the estimates of the tail regression parameters. As a baseline model, we rely on a three-component distribution model advocated in Debbabi et al. (2017), combining Gaussian (G) and GPD distributions, bridged with an exponential (E) distribution. This model is referred to as G-E-GPD.

Although splicing distributions are attractive thanks to their flexibility, estimating their parameters is challenging in practice: Standard maximum likelihood estimation often fails or is biased because of the difficulty of choosing simultaneously correctly-specified parametric models for both the bulk and the upper tail of the data. This is particularly the case with hedge funds (Agarwal et al., 2017), for which the lower tail is badly approximated by a Gaussian distribution. Therefore, with the goal to allow for the estimation of a simple but potentially misspecified splicing regression model, we introduce a robust[3] estimation strategy based on a censored likelihood principle advocated by Diks et al. (2011). Our estimator belongs to the class of M-estimators, and lowers the impact of a misspecified distribution below the extreme value thresholds. In this estimation procedure, we replace some of the likelihood contributions of the data in the lower tail region by a conditional probability of belonging to the said region of the density. With this approach, we keep the "best of both worlds": On the one hand, we use the asymptotically justified GPD

[3]The term "robust" refers to an estimation procedure that suffers from limited bias when the data suffer from arbitrary contamination with respect to our distributional assumptions.

model for the upper tail, without the need to know its exact distribution to infer on the tail regression parameters. On the other hand, we exploit data belonging to the body of the distribution to reduce the variability of our estimates, without the need for our model to be correctly specified in the lower tail, where the Gaussian distribution may be a bad approximation. The result is a more efficient estimation of the regression parameters in the upper tail, compared to classical POT-EVR approaches. The use of censored likelihood methods and M-estimators has been previously advocated in the study of extreme values, although in unconditional and multivariate contexts, and not in a regression setting. See, e.g., Hüser and Davison (2014) and Einmahl et al. (2016).

We discuss the theoretical properties of our estimator, and provide a data-driven method to select the robustness parameter determining the degree of censoring. We illustrate the good behavior of this approach in practice with realistic simulation studies. In particular, we demonstrate that our splicing approach is much more efficient at estimating regression effects than classical POT-based methods, even under several types of misspecification of the G-E-GPD model.

We then exploit this approach to study the conditional tail distribution of *Long/Short Equity*, *CTA/Managed Futures*, *Macro* and *Fixed Income* hedge funds, using returns of around 2,000 funds over a period of 20 years. Motivated by considerations outlined in Agarwal et al. (2017) and Avramov et al. (2013), we address the questions as to whether the tail indices of hedge funds' cross-sectional distributions fluctuate with the state of the market, and whether this tail risk measure carries information about the return-generating process of the funds. We answer these two questions positively and contrast the results of the proposed approach with those obtained with classical POT-EVR models, as well as with the semi-parametric approach of Kelly and Jiang (2014). We find that the hedge funds tail index of the cross-sectional distribution is positively related to favorable funding conditions, a decrease in financial uncertainty and a booming stock market. A strategy-

specific analysis highlights however some marked differences for the *CTA/Managed Futures* and *Fixed Income* funds. Then, studying the predictive content of lagged values of our tail risk measure for hedge funds excess returns, we find it to be significant for around 25% of the funds. We also find that funds with excess returns that exhibit a low sensitivity to our tail risk measure display a significantly higher average alpha than high-sensitivity funds. These two results underline the potential of the tail risk measure to explain hedge fund abnormal returns or to carry out portfolio selection.

The rest of the paper is organized as follows. In Section 2, we introduce our model and estimation strategy. In Section 2.3, we derive several theoretical results associated with our model. In Section 3, we conduct an extensive simulation study. In Section 4, we conduct our empirical analysis and conclude in Section 5.

2 Methodology

Similarly to Kelly and Jiang (2014), we assume that the conditional distribution of the stationary loss process $Y_{it} \in \mathbb{R}$ at time $t \in \{1, \dots, T\}$ for a given entity i (e.g. a hedge fund) belongs to the maximum domain of attraction of a Fréchet distribution. That is, the upper tail of the conditional loss distribution above a threshold u_{it} is well approximated by a GPD:

$$\mathbb{P}\left(Y_{it} \leq y_{it} | Y_{it} > u_{it}\right) \xrightarrow[u_{it} \to y_{it}^F]{} 1 - \left(1 + \frac{\xi_{it}(y_{it} - u_{it})}{\sigma_{it}}\right)^{-1-1/\xi_{it}}, \tag{1}$$

where y_{it}^F denotes the upper end point of Y_{it} while ξ_{it} and σ_{it} are the conditional shape and scale parameter of the tail distribution, respectively. Both are assumed to be strictly positive (Pickands, 1975; Balkema and de Haan, 1974). However, differently from Kelly and Jiang (2014), we further assume that $\xi_{it} = \xi(\mathbf{x}_{it})$ and $\sigma_{it} = \sigma(\mathbf{x}_{it})$, with $\mathbf{x}_{it}$ being a d-dimensional vector of covariates observed over time. See also Coles (2001); Chavez-Demoulin and Davison (2005) and Embrechts et al. (1997) for theoretical details; and

Chavez-Demoulin et al. (2016); Hambuckers et al. (2018) and Mhalla et al. (2022) for examples of applications in the field of finance. Here, we specify ξ_{it} and σ_{it} as explicit functions of $\mathbf{x}_{it}$: as in classical generalized linear models (GLM), we connect $\mathbf{x}_{it}$ to the parameters of the GPD with a log-link function, specifying

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

where $\boldsymbol{\beta}^{\xi}$ and $\boldsymbol{\beta}^{\sigma}$ are the vectors of regression coefficients, including the constants. Our primary interest is in $\boldsymbol{\beta}^{\xi}$, which captures the marginal effects of changes in covariates on the tail risk, and in ξ_{it}, which is used as a tail risk measure[4]. Since $\boldsymbol{\beta}^{\xi}$ and $\boldsymbol{\beta}^{\sigma}$ are constant across time and entities, estimates may be obtained by pooling all observations across these two dimensions, then by applying the POT-EVR approach. Thus, contrary to Kelly and Jiang (2014), we do not need to estimate nonparametrically monthly ξ_{it} before regressing it on $\mathbf{x}_{it}$ in a two-step procedure to obtain an estimate of $\boldsymbol{\beta}^{\xi}$. Notoriously, the choice of u_{it} is known to be often arbitrary and the POT-EVR approach to be inefficient, since it discards a large portion of the data. In the following, we detail our procedure to improve on these issues.

2.1 Splicing auxiliary regression model

To bypass the need to select *ex ante* u_{it} and to discard a large portion of the data, we propose extending the EVR model below u_{it}, thus defining a flexible auxiliary regression model for the full range of the data. Hence, instead of relying solely on tail data to estimate $\boldsymbol{\beta}^{\xi}$, we also exploit the informational content of data in the body of the distribution. Of course, this extended modeling comes at a price: while the POT-EVR approach is immune to misspecification issues if u_{it} is large enough, it is not the case for the auxiliary regression

[4]More precisely, in line with Kelly and Jiang (2014) framework, we assume later on that $\xi_{it} = \xi_{jt} = \xi_t, \forall i, j$, and use this quantity as our tail risk measure. For the sake of generality, we keep the fund-specific notation in the description of the methodology.

model if the bulk of the data is not close enough to the true distribution. Nevertheless, contrary to tail events that are difficult to model, a reasonable assumption such as an exponentially-decreasing density function is known to provide a decent approximation of the true density in the body of the data (e.g. Gaussian, by the central limit theorem). In addition, in the next subsection, we detail an improved maximum likelihood estimation strategy of the auxiliary regression model that reduces estimation bias in case of misspecification in the lower tail of the auxiliary distribution, an issue encountered with hedge funds data when relying on the Gaussian distribution in our modelling strategy. Thus, our approach can be seen as a compromise between an asymptotically correct EVR model based on too few data, and an over-parametrized model for the full range of the data[5].

We therefore assume that the *full* conditional density of Y_{it} can be approximated by a *splicing regression model* of the G-E-GPD type, denoted by

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

and associated with a vector of d predictors denoted $\mathbf{x}_{it}$, for $i = 1, \cdots, I$ and $t = 1, \cdots, T$. We assume that the same set of predictors drives the different distribution parameters, but this assumption can be easily relaxed. $\xi(\mathbf{x}_{it})$ and $\sigma(\mathbf{x}_{it})$ are those of the GPD in (1). We acknowledge that financial data usually exhibit heteroscedasticity via the specification of the variance parameter $s(\mathbf{x}_{it})$ as a function of the covariates, while μ_0 is a location parameter assumed to be constant. This last hypothesis reflects that usually the conditional mean has been removed in a preliminary step, e.g., with an asset pricing model, as in Kelly and Jiang (2014). Dropping the explicit reference to $\mathbf{x}_{it}$, the pdf of the G-E-GPD model is then

[5]See also Beirlant et al. (2004) for an earlier discussion.

defined by:

$$f(y_{it};\boldsymbol{\theta},\mathbf{x}_{it}) = \begin{cases} \gamma_{1,it}\,\phi(y_{it};\mu_0,s_{it}^2) & \text{if} \quad y_{it} \le u_{it}^* \\ \gamma_{2,it}\,e(y_{it};\lambda_{it}) & \text{if} \quad u_{it}^* \le y_{it} \le u_{it} \\ \gamma_{3,it}\,h(y_{it}-u_{it};\xi_{it},\sigma_{it}) & \text{if} \quad u_{it} \le y_{it} \end{cases} \tag{2}$$

where $\gamma_{j,it}$, $j = 1, 2, 3$, are the respective weights of the components, $\phi(\cdot;\mu,s^2)$ refers to the Gaussian pdf with mean μ and variance s^2, $e(\cdot;\lambda)$ to the exponential pdf with parameter $\lambda > 0$ defined by $e(y;\lambda) = \lambda\, e^{-\lambda y}$, and $h(z;\cdot)$ denotes the pdf of the GPD given by

$$h(z;\xi,\sigma) = \begin{cases} (1+\xi\, z/\sigma)^{-1-1/\xi}/\sigma & \text{if} \quad \xi \neq 0 \\ e^{-z/\sigma}/\sigma & \text{if} \quad \xi = 0 \end{cases}$$

where $\sigma > 0$, $z \in \mathbb{R}_+$ if $\xi \geq 0$ and $z \in [0, -\sigma/\xi]$ if $\xi < 0$. Hence, in this model, an observation y_{it} will be considered extreme if $y_{it} \geq u_{it}$, and its exceedance above u_{it} assumed to follow a GPD with tail parameter ξ_{it}. The parameter u_{it} plays the role of the conditional threshold used in the POT approach. Second, we assume that, for observations $y_{it} \leq u_{it}^*$, the true distribution of the data is close to a Gaussian distribution. Although it is a debatable assumptions for observations far in the left tail (we address extensively this point in the next subsection), the normal distribution provides virtually indistinguishable fit compared to most unimodal heavy-tailed distributions between the 20% and 80% quantiles, this latter quantile roughly corresponding to our estimates of u_{it}^* in Section 4. As further evidence of the correctness of this assumption, in Section 3 of the Supplement, we report goodness-of-fit tests of the Gaussian distribution for our data, truncated at these levels. We do not reject the normality assumption for all months but two. Thus, we may safely use the Gaussian distribution for data in this range, and benefit from its analytical and computational tractability compared to alternative distributions such as mixed Erlang (Reynkens et al., 2017) or Student. Finally, the truncated exponential distribution for observations in the range $u_{it}^* \leq y_{it} \leq u_{it}^*$ allows for a smooth transition between the two main compo-

nents of the density, addressing some computational limitations documented in Carreau and Bengio (2009) and encountered when building a splicing distribution solely on a GPD and a Gaussian distribution (Dacorogna et al., 2023): While the G-GPD formulation would enforce an extremely pronounced and unrealistic asymmetry of the density, the introduction of an exponential bridge provides an additional degree of freedom (via u_{it}^*). As such, it also allows for a reasonable level of asymmetry, more in line with documented empirical characteristics of hedge funds returns (Elyasiani and Mansur, 2017). This assumption is also akin to assume that the Gaussian distribution extends over the range $[u_{it}, u_{it}^*[$, however with a dispersion parameter different from the one of the lower observations. Similarly to the tests conducted for the Gaussian assumption, we perform ad-hoc goodness-of-fit tests of the exponential distribution for our data between the quantiles at levels 80% and 95%. The exponential distribution could not be rejected as well (see Section 3 in the Supplement). From (2), we can therefore devise a likelihood-based estimation procedure for ξ_{it} that uses all the data at hand, thus expected to reduce greatly the variance of our estimates compared to the POT. Finally, notice also that the proposed structure circumvents some of the conceptual issues associated with the absence of threshold stability property in EVR models (Eastoe and Tawn, 2009): By estimating directly $\xi(\mathbf{x}_{it})$ and $\sigma(\mathbf{x}_{it})$ without selecting *ex ante* a threshold, it becomes irrelevant to investigate the effect of a change in a preliminary threshold choice on the estimated regression model.

Although the density (2) is conceptually straightforward since we simply combine three well-known densities, the weights $\gamma_{j,it}$ and the truncation points (u_{it}, u_{it}^*) complexify the expressions. To reduce the dimensionality of the problem, we therefore assume that f is continuous and differentiable ($\mathcal{C}^1$) at the two junction points between the components, as advocated in Debbabi et al. (2017) and Dacorogna et al. (2023). Moreover, since f is a pdf, we have that $\int_{\mathcal{R}} f(y; \boldsymbol{\theta}, \mathbf{x}) dy = 1$. From these equalities, we reduce the vector of free

distribution parameters to $[\mu_0\,, s_{it}\,, \sigma_{it},\, \xi_{it}]$, the other parameters satisfying

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

See Debbabi et al. (2017) for details. In addition, to enforce $u^*_{it} \leq u_{it}$, $\forall i, t$, we use Theorem 3.4.5 in Embrechts et al. (1997), pg. 160 (Eq. 3.46), and define the threshold u_{it} as

$$u_{it} = u^*_{it} + \sigma_{it}/\xi_{it}. \tag{3}$$

Indeed, the model is formally identified for $u_{it} = c\sigma_{it}/\xi_{it} \geq u^*_{it}$ with c being any positive constant (see Embrechts et al., 1997). Eq. (3) is therefore similar to setting c as $u^*_{it}\, \xi_{it}/\sigma_{it}+1$. Notice that, for $\sigma_{it} \approx 0$, $u^*_{it} \approx u_{it} \to +\infty$, and the model is equivalent to the Gaussian location-scale regression model. This limiting model is ruled out since we enforce $\sigma_{it} > 0$.

From (3), we observe that as soon as we allow at least ξ_{it} to be driven by covariates, we also implicitly define a model in which the threshold parameter u_{it} depends on the covariates, a feature in line with the hypotheses of EVR models. As a corollary, estimating model (2) automatically sets the threshold u_{it}, lifting the need to estimate this conditional threshold *ex ante* and in isolation. To ensure sufficient flexibility and account for heteroscedasticity, we allow s_{it} to be a function of the covariates as well, specifying

$$\log(s_{it}) = \mathbf{x}^T_{it}\boldsymbol{\beta}^s, \tag{4}$$

and $\boldsymbol{\theta} = \left(\mu_0, \boldsymbol{\beta}^s, \boldsymbol{\beta}^\sigma, \boldsymbol{\beta}^\xi\right) \in \Theta \subset \mathbb{R}^{3d+4}$ is the vector of all parameters to be estimated. Notice that we can easily assume that different sets of covariates enter the equations of each distribution parameter, or that some of them are left constant.

2.2 Estimation under misspecification

A recurrent challenge with splicing models involves defining parametric models valid for the complete support of the data, and not only for the upper tail. In particular, if the Gaussian

distribution chosen to approximate the lower tail of the data is too far from the true data-generating process, we quickly encounter numerical issues with likelihood-based estimation procedures. Hedge funds data are particularly exposed to this issue, due to the presence of skewness and tail asymmetry, making the Gaussian distribution an inadequate distribution for the left tail. On the contrary, this is a problem to which POT-EVR is immune, since it discards the data in the lower tail. Thus, to retain simultaneously the computational ease of using our G-E-GPD in a likelihood-based estimation procedure, and the robustness of EVR to specification problems, we propose using a robust weighted maximum likelihood estimator (WMLE) instead of the classical MLE to estimate $\boldsymbol{\theta}$, as suggested by, e.g., Wang and Zidek (2005). The general idea of this estimator involves controlling for potential specification problems by lowering the importance of some observations in the likelihood function. To do so, we use a weighting scheme based on an *artificial censoring procedure*, reducing our problem to the maximization of a *censored likelihood function*, a proper scoring function introduced in Cuesta-Albertos et al. (2008) and Diks et al. (2011). First, let us start by defining the maximum likelihood estimator:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{I} \sum_{t=t_{i,1}}^{t_{i,n_i}} \log(f(y_{it};\boldsymbol{\theta},\mathbf{x}_{it})), \tag{5}$$

where y_{it} denotes the observation at time t for entity i reporting over n_i periods, with $t_{i,1}$ the first period for which entity i reports in the database, and t_{i,n_i} the last period. Remember that since our application of interest relates to hedge funds, hardly any entities (i.e. funds) report over the complete period under study. We are therefore dealing with an unbalanced panel. Our WMLE is then given by:

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{w}(\tau) = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{I} \sum_{t=t_{i,1}}^{t_{i,n_i}} & \mathbb{1}(y_{it} \geq q(\tau)) \log(f(y_{it};\boldsymbol{\theta},\mathbf{x}_{it})) \\ & + \mathbb{1}(y_{it} < q(\tau)) \log(F(q(\tau);\boldsymbol{\theta},\mathbf{x}_{it})),\end{aligned} \tag{6}$$

where $q(\tau)$ is a censoring threshold and $F(q(\tau);\boldsymbol{\theta},\mathbf{x}_{it})$ is the cumulative distribution function of y_{it} at point $q(\tau)$. The idea behind this artificial censoring mechanism is to give less weight to observations $y_{it} < q(\tau)$ in the left tail (and potentially also the left part of the body of the distribution), where our distributional assumptions are probably wrong and of less importance. On the contrary, in the right tail, since our model relies on EVT, we have an asymptotically justified model holding under fairly general conditions and which should not be problematic in the estimation procedure. The center of the distribution and it right neighborhood are also assumed to be well approximated by the combination of the Gaussian and the exponential distributions. The principle of estimator (6) is also known as *minimum scoring rule inference* (Dawid et al., 2016) and guarantees unbiased estimation under limited assumptions. The theoretical properties of (6) are discussed in Section 2.3. For the censoring threshold $q(\tau)$, we propose using either an unconditional threshold at level τ, or observation-specific thresholds, such as conditional quantiles. In the latter case, our estimator of $\boldsymbol{\theta}$ becomes a conditional weighted maximum likelihood estimator (CWMLE) and is given by:

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{cw}(\tau) = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{I}\sum_{t=t_{i,1}}^{t_{i,n_i}} \mathbb{1}(y_{it} \geq q_{it}(\tau)) \log(f(y_{it};\boldsymbol{\theta},\mathbf{x}_{it})) \\ + \mathbb{1}(y_{it} < q_{it}(\tau)) \log(F(q_{it}(\tau);\boldsymbol{\theta},\mathbf{x}_{it})),\end{aligned} \tag{7}$$

where $q_{it}(\tau)$ denotes the *conditional* quantile of y_{it} at level τ, estimated, e.g., via quantile regression (Koenker and Bassett, 1978) and defined as $\mathbb{P}(y_{it} \leq q_{it}(\tau)|\mathbf{X} = \mathbf{x}_{it}) = \tau$. To choose τ (termed the robustness parameters), we describe a data-driven approach in Section 2.4. Final estimation is conducted with numerical optimization routines. Starting values are described in the Supplement.

Remark that, to circumvent the issue of estimating $q(\tau)$ or $q_{it}(\tau)$ in a preliminary step, one would be tempted to set directly these quantities equal to u_{it} in eq. (6) or (7), and to perform a direct maximization. However, this is not a computationally feasible strategy: In

this setting, the value of the log-likelihood function for $u_{it} \to +\infty$ may be superior to the ones obtained for any finite u_{it}. Thus, maximizing (6) or (7) would lead all observations to be censored, and the algorithm may not converge due to some non-identified parameters.

2.3 Theoretical properties and inference

In order to make inference on the parameters vector $\boldsymbol{\theta}$, and propose confidence intervals for each of its components, we focus in this section on the asymptotic properties of the estimators introduced in the previous section. It is worth noticing that most of these properties may be derived from the existing literature. Indeed, let us recall that our WMLE is defined as:

$$\hat{\boldsymbol{\theta}}^{w}(\tau) = \arg\max_{\boldsymbol{\theta}\in\Theta} M_n(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}\in\Theta} \frac{1}{n}\sum_{i=1}^{n} m_{\boldsymbol{\theta}}(y_i, \mathbf{x}_i),$$

where $\boldsymbol{\theta} = \left(\mu_0, \boldsymbol{\beta}^s, \boldsymbol{\beta}^\sigma, \boldsymbol{\beta}^\xi\right) \in \Theta \subset \mathbb{R}^{3d+4}$, $n = \sum_{i=1}^{I} n_i$ and

$$m_{\boldsymbol{\theta}}(y, \mathbf{x}) = \mathbb{1}(y \geq q(\tau)) \log(f(y; \boldsymbol{\theta}, \mathbf{x})) + \mathbb{1}(y < q(\tau)) \log(F(q(\tau); \boldsymbol{\theta}, \mathbf{x})).$$

By considering for convenience $q(\tau)$ as known, this estimator is an M-estimator, as defined in Van der Vaart (2000). The asymptotic normality of this estimator may thus be derived from the literature on M-estimation, as follows. Introducing the vector-valued function $\psi_{\boldsymbol{\theta}}(.,.) = \frac{\partial}{\partial \boldsymbol{\theta}} m_{\boldsymbol{\theta}}(.,.)$, then, under the following assumptions:

(I) $\psi_{.}(y, \mathbf{x})$ fulfills a locally Lipschitz-type condition, i.e. for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in the neighborhood of $\boldsymbol{\theta}_0$, $||\psi_{\boldsymbol{\theta}_1}(y, \mathbf{x}) - \psi_{\boldsymbol{\theta}_2}(y, \mathbf{x})|| \leq \overline{\psi}(y, \mathbf{x})\, ||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||$, where $\overline{\psi}$ is a measurable function such that $\mathbb{E}[\overline{\psi}(y, \mathbf{x})^2] < \infty$,

(II) $\mathbb{E}\left[||\psi_{\boldsymbol{\theta}_0}(y, \mathbf{x})||^2\right] < \infty$ and the map $\boldsymbol{\theta} \to \mathbb{E}\left[||\psi_{\boldsymbol{\theta}_0}(y, \mathbf{x})||\right]$ is differentiable at $\boldsymbol{\theta}_0$ with nonsingular derivative matrix $\boldsymbol{V}_{\boldsymbol{\theta}_0}$,

(III) $\hat{\boldsymbol{\theta}}^{w}(\tau) \xrightarrow{P} \boldsymbol{\theta}_0$ and $\frac{1}{n}\sum_{i=1}^{n} \psi_{\hat{\boldsymbol{\theta}}^{w}(\tau)}(y_i, \mathbf{x}_i) = o_{\mathbb{P}}\left(n^{-1/2}\right)$,

we have, as $n \to \infty$,

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{w}(\tau) - \boldsymbol{\theta}_0\right) \to \mathcal{N}\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\theta}_0}^{-1}\mathbb{E}\left[\psi_{\boldsymbol{\theta}_0}(y,\mathbf{x})\psi_{\boldsymbol{\theta}_0}(y,\mathbf{x})^\top\right]\left(\boldsymbol{V}_{\boldsymbol{\theta}_0}^{-1}\right)^\top\right). \tag{8}$$

This asymptotic normality may thus be used to construct confidence regions for our estimators, under the assumption that Conditions (I), (II), and (III) are fulfilled. These conditions are difficult to check theoretically for our model, which is why we turn to their empirical counterparts. Nevertheless, in Section 4 of the Supplement, we illustrate this study in a simplified setting, considering a Pareto distribution as a representative example of the maximum domain of attraction of a Fréchet distribution.

Remark 1

(i) The term $\mathbb{E}\left[\psi_{\boldsymbol{\theta}_0}(y,\mathbf{x})\psi_{\boldsymbol{\theta}_0}(y,\mathbf{x})^\top\right]$ *in the asymptotic variance above may be estimated by its empirical counterpart* $\dfrac{1}{n}\displaystyle\sum_{i=1}^{I}\sum_{t=t_{i,1}}^{t_{i,n_i}} \psi_{\hat{\boldsymbol{\theta}}^{w}(\tau)}(y_{it},\mathbf{x}_{it})\psi_{\hat{\boldsymbol{\theta}}^{w}(\tau)}(y_{it},\mathbf{x}_{it})^\top$, *where* ψ *is obtained numerically. The derivative matrix* $\boldsymbol{V}_{\boldsymbol{\theta}_0}$ *can also be approximated numerically.*

(ii) Condition (III) requires the consistency of the estimator $\hat{\boldsymbol{\theta}}^{w}(\tau)$ *of* $\boldsymbol{\theta}_0$*. Conditions can be introduced to ensure this consistency, namely:*

i) $\sup\limits_{\boldsymbol{\theta}\in\Theta}\left|\left|\frac{1}{n}\sum\limits_{i=1}^{n}\psi_{\boldsymbol{\theta}}(y_i,\mathbf{x}_i) - \mathbb{E}\left[\psi_{\boldsymbol{\theta}}(y,\mathbf{x})\right]\right|\right| \xrightarrow{P} 0$, *as* $n \to \infty$*;*

ii) For all $\varepsilon > 0$, $\inf\limits_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0||\geq\varepsilon} ||\mathbb{E}\left[\psi_{\boldsymbol{\theta}}(y,\mathbf{x})\right]|| > 0 = ||\mathbb{E}\left[\psi_{\boldsymbol{\theta}_0}(y,\mathbf{x})\right]||$.

These conditions may also be replaced by similar conditions on the empirical criterion function $M_n(\boldsymbol{\theta})$*:*

$$\sup_{\boldsymbol{\theta}\in\Theta}|M_n(\boldsymbol{\theta}) - \mathbb{E}[m_{\boldsymbol{\theta}}(y,\mathbf{x})]| \xrightarrow{P} 0 \quad \textit{and} \quad \sup_{\boldsymbol{\theta}:||\boldsymbol{\theta}-\boldsymbol{\theta}_0||\geq\varepsilon}\mathbb{E}[m_{\boldsymbol{\theta}}(y,\mathbf{x})] < \mathbb{E}[m_{\boldsymbol{\theta}_0}(y,\mathbf{x})].$$

Notice that, in practice, $q(\tau)$ is unknown and we plug in an empirical unconditional quantile of $(y_{it})_{i,t}$ instead. The impact of this random (instead of known) threshold on the asymptotic properties should be theoretically taken into account. However, it is expected

to remain limited for several reasons: first, τ is fixed and its optimal value was found relatively central in practice (around 0.2), such that estimation uncertainty is small. Second, when using an empirical unconditional quantile as estimator of $q(\tau)$, we use all the available data, while for the parameters of the model we only use $100 \times (1-\tau)\%$ of them. Therefore, we can expect a faster convergence to the true $q(\tau)$, and a limited impact on $\hat{\boldsymbol{\theta}}^{w}(\tau)$. As such, one can rely on the above results to construct approximate confidence intervals. Empirical coverage rates of these intervals presented in the simulation study of Section 3 confirm the correctness of this approximation for the considered data-generating processes.

2.4 Data-driven choice of τ

Our estimators $\hat{\boldsymbol{\theta}}^{w}(\tau)$ and $\hat{\boldsymbol{\theta}}^{cw}(\tau)$ depend on the robustness parameter τ. This quantity determines the intensity of the censoring mechanism: since $q(\tau)$ and $q_{it}(\tau)$ are defined as quantiles and conditional quantiles, respectively, a large τ implies that a large fraction of the observations is censored in the likelihood function. If a large fraction of the lower data exhibit important deviations from the G-E-GPD model, a large τ may be thus required to overcome misspecification issues. However, there is simultaneously a need to set τ small if the G-E-GPD model is correctly specified, to avoid losing valuable information from relevant data points.

With the ambition to respect the features of the data, we introduce in this section a data-driven selection method, relying on the modified Anderson-Darling (AD^m) statistic. This quantity assesses the goodness-of-fit of a parametric distribution and gives a special weight to extreme values in the upper tail of a distribution (Babu and Toreti, 2016). The use of this statistic continues the approach advocated by Jun-Haeng Heo et al. (2013) and Bader et al. (2018) to select an appropriate threshold in classical EVT.

Denoting by $\boldsymbol{\theta}^{w}(\tau)$ or $\boldsymbol{\theta}^{cw}(\tau)$ the (C)WMLE of the splicing model obtained with a specific

τ, we compute the corresponding pseudo-residuals obtained from the probability integral transform (PIT), namely $\hat{\epsilon}_{it}(\tau) = \Phi^{-1}(F(y_{it}; \hat{\boldsymbol{\theta}}^{w}(\tau), \mathbf{x}_{it}))$, with Φ^{-1} being the quantile function of the standardized normal. Under a correct specification and a good estimation of the splicing model, we expect these residuals to be approximately normally distributed, in particular in the upper tail. The AD^m statistic is then given by

$$AD^m(\tau) = n \int_{-\infty}^{+\infty} \frac{\Big(\Phi(x) - \tilde{F}_n(x, \tau)\Big)^2}{1 - \Phi(x)} d\Phi(x), \tag{9}$$

where $\tilde{F}_n(\cdot, \tau)$ denotes the empirical distribution function of the re-indexed pseudo-residuals $\hat{\epsilon}_k(\tau)$, for $k = 1, \ldots, n$ and $n = \sum_{i=1}^{I} n_i$. Eq. (9) measures departures from the expected Gaussian distribution. The sample version of Eq. (9) is given in Jun-Haeng Heo et al. (2013) (equation (9)). Computing $AD^m(\tau)$ for various values of τ, we select τ^{opt}, such that

$$\tau^{opt} = \arg\min_{\tau} AD^m(\tau).$$

From a practical standpoint, both the simulation study and the empirical analysis presented in the next sections suggest that choosing $\tau \in [0.10, 0.30]$ improves numerical stability and delivers improved estimations for the considered data.

3 Simulation evidence

In this section, we assess the finite sample properties of the proposed methodology under realistic data-generating processes (DGP), as a "proof-of-concept" of the proposed method[6]. We compare the performance of the POT-based method with the G-E-GPD model estimated either with the MLE or our censoring procedure. This extensive comparison allows us to distinguish the advantages brought solely by the auxiliary splicing regression formulation (aiming at estimating automatically the threshold parameter, and reducing the variance of the estimates), from the ones obtained with the addition of the censoring pro-

[6]The method was implemented in MATLAB R2021b. Codes are available upon request to the corresponding author.

cedure (targeting a reduction of a potential bias caused by the misspecification of the likelihood function).

We consider three settings: (i) correct specification, (ii) contaminated lower values, and (iii) misspecified model. In the first setting, the data are exactly distributed as the splicing model, whereas in the second and third settings the splicing model suffers from misspecifications in the lower tail of the distribution. In addition, in the third setting, the splicing model is only asymptotically correct in the upper tail. The selection of τ is made over a grid of 20 values equally spaced between 0.05 and 0.5. We compare the quality of the estimated parameters when the weight function is taken either as the empirical quantile of the data at level τ, or as estimated conditional quantiles obtained from parametric quantile regression at level τ. As performance measure, we report the root mean squared error (RMSE) of the estimated parameters $\hat{\boldsymbol{\beta}}^{\xi}$, given as (we denote generically the j-th element of the vector $\boldsymbol{\beta}^{\xi}$ by β_j^{ξ}, $j = 0, \ldots, d$, with β_0^{ξ} referring to the constant term):

$$RMSE_j = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\frac{\hat{\beta}_j^{\xi,(b)} - \beta_j^{\xi}}{|\beta_j^{\xi}|}\right)^2}, \tag{10}$$

where $\hat{\beta}_j^{\xi,(b)}$ denotes an estimate computed on sample b, for $b = 1, \ldots, B$ with $B = 200$. As a second criterion, we compute the empirical coverage probability and power of the 95% confidence intervals for β_j^{ξ}, $j = 0, \ldots, d$, obtained from Eq. (8). We also report the average length of these confidence intervals in the Supplement. We compare the results obtained with the proposed WMLE with those of the MLE, and classical POT-EVR methods with different threshold selection methods: a conditional quantile regression approach (using a quadratic polynomial in the covariates), and an unconditional quantile approach. We use quantiles at levels 90%, 95%, and 99%.

To mimic the fact that we work with panel data in our application, we simulate realizations of the DGPs, at each time point $t = 1, \ldots, T$ with $T = 250$, as if we had $I = 40$ reporting entities over time. As such, the total sample size $n = TI$ is equal to $10,000$.

Although appearing large, this sample size is much smaller than the total sample size in our empirical application, in which we use around 189,000 month-fund observations[7]. It also gives us the opportunity to test the POT approaches in a setting with a sufficiently large number of extreme observations.

3.1 Data-generating processes

In this section, we describe the three DGPs used to simulate the data. Throughout all DGPs, we assume that up to $d = 5$ covariates may drive each distribution parameters (ξ_{it}, σ_{it} and s_{it}), reflecting the fact that financial data display heteroscedasticity and tail heterogeneity. Four covariates are simulated from AR(1) processes with location-scale t-distributed error terms (denoted by $\mathcal{T}(\cdot)$), to mimic the fact that our explanatory variables (e.g. the VIX) are collected over time, exhibit autocorrelation, and display heavy tails:

$$x_{jt} = .2 + .5x_{jt-1} + \epsilon_{jt}, \text{ with } \epsilon_{jt} \sim \mathcal{T}(0, 0.1, 8), \ j = 1, \ldots, 4.$$

These covariates are assumed to be common to all entities at a given point in time (thus we drop the index i). In addition, we assume one time-specific and entity-specific covariate, $x_{i5t} \sim N(0, .1)$. This covariate reflects the possibility to include time-varying and fund-specific information in the set of covariates, although this avenue is not pursued in our empirical analysis in light of the tail risk definition formulated by Kelly and Jiang (2014).

We choose $\beta^{\xi} = [\log(0.2),\ 1,\ -1.3,\ 1.3,\ 0,\ -.8]$[8], and set $\mu_0 = 0$. Intercept values were chosen in accordance with estimates obtained in our empirical analysis. For β^s and β^{σ}, we detail the chosen values in Section 1 of the Supplement.

Correct specification (DGP I) First, assume that the response variable y_{it} is simulated from the splicing regression model (2), making the model perfectly specified. We thus

[7] In the Supplement, we report the obtained results with $T = 1,000$ and a final sample size of $40,000$ for DGP I to III. Results are qualitatively similar.

[8] In terms of effect sizes, these coefficients imply that a one-standard deviation change in the corresponding covariate (which stands between .1 and .13) increases or decreases ξ_{it} between 8% and 13%.

expect the censoring threshold τ to be close to 0. For this DGP, we expect the MLE of the splicing regression model to be the best estimator. Furthermore, we expect the POT estimator to be much less efficient, since it uses only a fraction of the data.

Contaminated lower values (DGP II) To consider a more realistic situation, we introduce in DGP II a misspecification in the lower values of the splicing distribution. To do so, we start by generating the data following the same splicing model as in DGP I. Then, we apply the following transformations to the simulated data:

- We identify the 30% smallest observations y_{it} in the sample, and we denote by t^* the empirical quantile at level 30%.
- These observations are then replaced by $y_{it}^* = t^* - (H^{-1}(z_{it}^*, \sigma^*, \xi^*))$, with z_{it}^* being a random number simulated from a uniform distribution, H^{-1} being the inverse of the GPD, $\sigma^* = 0.08$ and $\xi^* = .1$. The data are therefore "doubly GP" distributed and heavily asymmetric.

Thus, the MLE is misspecified (although only a fraction of the data are contaminated), and we should observe an increase in its RMSE. On the contrary, the POT estimator remains correctly specified for thresholds large enough, and we do not expect to observe major differences with respect to DGP I. In the Supplement, we provide additional results for different proportions of contaminated data.

Contaminated t-locations scale distribution (DGP III) Finally, we turn to the most realistic case in which the underlying model is only asymptotically valid in the right tail and misspecified in the left tail. We consider a t-location scale distribution, characterized by 3 parameters: a mean parameter μ_0, a scale parameter $s(\mathbf{x}_{it})$, and a shape parameter $\nu(\mathbf{x}_{it})$. As for the classical t-distribution, the equivalent shape parameter of the limiting GP regression model is simply $\xi(\mathbf{x}_{it}) = \dfrac{1}{\nu(\mathbf{x}_{it})}$ (McNeil et al., 2015). Consequently,

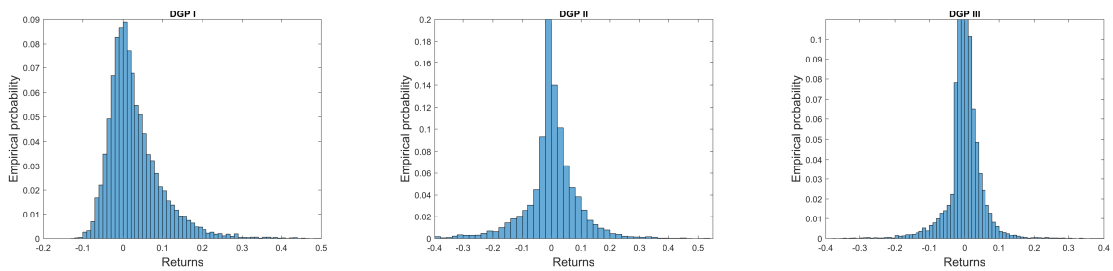
the true limiting GP regression model has a shape parameter given by

$$\xi(\mathbf{x}_{it}) = \exp(-\mathbf{x}_{it}^T \tilde{\boldsymbol{\beta}}^{\xi}) = \exp(\mathbf{x}_{it}^T \boldsymbol{\beta}^{\xi}), \tag{11}$$

with $\tilde{\boldsymbol{\beta}}^{\xi} = -\boldsymbol{\beta}^{\xi}$. We use these values of the regression parameters to simulate the data, such that the GP regression model has a dynamic of the shape parameter similar to DGP I and DGP II. Here, the MLE, WMLE and the POT estimators will suffer from an additional bias compared to DGP I and DGP II. Finally, we introduce an additional misspecification in the left tail, similar to DGP II, by resampling the smallest 30% of the observations from a GPD. This procedure favores the POT estimator. Further simulation evidence may also be found in the Supplement.

3.2 Description of the simulated data

We start by briefly describing the features of the simulated data. In Figure 2, we display the histogram of the pooled data for each DGP. For DGP I, the data exhibit a heavy right tail and positive skewness. In this setup, the true conditional threshold, above which our random variable follows the extreme value regression model, is given by Eq. (3), and is a nonlinear function of $\mathbf{x}_{it}$. On average, around 1% of the simulated data are larger than their corresponding u_{it}. For DGP II, we introduce a misspecification in the left tail, which is now heavier compared to DGP I. Similar fractions of observations are larger than u_{it}, since the models are identical in the right tail. Finally, for DGP III, both tails are heavier than the normal as well, and the right tail is not exactly GP-distributed.

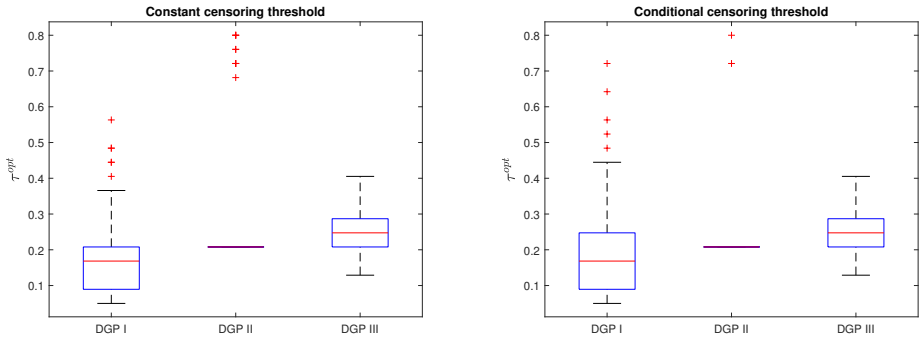


(i) DGP I (ii) DGP II (iii) DGP III

Figure 2: Simulated data from the different DGPs.

3.3 Results

We report first information on the selected τ^{opt} (Figure 3) under the different DGPs. For DGP I (correct specification), the selected τ^{opt} values vary mostly between 0.05 and 0.2. Inspecting the sequences of the obtained AD^m statistics for various τ, we only observe small differences across AD^m. This result is in line with the fact that, under a correct specification of the model, our censored estimate of $\boldsymbol{\beta}^{\xi}$ is unbiased for any $0 \leq \tau < 1$. For DGP II, we observe values of τ^{opt} systematically around 0.25, while for DGP III, we consistently select τ^{opt} between 0.2 and 0.3. Hence, the selection procedure succeeds in identifying the existence of a contamination in the left tail[9].



(i) Constant censoring threshold (ii) Conditional censoring threshold

Figure 3: Selected censoring threshold τ^{opt}, over 200 samples, with the censoring threshold being either (i) the empirical quantile ($q(\tau)$) or (ii) the estimated conditional quantile at level τ ($q_{it}(\tau)$).

We now compare the quality of the estimates obtained with the proposed estimation method to those obtained with the POT alternatives and the MLE. In Figure 4, we first display the estimated parameters $\hat{\beta}_0^{\xi}$, $\hat{\beta}_1^{\xi}$ and $\hat{\beta}_5^{\xi}$, reflecting the various types of covariates included in the regression model (we restrict our attention to these three parameters for the sake of brevity, results for the other covariates are identical and available upon demand). For DGP I, unsurprisingly, our approach (columns `WMLE` and `CWMLE`) delivers unbiased and precise estimates, comparable to the MLE (column `MLE`). On the contrary, the POT

[9]Repetitions of the simulation experiment with different contamination rates for DGP II can be found in the Supplement.

approaches exhibit much more variability, and display a significant bias when the chosen threshold is too small (e.g. columns `CPOT90` and `POT90`, relying on conditional and unconditional quantiles at level 90%). Turning to DGPs II and III, we observe similar results despite the misspecification of the splicing model: `WMLE` and `CWMLE` deliver better results than the POT. In addition, for these DGPs, our approach is superior to the classical MLE, which exhibits larger bias and an important dispersion. We do not observe major differences across the different parameters. In the Supplement (Section 3), we report detailed estimates of the bias. In Figure 5, we compare the RMSE associated to each estimated regression parameter. For clarity, we display the quantities $RMSE/RMSE^* - 1$, with the superscript "$*$" referring to the errors obtained with the MLE. As such, negative values indicate a lower error rate than the MLE. In terms of RMSE on the regression coefficients, the splicing approach is vastly superior to the POT approaches. For DGP I, the censored likelihood estimators (columns `WMLE` and `CWMLE`) are comparable to the MLE (with a difference of around 1%). Moreover, for DGPs II and III, the censored likelihood estimators are much better than the MLE, with RMSE reductions ranging between 15% and 65%. The POT approaches perform relatively better for DGP III, especially for the constant term, but the RMSE is still much larger than those of the censored approaches. Thus, while the addition of the auxiliary regression model decreases the variability of the estimates (column `MLE`) compared to the POT, it is its combination with the censoring procedure that leads to a simultaneous decrease in bias and variability (columns `WMLE` and `CWMLE`).

We now turn to the performance of the confidence intervals obtained with the different methods. In Figure 6, we display the power curves obtained from testing $H_0 : \beta_j^{\xi} = \beta_j^{H0}$, $j = 0, 1, 5$ with the different methods, for various values of β_j^{H0}. Results for the other coefficients are comparable, and are available upon demand. The power curves indicate a clear superiority of the splicing approaches (dashed and solid red curves in Figure 6): these methods exhibit respectable sizes (however slightly above the nominal level - between 6.5%

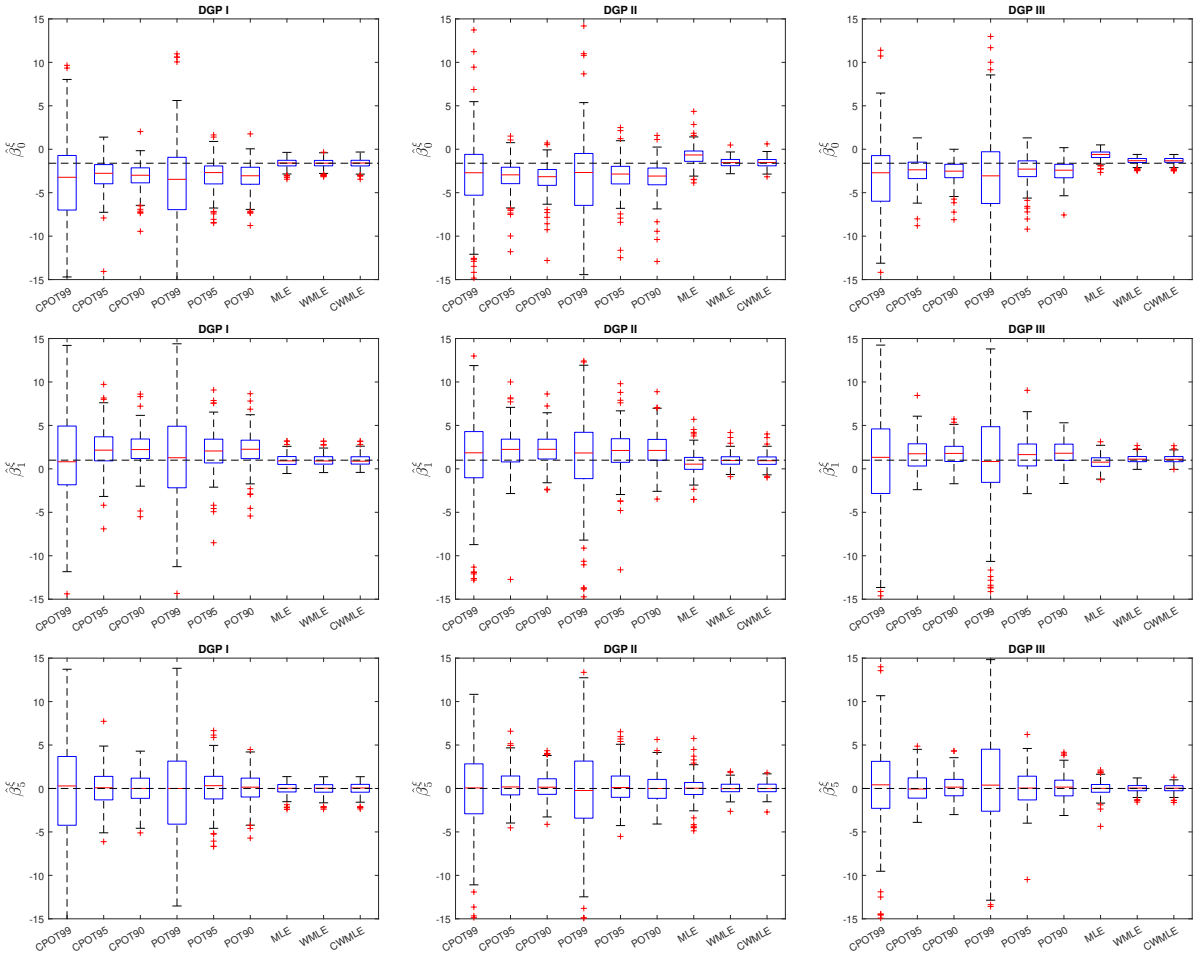


Figure 4: Estimated parameters $\hat{\beta}_0^{\xi}$ (top), $\hat{\beta}_1^{\xi}$ and $\hat{\beta}_5^{\xi}$ (bottom) obtained with the proposed method and various POT approaches. Dashed: values of the true parameters. Columns `CPOT99`, `CPOT95` and `CPOT90` display the results obtained with the (conditional) 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. Details on the POT computations can be found in the Supplement.

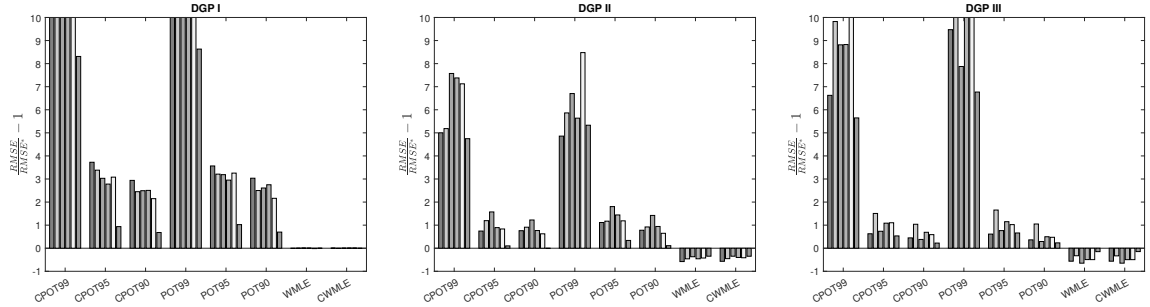


Figure 5: $RMSE/RMSE^* - 1$ for the various regression parameters, for DGP I to III.

and 8% for the regression coefficients), and powers that are quickly increasing as soon as we are departing from the true null hypothesis. The POT approaches, although exhibiting good coverage rates of the true null hypothesis for high thresholds, fare poorly in terms of power. The use of a smaller threshold in the POT leads to power improvements, but it comes at the cost of worse coverage rates. In the Supplement, we also report the median length of the 95% confidence intervals. The censored estimators are clearly favored on this criterion as well: as soon as the true DGP does not exactly correspond to the splicing regression models (DGPs II and III), the MLE suffers from significant biases, coverage rates well below the nominal level, and a loss of power (see, e.g., bottom right on Figure 6).

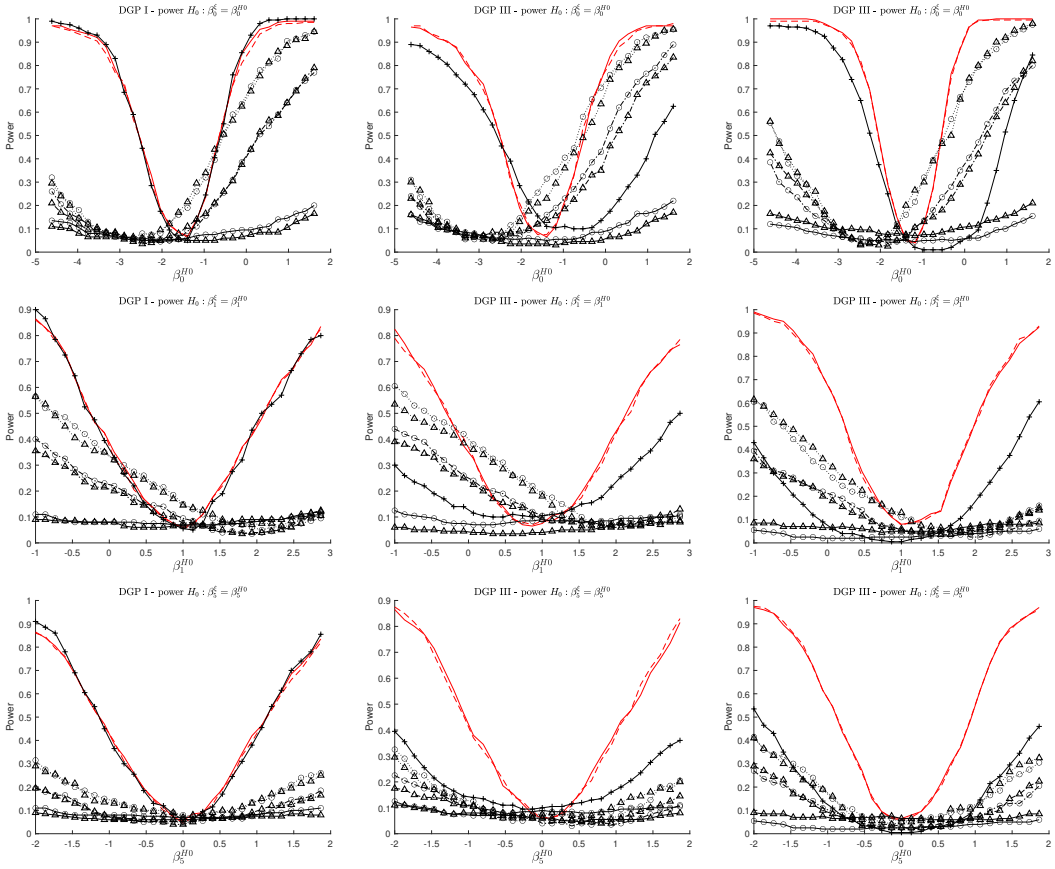


Figure 6: Power curves for the null hypotheses $H_0 : \beta_j^\xi = \beta_j^{H0}$, $j = 0, 1, 5$, for different values of β_j^{H0}. From left to right: DGP I to DGP III. For β_j^{H0} equal to the true parameter, we expect a rejection rate equal to the 5% test level. Red solid and dashed: `WMLE` and `CWMLE`. +: `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.

To conclude this section, we have demonstrated that the censored likelihood estimation strategy clearly improves the bias-variance trade-off in the estimation of the tail index regression parameters. In particular, when the true DGP is misspecified with respect to the splicing model (i.e. DGP III), the bias and the variance of the estimated regression effects are relatively small, while the power and the coverage rate remain satisfactory. Additional simulation evidence can be found in Section 1 of the Supplement. Notably, we investigate different contamination rates in DGP II, as well as a case without heteroscedasticity. Our results and conclusions stay unchanged. We also consider a DGP based on a contaminated Fréchet distribution, where the 30% smallest observations are replaced by an inverse GPD. In this setting less favorable to the proposed method, the relative efficiency of the POT increases significantly, but our WMLE stays largely superior.

4 Hedge funds tail risks

In this section, we study the determinants of *hedge funds tail risks*, relying on the proposed approach. As tail risk measure, we use the conditional tail index of the cross-sectional distribution of hedge funds' returns, i.e. $\xi(\mathbf{x}_t)$. That is, we are interested in a tail index characterizing the tail distribution of a given group of funds at a given time t, and not in fund-specific tail indices. Therefore, we only use time-varying covariates $\mathbf{x}_t$, and no fund-specific covariates. Notice that this definition is conceptually similar to what Kelly and Jiang (2014) coined under the term 'tail risk', however extracted here from hedge funds returns instead of stock returns, and including a regression structure. Therefore, the present analysis differs drastically from Jiang and Kelly (2012), who study hedge funds exposure to the stock-based tail risk measure (or factor) of Kelly and Jiang (2014). Here, we employ similar extreme value theory concepts to extract time-varying tail indices, but we do so solely using hedge funds returns and a set of covariates. Moreover, we are primarily interested in inferring on the *determinants* of the tail index dynamic over time, and not

in constructing a specific asset pricing factor. Thus, as advocated in Sections 1 and 2, we resort to an extreme value regression model explicitly allowing for regression effects estimated on pooled hedge funds returns, and not to monthly nonparametric tail index estimates, as in Kelly and Jiang (2014). The pooling of exceedances over time is essential here since it allows incorporating the covariates in the model. While it is well acknowledged that hedge funds exhibit tail risks, there is no clear characterization of the link between economic risk factors and the conditional tail distribution of hedge funds (Bali et al., 2007). Differently, previous studies relying on EVT to measure tail risks focus either on stocks (Kelly and Jiang, 2014; Huang et al., 2012) or assess hedge funds exposure to stock tail risks (Jiang and Kelly, 2012), but do not study hedge funds tail risk determinants[10]. We fill this gap in the present section and also estimate $\xi(\mathbf{x}_t)$ from hedge funds split between investment strategies (Section 4.3), with the idea of highlighting the heterogeneity in risk profiles along these strategies. Finally, we briefly illustrate how the computed hedge funds tail risk can be used to provide novel insights on the return-generating process of hedge funds.

4.1 Data preparation

We conduct our analysis using monthly net-of-fees returns of hedge funds reporting in US dollars over the period 01/1995-09/2021 in the EurekaHedge database, and declaring a strategy of the types *Long/Short Equity*, *CTA/Managed Futures*, *Fixed Income* or *Macro*. Both dead and live funds are included in the final sample. To address backfill biases, we applied the procedure of Jorion and Schwarz (2019), effectively removing returns from months prior to a recomputed add date[11] in the database, and solely keeping funds with

[10]See also Bali (2003) and Gupta and Liang (2005) for the use of extreme value theory to measure value-at-risk at the fund level.

[11]The earliest recomputed add date in our sample is September 2002, which causes us to remove observations prior to that date.

at least 60 months of remaining uninterrupted history. In addition, to remove the effect of return smoothing, we used the Getmansky et al. (2004) moving average model with two lags to reconstruct unsmoothed returns. Our final sample consists of 208,345 monthly observations spanning 1,982 funds.

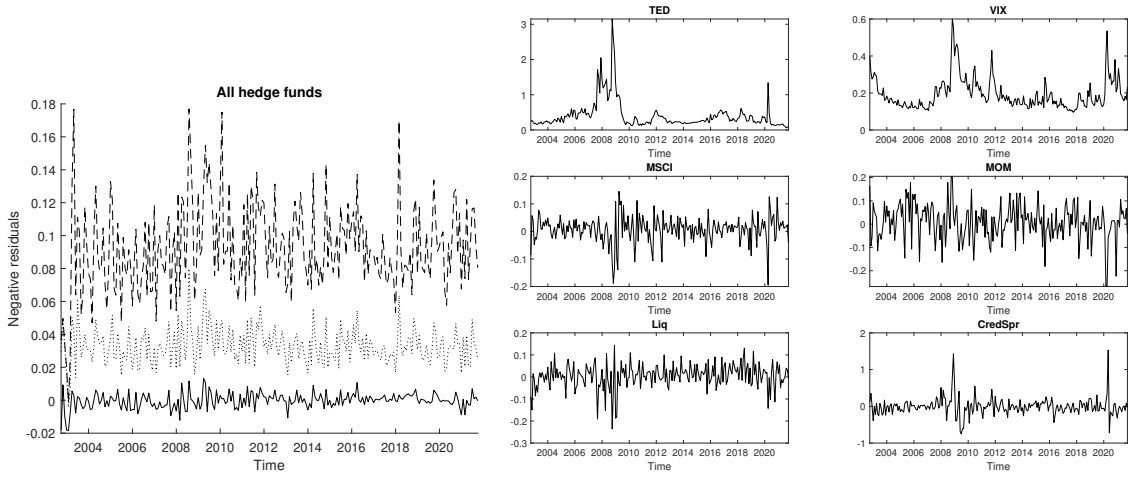
In a preliminary filtering step, we removed time variations in the mean of the unsmoothed returns with an asset pricing model (Patton and Ramadorai, 2013), following Kelly and Jiang (2014). Details of this step are given in the Supplement (Section 2). We then use the negative residuals of this model (i.e. the residuals multiplied by -1) to estimate the tail index $\xi(\mathbf{x}_t)$ and the associated regression effects. We denote by r_{it} the observed return at time t of fund $i = 1, \ldots, I$, reporting during n_i months. The total number of observations is thus $n = \sum_{i=1}^{I} n_i$. The negative residual associated with r_{it} is denoted $\hat{y}_{it}$ to stay consistent with the notation of Section 2. We first study the cross-sectional distribution for the four strategies together, before discussing strategy-specific tail risk measures to account for potential heterogeneity[12] in Subsection 4.3.

In Figure 7, panel (a), we display the empirical mean, 90% and 99% quantiles of the cross-section of negative residuals at each point in time. We observe significant variations over time, in particular at the onset of crisis periods. A study of the tail heaviness of the data is provided in the Supplement and validates our assumption that $\xi(\mathbf{x}_t) > 0$.

As contemporaneous covariates for $\xi(\mathbf{x}_t)$, we consider the following factors: the TED spread in level, the monthly change of the TED spread ($\Delta\texttt{TED}$), the CBOE volatility index VIX, the returns of the MSCI world index, the global equity momentum factor of Moskowitz et al. (2012), the liquidity factor of Pastor and Stambaugh (2003), and the credit spread factor of Fung and Hsieh (2004). The time series of the covariates are displayed in Figure 7,

[12]See, e.g., Mhalla et al. (2022) and Dupuis et al. (2023) for discussions on this pooling approach. Notice that we are interested in inferring on the tail of the cross-sectional distribution of the funds' returns. Therefore, we do not try to control further for fund-specific heterogeneity. Nevertheless, the preliminary filtering procedure reduces potential heterogeneity problems, and this point is tackled further in the strategy-specific analysis.

panel (b). These variables are suggested by Bali et al. (2014), Agarwal et al. (2017), and Patton and Ramadorai (2013) as capturing tail risk over time. In particular, market returns can be seen as capturing systematic tail risk of hedge funds and benchmark pressures. The other factors span various dimensions of disturbances on the financial market, found to influence strongly the investment strategy adopted by hedge funds: Uncertainty, volatility, momentum, market liquidity, and funding constraints. $\Delta\texttt{TED}$ is used to proxy funding liquidity shocks, in the spirit of Agarwal et al. (2017). Indeed, from a theoretical standpoint, Brunnermeier and Pedersen (2009) suggest that funding liquidity and market liquidity interact, influencing the margin requirements of hedge funds, thus their trading behaviour: In times of costly margin requirements (high funding costs), hedge funds reduce their positions, thus reducing tail risks. On the other hand, lower market liquidity may also imply higher margin requirements, often leading funds to close positions during a crisis and to incur large losses, suggesting a higher tail risk in these situations.



(a) (b)

Figure 7: (a): Mean (solid), 90% (dotted) and 99% empirical quantile (dashed) of the cross-section of negative residuals $\hat{y}_{it}$. (b): Time series of the covariates.

In the next section, we focus solely on the tail index regression parameters $\hat{\boldsymbol{\beta}}^{\xi}$, as it is not possible to have a clear interpretation of those related to s and σ, treated here as "nuisance parameters" and having a similar role as control variables. Nevertheless, to give as much

flexibility as possible to our model, we let s and σ be functions of the same covariates. Estimates can be found in the Supplement. Remind that, since our covariates are solely time-specific and not fund-specific, we have $\xi(\mathbf{x}_{it}) = \xi(\mathbf{x}_t)$, $\forall i = 1, \cdots, I$. Moreover, each column of the predictor matrix has been standardized.

4.2 Main results

First, we estimate the marginal regression effects of the covariates on $\xi(\mathbf{x}_t)$ for the complete sample of funds. We consider both conditional and unconditional censoring procedures, but focus on the latter approach as the results are similar in the two cases. In Figure 8, left panel, we display the AD^m statistic as a function of τ for both approaches. We find an optimal value of 0.216, a result consistent with our simulation study and suggesting a misspecification of the G-E-GPD model in the left tail. On the right panel of the same

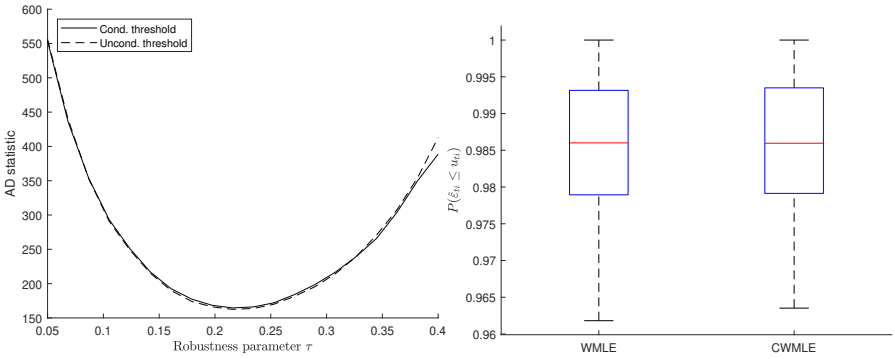


Figure 8: Left: AD^m statistics as a function of τ, using either an unconditional quantile (dashed) or a conditional quantile (solid) at level τ as censoring threshold. Right: boxplot of the distribution $\mathbb{P}(\hat{y}_{ti} \leq u_{ti})$).

figure, we display the boxplot of the quantile levels corresponding to the implicit thresholds u_{it} obtained with `WMLE` and `CWMLE` (i.e. $\mathbb{P}(\hat{y}_{it} \leq \hat{u}_{it})$). This quantity indicates how far in the tail the GPD approximation starts under the G-E-GPD model. We find that conditional quantiles at levels between 0.98 and 0.99 (for a median around 0.985) would be suitable thresholds for most combinations of the covariates. For some combinations, though, lower threshold levels (as small as 0.965) are selected. Overall, around 1.5% of the observations in our sample are found to exceed these thresholds.

Table 1: 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.

Covariate	WMLE	CWMLE	MLE	POT95	POT97.5
β_0^ξ	-.998***	-.998***	-0.47***	-2.15***	-2.25***
	$[-1.07, -0.92]$	$[-1.08, -0.92]$	$[-0.53, -0.41]$	$[-2.36, -1.94]$	$[-2.59, -1.92]$
β^ξ(TED)	-0.26***	-0.24***	0.04	-0.03	-0.07
	$[-0.36, -0.16]$	$[-0.34, -0.15]$	$[-0.06, 0.14]$	$[-0.28, 0.23]$	$[-0.43, 0.3]$
β^ξ(VIX)	-0.20***	-0.20***	-0.10**	0.15	0.15
	$[-0.30, -0.10]$	$[-0.31, -0.09]$	$[-0.19, -0.01]$	$[-0.02, 0.33]$	$[-0.12, 0.42]$
$\beta^\xi(\Delta$TED)	-0.43***	-0.44***	-0.43***	0.02	-0.04
	$[-0.5, -0.36]$	$[-0.51, -0.37]$	$[-0.52, -0.35]$	$[-0.22, 0.26]$	$[-0.41, 0.32]$
$\beta^\xi(\Delta$MSCI)	0.12**	0.12**	0.04	0.02	-0.14
	$[0.04, 0.2]$	$[0.04, 0.21]$	$[-0.02, 0.1]$	$[-0.12, 0.16]$	$[-0.38, 0.09]$
β^ξ(MOM)	0.19***	0.20***	0.17***	-0.08	-0.08
	$[0.12, 0.25]$	$[0.13, 0.27]$	$[0.11, 0.23]$	$[-0.25, 0.09]$	$[-0.31, 0.16]$
β^ξ(Liq)	0.01	0.01	0.15***	0.21**	0.29**
	$[-0.05, 0.06]$	$[-0.05, 0.06]$	$[0.08, 0.22]$	$[0.03, 0.4]$	$[0.02, 0.56]$
β^ξ(CredSpr)	-0.13***	-0.14***	-0.03	-0.11	-0.07
	$[-0.18, -0.08]$	$[-0.19, -0.09]$	$[-0.1, 0.04]$	$[-0.28, 0.06]$	$[-0.31, 0.16]$
β_0^σ	-4.03***	-4.03***	-3.46***	-3.67***	-3.56***
	$[-4.04, -4.02]$	$[-4.04, -4.02]$	$[-3.46, -3.45]$	$[-3.65, -3.59]$	$[-3.56, -3.47]$
Controls ?	Yes	Yes	Yes	Yes	Yes
β_0^s	-3.56***	-3.56***	-3.47***	-	-
	$[-3.58, -3.55]$	$[-3.58, -3.55]$	$[-3.49, -3.45]$	-	-
Controls ?	Yes	Yes	Yes	-	-
β_0^m	-0.002***	-0.002***	-0.001***	-	-
	$[-0.002, -0.002]$	$[-0.002, -0.002]$	$[-0.001, -0.001]$	-	-
τ^{opt}	0.216	0.216	-	-	-

In Table 1, we report point estimates of the regression coefficients and their confidence intervals, for the G-E-GPD methods and the POT-based methods[13]. Looking at the results for WMLE (first column), we find the TED spread (TED and ΔTED), the credit spread factor (CredSpr), the VIX (VIX), the momentum factor (MOM) and the market return (ΔMSCI) to display regression coefficients significantly different from zero. A decrease in TED, ΔTED CredSpr, and VIX are associated with an increase in tail risk, suggesting a propensity for the cross-section of hedge funds to be particularly exposed to tail risk in a context of cheap funding conditions and low uncertainty. Similarly, an increase in time series equity momentum and market returns are associated with an increase in tail risk, consistent with the idea that tail risk increases in booming market conditions.

[13]We use the same set of covariates for both estimation methods, and an unconditional threshold in the POT. Results with conditional thresholds are qualitatively alike and available upon demand, while regression estimates for σ obtained with WMLE, MLE and POT97.5 can be found in the Supplement.

The signs of these estimates are in line with the theoretical mechanism described by Brunnermeier and Pedersen (2009), which sees funds reducing their risk exposures in times of expensive margin requirements, funding liquidity shocks and high expected volatility. Empirically, our findings may be also related to the literature on time-varying risk exposure of hedge funds and market-timing abilities. In particular, Patton and Ramadorai (2013) highlight the role of funding liquidity, market returns and market uncertainty in influencing risk exposure over time (via allocation decision) at the fund level. More precisely, they show that hedge funds tend to abruptly cut their positions in response to an abrupt rise in volatility (proxied by intra-month realized volatility or the VIX), rise in funding costs (proxied by the TED spread), and fall in market returns (measured by the S&P 500). In addition, Darolles and Roussellet (2024) show both theoretically and empirically that when funding liquidity deteriorates, funds increase their cash holdings and decrease their exposure to illiquid assets, in the perspective of avoiding liquidation triggered by client withdrawals. They highlight that this behavior is consistent with funds actively managing liquidity risks on both the asset and liability sides of their balance sheet and that, counterintuitively, portfolio composition is informative about liquidity exposure on the liability (i.e. client) side. Their model implies that when funding liquidity becomes poorer, the fund adjusts its cash holding upwards such that its liquidation probability stays the same and generates lower expected returns because of the portfolio composition shift (see also Liu and Mello, 2011). Such a shift to cash is consistent with the decrease in tail risk associated with an increase in `TED`, Δ`TED` and `CredSpr`: *ceteris paribus*, considering that the effect on expected return has been removed in the preliminary filtering step, the limit probability of extreme losses (as a percentage of total assets) decreases, since a larger fraction of the funds' assets (the one in cash) is immune to a decrease in value. Moreover, their framework offers an interesting potential explanation regarding the limited effect of `Liq`: at equilibrium, they predict that the optimal amount of cash holdings is insensitive to marginal changes

in market liquidity because the higher liquidation probability is compensated by a higher return on the illiquid assets, with the latter effect dominating the former. There is therefore no change in risk exposure, which in our model would translate into an absence of marginal effect on the speed of decay of the probability of extreme losses, i.e. on the tail index. More generally, our results seem to indicate that hedge fund tail risk dynamic is related to the limits to arbitrage (Shleifer and Vishny, 1997) and the inner economic drivers leading funds to change their portfolio composition to ensure that their survival likelihood stays at acceptable levels (Lan et al., 2013).

Note that one might be tempted to also compare our results directly with those of Agarwal et al. (2017). However, such a comparison is misplaced, due to the difference in tail risk definition between the two studies: While we focus on estimating the tail index of the cross-sectional distribution of hedge funds returns, conditional on covariates (such as market returns), Agarwal et al. (2017) estimate fund-by-fund tail dependence measures with the market, multiply them by the ratio between the expected shortfalls (ES) for the fund and for the market, then average them over all funds alive at specific times. This difference is crucial since their measure is akin to calculating weighted averages of the funds ES, expressed in units of market risk[14]. Thus, if in a given month the ES of the market increases faster, *ceteris paribus*, than the average ES of the funds, their aggregate tail risk measure decreases, a reflection of their measure being negatively related to the ES of the market. On the contrary, our measure does not directly depend on market risk, but is simply computed conditionally to realized market returns. Therefore, we estimate the tail index of the cross-section of funds (and the related marginal regression effects) under specific market conditions without imposing a negative relation between tail risk and market risk, and not how tightly market and funds risks are co-moving together. A

[14]The weights are given by the tail dependence measures - named tail sensitivity in their article. Since the tail sensitivity is calculated with a 24-month rolling window and is based on at most two exceedances above high thresholds, these weights only take values 0, 0.5 or 1 - see Section 3.2 in Agarwal et al. (2017).

second fundamental difference is that we use the tail index as a risk measure, and not ratios of ES. Our measure is therefore invariant to simple scaling effects that arise, e.g., with an increase in cross-sectional volatility. While an increase in Agarwal et al. (2017) tail risk measure might simply be due to a larger increase in the volatility of hedge funds returns compared to the one of market returns, this is not the case with our tail risk definition, based on the tail index (Embrechts et al., 1997). An increase in our tail risk measure will be solely due to a higher limit probability of extremely large losses. In Agarwal et al. (2017) setting, this effect cannot be disentangled from changes in volatility. One solution to reconcile our findings with theirs would be to investigate the signs of $\boldsymbol{\beta}^{\sigma}$ (see Supplement, Section 3) as well, and not only $\boldsymbol{\beta}^{\xi}$. For example, we observe a positive effect of the TED for σ_t, indicating a higher dispersion in the tail in times of decreasing funding liquidity, consistent with the findings of Agarwal et al. (2017). But, as mentioned earlier, in doing so we would step away from our own tail risk definition, which is scale-invariant. Thus, we refrain from comparing the empirical results reported in Agarwal et al. (2017) with the present ones.

Since the use of financial leverage might modulate the effects of funding liquidity and market liquidity, we repeat our regression analysis on subsamples of funds grouped on their use of leverage. Estimates are reported in the Supplement (Section 3.2). We do not observe differences in terms of signs of the significant coefficients, which confirms our main analysis: All variables are found to have a significant effect, except market liquidity (`Liq`). However, in terms of effect size, tail risk of leveraged funds appears to be more sensitive to funding liquidity levels (`TED`). This result is once again in accordance with the margin channel described by Brunnermeier and Pedersen (2009).

Focusing on the POT approaches, no regression effects are found to be significantly different from zero, except for `Liq`. The G-E-GPD approach suggests thresholds in the range of a 98%-99% quantile, above which the GPD approximation should hold. However,

as already indicated by the EVR results obtained with a 97.5% quantile as threshold for the POT, the use of such high thresholds leaves us with very few data, and leads to no significant estimates. This result highlights therefore the efficiency gains obtained with the G-E-GPD approach.

Adopting a time-series view, we compute $\hat{\xi}(\mathbf{x}_t)$ for $t = 1, \ldots, T$ using our `WMLE` estimates (named HKU hereafter). We display these quantities and the associated pointwise 95% confidence intervals in Figure 9. The HKU estimates (red solid) suggest a higher risk level of the funds' cross-sectional distribution than the POT estimates (black solid), and more variations over time than the month-by-month Hill estimator of Kelly and Jiang (2014) (dashed black, denoted KJ hereafter). In particular, the HKU estimate indicates a high tail risk level before large systemic events, such as the global financial crisis in 2007-2008 or the debt crisis in 2010, followed by a rapid decline. These variations appear to reflect well the dynamic nature of hedge funds' investment strategies, expected effectively to take more risk exposure in booming periods, and to wind down their risks in times of market turmoils (Patton and Ramadorai, 2013; Darolles and Roussellet, 2024). On the contrary, the POT and KJ estimators do not seem to capture these differences over time.

Finally, we compute the historical correlations between our tail risk measure and a series of financial and economic indicators (Table 2). We find a strong negative correlation with all volatility and financial stress measures, namely VIX, VVIX, VGOLD, VOIL, VFX, FSI and EPU (see the legend of Table 2 for more details). Conversely, we also observe a positive correlation with FX, crypto, global and US market levels or returns (BTC, FX Adv., FX Em., MSCI and SP500). Correlations are found stronger with the stock markets, and milder with the other markets. Finally, it correlates strongly with the credit spread factor (negatively) and the momentum and betting-against-beta factors (positively). Some of these results are not surprising, since several of these variables are explicitly included in the tail index equation. However, it is interesting to observe the coherence with the

interpretation of the results (in line with the deleveraging hypothesis) and the broad but simultaneous overlap with most economic and financial fundamentals. Only the GSCI commodity index and the value factor display correlations lower than 10%. In contrast, the POT and KJ estimators appear less broadly correlated. The POT (resp. KJ) exhibits 15 (resp. 12) correlations out of 22 to be smaller than 10%. The POT correlates strongly with VIX and VVIX for the financial indicators, and with the credit spread, momentum and liquidity factors. Differently, KJ correlates mostly with VGOLD and BTC, and the bond and currency trend-following factors of Fung and Hsieh (2004).

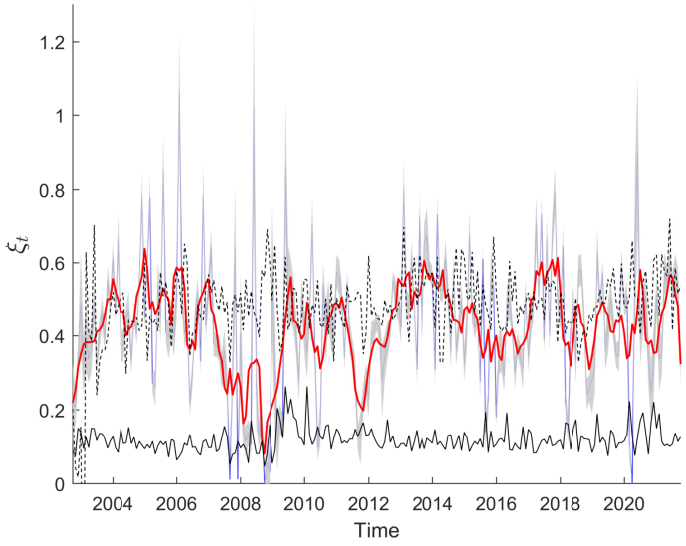


Figure 9: Time series of $\hat{\xi}(\mathbf{x}_t)$ (solid blue). Red: smoothed estimate of `WMLE` obtained from a moving average filter. Shades of grey: 95% pointwise confidence intervals for $\xi(\mathbf{x}_t)$. Solid black: POT estimates with 95% empirical quantile as threshold. Dashed black: Hill estimator proposed by Kelly and Jiang (2014) with the monthly 90% empirical quantile as threshold.

4.3 Strategy-specific analysis

Results presented in the previous subsection are obtained from the entire sample of hedge funds. However, funds with different investment strategies are expected to have risk profiles that are different (Sadka, 2010). To investigate further this casual observation, we repeat our main analysis, but pooling funds according to their reported investment strategy. We

Table 2: Correlation coefficients between the estimated tail risk measures, financial and economic indicators, as well as asset pricing risk factors. Size and Value refer to the factors of Fama and French (1993). BonT, CurT and ComT refer to the bond; currency and commodity trend-following factors of Fung and Hsieh (2004). CredSpr refers to the credit spread factor of the same authors. MOM, BaB and Liq refer to the time-series equity momentum factor of Moskowitz et al. (2012), the betting-against-beta factor of Frazzini and Pedersen (2014) and the liquidity factor of Pastor and Stambaugh (2003). For the economic and financial indicators, we use the VIX, the VVIX (vol of VIX), the CBOE gold ETF (VGOLD) and crude oil ETF (VOIL) volatility index, realized volatility computed from an advanced economies FX index (VFX), the MSCI and S&P500 returns, the S&P GSCI commodity index (Com.), the Bitcoin returns (BTC), two global exchange rate indices vis-a-vis the US Dollar (for advanced and emerging economies: `FX Adv.` and `FX Em.`), as well as the categorical economic policy uncertainty index on monetary policy (EPU) and the Financial Stress Index (FSI) of St Louis Fed.

$\hat{\xi}(\mathbf{x}_t)$	VIX	VVIX	VGOLD	VOIL	VFX	BTC	Com.	FX Adv.	FX Em.	FSI	EPU
WMLE	-0.53	-0.25	-0.45	-0.34	-0.32	0.19	-0.01	0.11	0.10	-0.41	-0.28
POT95	0.26	0.21	0.08	0.07	0.05	-0.02	-0.06	0.02	0.11	0.07	-0.01
KJ	-0.09	0.02	0.22	0.07	0.08	0.28	0.20	0.08	0.06	0.01	-0.12
$\hat{\xi}(\mathbf{x}_t)$	MSCI	SP500	Size	Value	BondT	CurT	ComT	CredSpr	MOM	BaB	Liq
WMLE	0.52	0.48	0.17	0.07	-0.24	-0.11	-0.13	-0.33	0.43	0.29	0.20
POT95	0.01	-0.01	0.03	-0.00	0.05	-0.04	0.27	-0.35	-0.50	0.06	0.63
KJ	-0.03	-0.04	-0.03	-0.05	0.18	0.15	0.02	0.00	0.11	0.03	0.13

then estimate one regression model per strategy.

In Table 3, we report descriptive statistics of $\hat{y}_{it}$ for each of the four investment strategies (*Long/Short Equity*, *Macro*, *Fixed-income* and *CTA*). See Pedersen (2019) for a more detailed description of the strategies. Our sample is largely dominated by *Long/Short Equity* funds. In terms of distributional characteristics, *Fixed-income* and *Long/Short Equity* funds display more kurtosis than the other two strategies. Moreover, *Fixed-income* funds exhibit the smallest standard deviation, while *Long/Short Equity* funds display the strongest skewness. Subsample sizes vary roughly between 120,600 and 17,000 observations, with a monthly average varying from 527 observations for *Long/Short Equity* funds to 75 observations for *Macro* funds.

Table 3: Descriptive statistics for $\hat{y}_{it}$. Columns n_i and $\bar{n}_{it}$ report the total number and the average monthly number of observations. Columns $\bar{\mu}_t$, $\bar{s}_t$, $\bar{sk}_t$, $\bar{ku}_t$ report averages of the monthly mean, standard deviation, skewness and kurtosis computed from cross-sections of $\hat{y}_{it}$.

Strategy	# funds	n_i	$\bar{n}_{it}$	$\bar{\mu}_t$	$\bar{s}_t$	$\bar{sk}_t$	$\bar{ku}_t$
All	1,972	208,347	910	-0.001	0.033	-0.58	17.30
L/S Equity	1,102	120,597	527	-0.000	0.032	-0.53	15.41
CTA	404	42,645	186	-0.001	0.041	-0.19	8.32
Fixed-Income	295	28,029	123	-0.001	0.016	-0.08	17.24
Macro	171	17,076	75	-0.001	0.033	-0.11	8.52

In Figure 10, we report the estimated regression effects $\hat{\boldsymbol{\beta}}^{\xi}$ obtained with either our method or the POT approach (for simplicity, we focus on `WMLE` and `POT97.5`). We always select $\tau^{opt} = 0.216$ as censoring threshold for the WMLE. Similarly to the main analysis, almost no regression coefficients are found to be significant with the POT approach, due to the huge estimation uncertainty. Regarding the WLME, for *Long/Short Equity*, the estimated effects are overall in accordance with the main regression analysis, with only Δ`TED` becoming insignificant. Some noticeable differences are found for the other strategies: The effect of Δ`TED` is found to be positive and significant for *CTA*, while for *Fixed-Income*, only `CredSpr` and `VIX` are found to be significant, the latter with a positive effect contrary to the other strategies. The tail index for *Macro* seems to be driven mostly by `TED`, Δ`TED` and Δ`MSCI`, but not by the other variables.

In Figure 11, we display the estimated tail risk measures over time, for each strategy. As for the regression estimates, *Long/Short Equity* features a similar pattern than the main analysis while the other strategies exhibit more specific dynamics. For example, *Fixed Income* has a rather constant tail index, except in 2009-2010 when it reached its maximum (close to 1). On average, *Fixed Income* appears to have the highest tail risk over the period 2002-2020. *CTA/Managed Futures* and *Macro* feature more variations over time, with a strong decrease before the 2008 crisis. All strategies except *Long/Short Equity* display a marked tail risk increase at the beginning of the COVID crisis.

These results highlight therefore a relative heterogeneity between funds with different investment strategies. Importantly, the use of the WMLE allows us to document these differences, while with the POT we cannot overcome the small-sample issues stemming from analyzing strategy-specific subsamples.

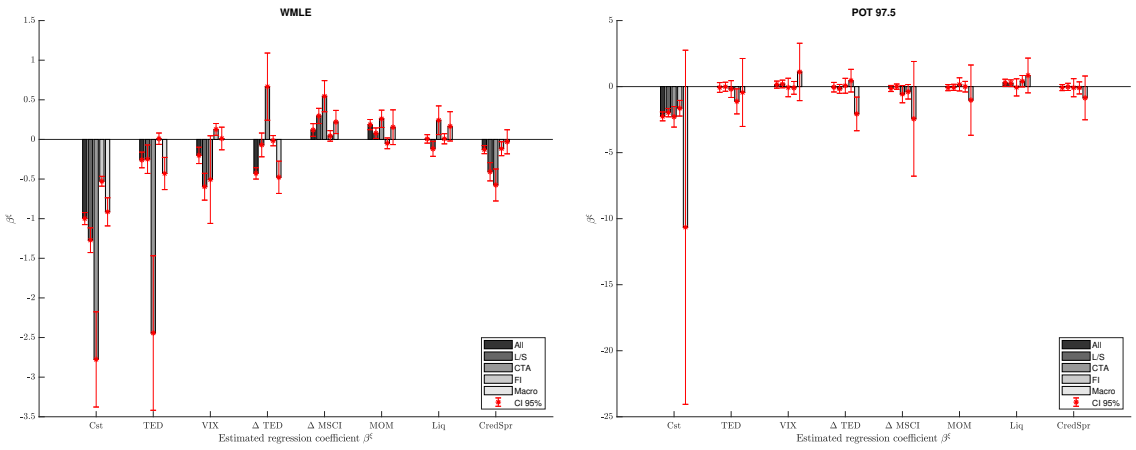


Figure 10: Estimated extreme value regression effects obtained from strategy-specific residuals, and associated confidence intervals (red). Left: WMLE estimates. Right: POT estimates with a 97.5% threshold. Notice that the graphs do not have the same scale on the y-axis.

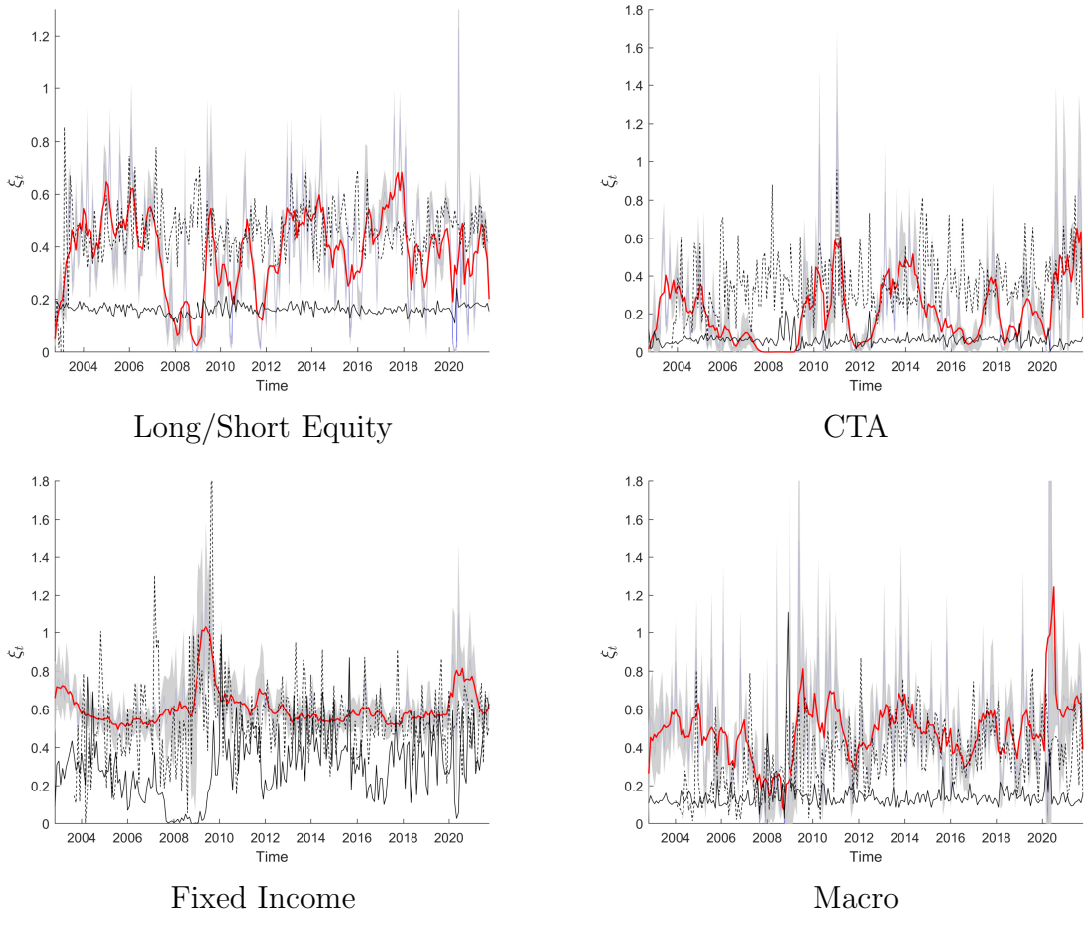


Figure 11: Time series of $\hat{\xi}(\mathbf{x}_t)$ (solid blue) for the four different strategies. Red: smoothed estimate of `WMLE` obtained from a moving average filter. Shades of grey: 95% pointwise confidence intervals for $\xi(\mathbf{x}_t)$. Solid black: POT estimates with 95% empirical quantile as threshold. Dashed black: Hill estimator proposed by Kelly and Jiang (2014) with the monthly 90% empirical quantile as threshold.

4.4 Implications for hedge funds return-generating processes

Can our hedge fund tail risk measure help us better understand the return-generating process at the fund level? To answer this question and further illustrate the potential of our estimation method, we now investigate (i) whether our conditional hedge fund tail risk measure conveys information about future hedge fund returns, in the sense of Avramov et al. (2013), and (ii) if the strength of the relationship between our tail risk measure and future returns at the fund level (i.e. tail risk sensitivity) is a feature of the fund that helps distinguish between high- and low-alpha funds, with implications for portfolio selection. We contrast these results with those obtained from two other tail risk estimation methods, namely POT and KJ.

Notice that, in this section, we do not ambition to construct a novel factor nor to explain the cross-section of hedge funds returns[15]. Instead, we first investigate if our tail risk measure at time t has a significant explanatory power for the observed return at time $t+1$ beyond usual contemporaneous asset pricing factors (Avramov et al., 2013). This question is motivated by Avramov et al. (2013), who showed that a small set of macro-financial variables effectively predict hedge funds returns, and by Jiang and Kelly (2012), who showed that stock tail risk is an important determinant of hedge funds returns over time. We therefore update their results for a more recent time period, and complement them by demonstrating that a similar ability is associated with our hedge funds tail risk measure embedding simultaneously macro-financial information and realized tail risk of hedge funds. As we will show, this ability is found to be superior to the POT and KJ measures. Second, we examine if the intensity of the correlation between our tail risk

[15]This is why we do not use contemporaneous tail risk estimates in our regressions. Moreover, in its present state, $\hat{\xi}(\mathbf{x}_{t-1})$ has not the requested economic rational to be interpreted as a factor in an asset pricing sense. Nevertheless, in the Supplement, we report additional results where we use contemporaneous tail risk measures in regressions (13). We report the proportion of funds with a significant coefficient related to contemporaneous measures. The predictive ability increases overall for all measures compared to the specification with lagged tail risks, with the HKU estimates slightly outperforming KJ and the POT.

measure at time t and excess return at time $t+1$ at the fund level is a fund characteristic sufficiently strong to indicate the presence of abnormal returns beyond usual asset pricing models (i.e. alpha). A similar perspective is adopted, e.g., in Fung et al. (2008), where the authors study if cross-sectional differences between funds allow an hypothetical investor to identify funds overperforming specific benchmarks. For simplicity, and since this section is mostly illustrative of the method developed in Section 2, we restrict our analysis to static in-sample predictive regression and selection exercises using the tail indices estimated on the entire panel of funds. A similar exercise is carried out by Avramov et al. (2013), Section IV-A. We let the out-of-sample exercise proposed by the same authors and based on rolling-window estimates, to further research.

4.4.1 Empirical strategy

We use first the conditional predictive regression framework described in Avramov et al. (2013). Starting from a classical asset pricing model, we have

$$r_{it} - r_t^f = \alpha_i^* + \boldsymbol{\beta}_i^* \tilde{f}_t + \eta_{it}^*, \tag{12}$$

where r_{it} is the return at time t for fund i, r_t^f is the 3-month treasury bill rate and $\boldsymbol{\beta}_i^*$ is the vector of loadings for the vector of risk factors $\tilde{f}_t$. We denote by α_i^* the i^{th} fund's unconditional alpha. As reference asset pricing models, we use the CAPM, the 7-factor model of Fung and Hsieh (2004) (FH) and the 6-factor model of Joenväärä et al. (2021) (JKKT). While an estimate $\hat{\alpha}_i^*$, $i = 1, \ldots, I$, may be obtained from (12), Avramov et al. (2013) pointed out that hedge fund risk premia can vary with changing economic conditions or that hedge funds have selection and timing skills that depend on the state of the economy, making fund alphas predictable. To capture this intuition, following the same authors, we therefore estimate the contribution of $\hat{\xi}(\mathbf{x}_{t-1})$ in explaining time-variations in the excess

return at the fund level, using the model

$$r_{it} - r_t^f = \alpha_i^{*0} + \alpha_i^{\xi}\hat{\xi}(\mathbf{x}_{t-1}) + \boldsymbol{\beta}_i^* \tilde{f}_t + \eta_{it}^*. \tag{13}$$

We then compute the proportion of funds with $\alpha_i^{\xi} \neq 0$, with π^0, π^+ and π^- denoting the proportions of funds with α_i^{ξ} equal, larger or smaller than zero, respectively. To control for multiple testing, we use the method of Barras et al. (2010) based on the false discovery rate. A large proportion $\pi^{+-} = \pi^+ + \pi^-$ is further evidence of significant in-sample predictive ability of the tail risk measure for hedge fund excess returns, answering our first question.

Second, to answer the question of whether our tail risk measure can be used to discriminate between high- and low-alpha funds stemming from Equation (12), we estimate the tail risk sensitivity γ_i from the single predictor equation discussed in Avramov et al. (2013):

$$r_{it} - r_t^f = \gamma_{i0} + \gamma_i\hat{\xi}(\mathbf{x}_{t-1}) + \delta_{it}, \tag{14}$$

where δ_{it} is an error term centered on zero. We then split the funds into deciles, according to their estimated γ_i, and compute their decile average unconditional alpha $\hat{\alpha}_i^*$. We hypothesize that a high γ_i is indicative of an inability to limit risk exposure, thus of low selection and timing skills, translating in low alpha. Consequently, deciles with low-γ_i funds should earn higher alpha (on average) than high-γ_i funds. This pooling approach has also the advantage to magnify the link between tail risk and alpha: by making use of the average alpha in the cross-section of funds, we decrease the estimation uncertainty compared to situations when one assesses, at the fund level, the contribution of $\hat{\xi}(\mathbf{x}_{t-1})$ in explaining the alpha. Notice that the use of (14) instead of (13) is necessary here to build the tail risk sensitivity, since we wish to assess the signal solely conveyed at time t by the tail risk measure about returns at time $t+1$, and not to further explain excess returns beyond specific asset pricing model. Having two different equations therefore ensures that the information sets used for the alpha and for the tail sensitivity computations do not overlap.

4.4.2 Results

Detailed descriptive statistics of the annualized $\hat{\alpha}_i^*$ obtained from (12) for our sample of funds are available in the Supplement, Section 3. As previously documented in Ardia et al. (2023), the CAPM and FH models provide very similar results, with an average annualized alpha around 2.1%. The JKKT model provides a much smaller average alpha at 0.9%, reflecting its better specification. Using the method of Barras et al. (2010), we calculate π^+ for α_i^*, and estimate the proportion of positive-alpha funds to be around 22% with CAPM and FH, and around 15% with JKKT.

We then compute $\hat{\alpha}_i^\xi$ for each fund, using either HKU, KJ or the POT estimates in Eq. (13). Using once again the method of Barras et al. (2010), we report first π^{+-}, i.e. the proportion of funds with $\alpha_i^\xi \neq 0$ (Figure 12, left panel). A larger portion of the funds (around 25% for CAPM and JKKT models) exhibits $\alpha_i^\xi \neq 0$ when using HKU compared to the other measures. The POT, and in particular KJ, are found to be less informative, with the latter displaying a predictive ability for at most 5% of the funds.

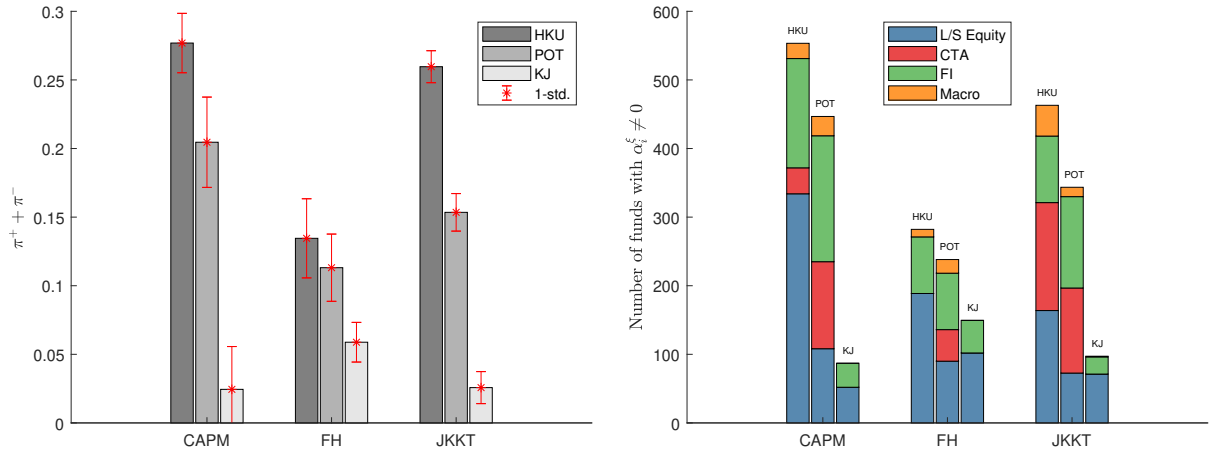


Figure 12: (a) Proportions π^{+-} of funds with $\alpha_i^\xi \neq 0$, obtained from eq. (13) with either HKU (dark grey), POT (grey) or KJ (light grey) estimates of $\xi(\mathbf{x}_{it})$. (b) Number of funds, split between investment strategies, found to exhibit $\alpha_i^\xi \neq 0$. These quantities 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.

Second, looking at strategy-specific results (Figure 12, right panel), we find in the case

of the CAPM model that mostly *Long/Short Equity* funds have predictable returns when using HKU, followed by *Fixed Income* then *CTA/Managed futures* funds. For JKKT, the number of *Long/Short Equity* funds decreases markedly, but we find a larger number of *CTA/Managed futures* and *Macro* funds that exhibit predictable returns. Using the KJ risk measure, little difference between asset pricing models is observed, with only a small number of *Long/Short Equity* and *Fixed Income* funds exhibiting predictable returns (around 5% of the funds). The POT measure appears to explain well mostly the returns of *CTA/Managed futures* and *Fixed Income* funds, but overall less often than HKU for all asset pricing models. These results suggest that our tail risk measure has the potential to improve the prediction of hedge funds returns - at least partly - compared to the two considered alternatives.

Finally, we compute $\hat{\gamma}_i$ using the different tail risk measures in Equation (14). The correlation between $\hat{\alpha}_i^*$ and $\hat{\gamma}_i$ is negative when using HKU for the three reference models, although marginally for CAPM (-4.2%) and FH (-2.5%). For the JKKT model, the correlation reaches -17%. Using POT and KJ instead of HKU, the correlation is found to be mild for all models, between 4% and 6% for KJ, and -2.7% and 6% for the POT.

Table 4: Average $\hat{\alpha}_i^*$ in the first (D1) and last (D10) decile of funds sorted on their $\hat{\gamma}_i$, with the different asset pricing models. The line D1-D10 reports the difference. *,** and ***: significant difference from 0 at the 10%, 5% and 1% test level, using Welch's t-test. In brackets: 95% bootstrap confidence intervals.

Model	HKU	POT	KJ
D1 $\hat{\alpha}_i^*$ (CAPM)	2.58	0.60	1.67
D10 $\hat{\alpha}_i^*$ (CAPM)	1.42	2.09	2.12
D1-D10	1.16	-1.49*	-0.45
Bootstrap CI	$[-0.28, 2.65]$	$[-2.88, -0.07]$	$[-1.91, 1.01]$
D1 $\hat{\alpha}_i^*$ (FH)	2.53	1.87	1.88
D10 $\hat{\alpha}_i^*$ (FH)	1.83	1.26	2.27
D1-D10	0.70	0.61	-0.39
Bootstrap CI	$[-0.69, 2.19]$	$[-0.73, 2.03]$	$[-1.94, 1.00]$
D1 $\hat{\alpha}_i^*$ (JKKT)	2.87	-0.59	0.01
D10 $\hat{\alpha}_i^*$ (JKKT)	-1.93	0.02	0.94
D1-D10	4.80***	-0.61	-0.93
Bootstrap CI	[3.19, 6.47]	$[-2.04, 0.85]$	$[-2.50, 0.60]$

Investigating the average $\hat{\alpha}_i^*$ per decile of funds sorted on their $\hat{\gamma}_i$ (Figure 13), we

observe a clear downward trend when using HKU with the JKKT model (Figure 13, left panel, dashed line): an increase in γ_i obtained with HKU is associated with a decrease in $\hat{\alpha}_i^*$. The bottom decile displays an average annualized $\hat{\alpha}_i^*$ of around 3%, while the top decile exhibits -2% annually. This result is remarkable given that the JKKT model delivers much smaller alphas than FH and the CAPM. In Table 4, we report differences in average alpha between bottom and top deciles, with positive values indicating a higher alpha for small-γ_i funds. Formally testing for differences between the average $\hat{\alpha}_i^*$ of the top and bottom deciles with Welch's t-test, we reject the null hypothesis of equal means for HKU at the 1% test level with the JKKT model. A similar result is obtained when repeating the test up to the 8^{th} decile. To stress the robustness of these results with respect to the additional uncertainty generated by the estimation error of $\hat{\alpha}_i^*$, we construct 95% confidence intervals with a residuals-based bootstrap procedure similar to the one advocated by Kosowski et al. (2006) (Table 4, last lines of the panels). We reach similar conclusions[16]. For CAPM and FH, although we do not find statistical significance at the 5% test level, differences between bottom and top deciles are always positive. Repeating this analysis with either KJ or the POT, we do not observe such clear-cut evidence along each decile.

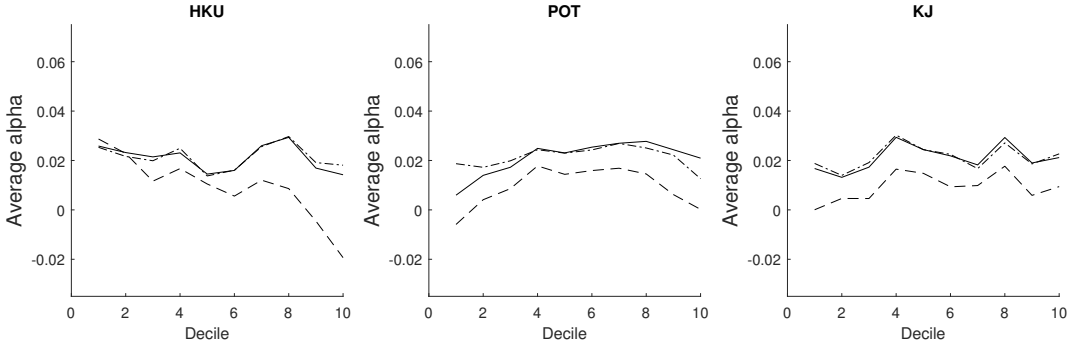


Figure 13: Average annualized $\hat{\alpha}_i^*$ for deciles of funds sorted on their $\hat{\gamma}_i$, obtained with the three different tail risk measures. Solid: CAPM alpha. Dashed-dotted: FH alpha. Dashed: JKKT alpha. Reported confidence intervals are the ones of Welch's t-tests.

Decomposing the performance per investment strategy (Figure 14), we find again a pos-

[16]The bootstrap confidence intervals for the other deciles are reported in the Supplement, Section 3.

itive difference for 11 out of 12 combinations asset-pricing model/investment strategy when using HKU (dark grey bars). This difference is found to be significantly positive in 8 combinations and is particularly large for the JKKT model, and the strategies *CTA/Managed futures* and *Macro*. For the POT and KJ, the patterns are more erratic across strategies and asset pricing models, with differences found significantly positive two times each, and negative 3 and 8 times, respectively. A reverse strategy with the POT would prove slightly more successful than HKU for *Fixed income* and equivalent for *Macro*, but overall inferior for *Long/Short Equity* and *CTA/Managed Futures*. These results further support the predictive potential of tail risk sensitivity obtained with the WMLE to select funds exhibiting abnormal returns, especially under the JKKT model.

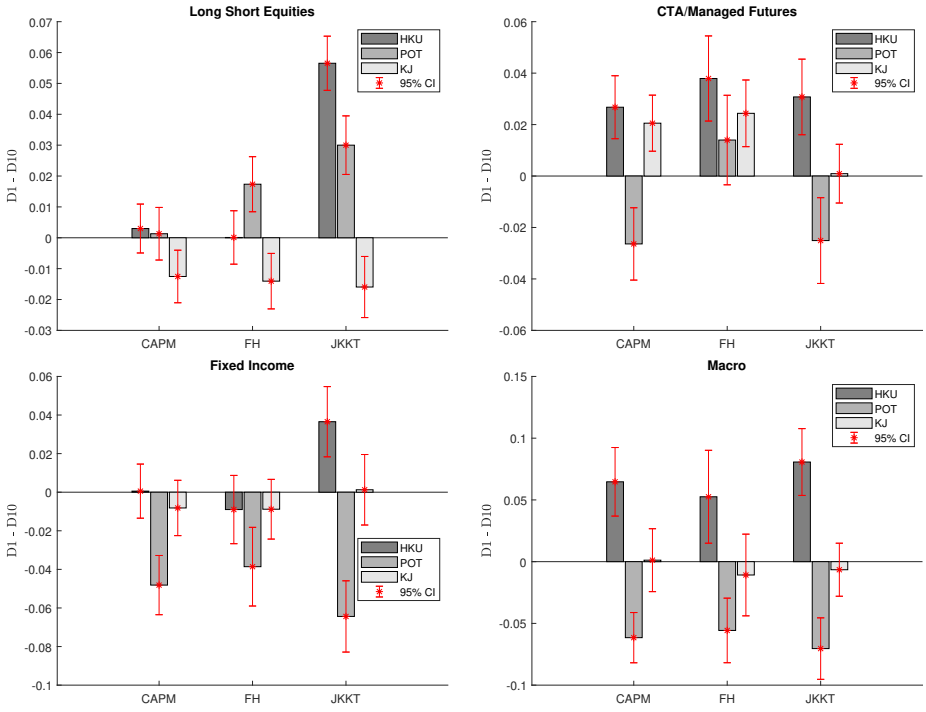


Figure 14: Differences in average alpha between first (D1) and last (D10) deciles of funds sorted on their $\hat{\gamma}_i$ computed with the different tail risk estimates: HKU (dark grey), POT (grey), KJ (light grey), and split along investment strategies.

We can conclude from this analysis that our tail risk measure can provide new insights on hedge fund performance and selection of high-alpha funds (although a formal out-of-sample assessment of this ability is left to further research). While the results are rather clear for

HKU, we cannot reach similar conclusions for KJ and POT, probably owing to their higher estimation uncertainty. These results highlight therefore once more the usefulness of an efficient estimation method for tail risk inference.

5 Conclusion

Measuring tail risk and identifying its determinants is a challenging task when studying hedge funds, due to their low reporting frequency. In particular, approaches based on the peaks-over-threshold (POT) discard an excessive portion of the data, rendering classical estimation techniques inefficient.

Our primary contribution is to develop a method that estimates the tail regression model simultaneously to the threshold parameter, drastically reducing estimation uncertainty. The proposed approach is based on an auxiliary splicing regression model, and uses the full sample to infer on the upper tail. To guard against misspecification issues in the opposite tail, we outline a censored maximum likelihood procedure that decreases the influence of these observations on our estimator. In an extensive simulation study, we demonstrate the superiority of this estimator over POT alternatives. Then, applying the proposed methodology to a database of 1,982 hedge funds, we identify several market factors in line with the funding liquidity theory of Brunnermeier and Pedersen (2009) and the limit to arbitrage of Shleifer and Vishny (1997) that are driving hedge funds tail risk. Contrasting our results with the POT and the approach of Kelly and Jiang (2014), we also find that our measure conveys greater explanatory power for abnormal excess returns of hedge funds.

In future work, in light of its demonstrated potential, we expect to explore the possibility to simultaneously estimate regression models for both tails, and to stress the out-of-sample predictive ability of our tail risk measure.

References

Agarwal, V.; Ruenzi, S., and Weigert, F. Tail risk in hedge funds: A unique view from portfolio holdings. *Journal of Financial Economics*, 125(3):610–636, 2017.

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

Avramov, D.; Barras, L., and Kosowski, R. Hedge fund return predictability under the magnifying glass. *Journal of Financial and Quantitative Analysis*, 48(4):1057–1083, 2013.

Babu, G.J. and Toreti, A. A goodness-of-fit test for heavy tailed distributions with unknown parameters and its application to simulated precipitation extremes in the euro-mediterranean region. *Journal of Statistical Planning and Inference*, 174: 11–19, 2016.

Bader, B.; Yan, J., and Zhang, X. Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *Annals of Applied Statistics*, 12(1):310–329, 2018.

Bali, T.G. An extreme value approach to estimating volatility and value at risk. *The Journal of Business*, 76(1):83–108, 2003.

Bali, T.G.; Gokcan, S., and Liang, B. Value-at-risk and the cross-section of hedge fund returns. *Journal of Banking & Finance*, 31(4):1135–1166, 2007.

Bali, T.G.; Brown, S.J., and Caglayan, M.O. Macroeconomic risk and hedge fund returns. *Journal of Financial Economics*, 114(1):1–19, 2014.

Balkema, A.A. and de Haan, L. Residual Life Time at Great Age. *The Annals of Probability*, 2(5):792 – 804, 1974.

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

Bee, M.; Dupuis, D.J., and Trapin, L. Realized Peaks over Threshold: A Time-Varying Extreme Value Approach with High-Frequency-Based Measures. *Journal of Financial Econometrics*, 17(2):254–283, 2019.

Beirlant, J.; Joossens, E., and Segers, J. "Generalized Pareto Fit to the Society of Actuaries' Large Claims Database", Ana C. Cebrián, Michel Denuit, and Philippe Lambert, July 2003. *North American Actuarial Journal*, 8(2):108–111, 2004.

Billio, M.; Getmansky, M.; W. Lo, A.W., and Pelizzon, L. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559, 2012.

Brunnermeier, M.K. and Pedersen, L.H. Market Liquidity and Funding Liquidity. *The Review of Financial Studies*, 22(6): 2201–2238, 2009.

Carreau, J. and Bengio, Y. A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes*, 12:53–76, 2009.

Castro-Camilo, D.; Huser, R., and Rue, H. A spliced gamma-generalized pareto model for short-term extreme wind speed probabilistic forecasting. *Journal of Agricultural, Biological and Environmental Statistics*, 24:517–534, 2019.

Chavez-Demoulin, V. and Davison, A. C. Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (App. Stat.)*, 54(1):207–222, 2005.

Chavez-Demoulin, V.; Embrechts, P., and Hofert, M. An extreme value approach for modeling operational risk losses depending on covariates. *Journal of Risk and Insurance*, 83(3):735–776, 2016.

Coles, S. *An introduction to statistical modeling of extreme values*. Springer, London, 2001.

Cuesta-Albertos, J. A.; Matràn, C., and Mayo-Iscar, A. Robust estimation in the normal mixture model based on robust clustering. *Journal of the Royal Statistical Society: Series B (Stat. Method.)*, 70(4):779–802, 2008.

Dacorogna, M.; Debbabi, N., and Kratz, M. Building up cyber resilience by better grasping cyber risk: A new algorithm for modelling cyber complaints filed at the *gendarmerie nationale*. *European Journal of Operation Research*, 311(2):708–729, 2023.

Darolles, S. and Roussellet, G. Managing hedge fund liquidity risks. *Working paper*, pages 1–38, 2024. URL `https://dx.doi.org/10.2139/ssrn.5043919`.

Davison, A. C. and Smith, R. L. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Stat. Method.)*, 52(3):393–425, 1990.

Dawid, A.P.; Musio, M., and Ventura, A. Minimum scoring rule inference. *Scandinavian Journal of Statistics*, 43(1), 2016.

de Carvalho, M.; Pereira, S.; Pereira, P., and de Zea Bermudez, P. An extreme value bayesian lasso for the conditional left and right tails. *Journal of Agricultural, Biological and Environmental Statistics*, 27:222–239, 2022.

Debbabi, M.; Kratz, M., and Mboup, M. A self-calibrating method for heavy tailed data modelling. application in neuroscience and finance. *ESSEC Working Paper & arXiv1612.03974v2*, 2017.

Diks, C.; Panchenko, V., and van Dijk, D. Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics*, 163(2):215–230, 2011.

Dupuis, D.; Engelke, S., and Trapin, L. Modeling panels of extremes. *Annals of Applied Statistics*, 17(1):498–517, 2023.

Eastoe, E.F. and Tawn, J.A. Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society. Series C (App. Stat.)*, 58(1):25–45, 2009.

Einmahl, J.H.J. and He, Y. Extreme value estimation for heterogeneous data. *Journal of Business & Economic Statistics*, 41(1):255–269, 2023.

Einmahl, J.H.J.; Kiriliouk, A.; Krajina, A., and Segers, J. An M-estimator of spatial tail dependence. *Journal of the Royal Statistical Society. Series B (Stat. Method.)*, 78(1):275–298, 2016.

Elyasiani, E. and Mansur, I. Hedge fund return, volatility asymmetry, and systemic effects: A higher-moment factor-egarch model. *Journal of Financial Stability*, 28:49–65, 2017.

Embrechts, P.; Klüppelberg, C., and Mikosch, T. *Modelling Extremal Events for Insurance and Finance*. Springer, 1997.

Fama, E.F. and French, K.R. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33 (1):3–56, 1993.

Frazzini, A. and Pedersen, L.H. Betting against beta. *Journal of Financial Economics*, 111(1):1–25, 2014.

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

Fung, W.; Hsieh, D.A.; Naik, N.Y., and Ramadorai, T. Hedge funds: Performance, risk,and capital formation. *Journal of Finance*, 63(4):1777–1803, 2008.
Getmansky, M.; Lo, A.W., and Makarov, I. An econometric model of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics*, 74(3):529–609, 2004.
Gupta, A. and Liang, B. Do hedge funds have enough capital? a value-at-risk approach. *Journal of Financial Economics*, 77(1):219–253, 2005.
Hambuckers, J.; Groll, A., and Kneib, T. Understanding the economic determinants of the severity of operational losses: A regularized generalized Pareto regression approach. *Journal of Applied Econometrics*, 33(6):898–935, 2018.
Hill, B. M. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3:1163–1174, 1975.
Huang, W.; Liu, Q.; Ghon Rhee, S., and Wu, F. Extreme downside risk and expected stock returns. *Journal of Banking & Finance*, 36(5):1492–1502, 2012.
Hüser, R. and Davison, A. C. Space–time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Stat. Method.)*, 76(2):439–461, 2014.
Jiang, H. and Kelly, B. Tail Risk and Hedge Fund Returns. *Chicago Booth Research Paper No. 12-44, Fama-Miller Working Paper*, November 2012.
Joenväärä, J.; Kauppila, M.; Kosowski, R., and Tolonen, P. Hedge fund performance: Are stylized facts sensitive to which database one uses? *Critical Finance Review*, 10:1–70, 2021.
Jorion, P. and Schwarz, C. The fix is in: Properly backing out backfill bias. *The Review of Financial Studies*, 32(12): 5048–5099, 2019.
Jun-Haeng Heo, J-H.; Shin, H.; Woosung Nam, W.; Om, J., and Jeong, C. Approximation of modified anderson–darling test statistics for extreme value distributions with unknown shape parameter. *Journal of Hydrology*, 499:41–49, 2013.
Kelly, B. and Jiang, H. Tail Risk and Asset Prices. *The Review of Financial Studies*, 27(10):2841–2871, 06 2014.
Kneib, T.; Silbersdorff, A., and Säfken, B. Rage against the mean – a review of distributional regression approaches. *Econometrics and Statistics*, 2021.
Koenker, R. and Bassett, G. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.
Kosowski, R.; Timmermann, A.; Wermers, R., and White, H. Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis. *The Journal of Finance*, 61(6):2551–2595, 2006.
Lan, Y.; Wang, N., and Yang, J. The economics of hedge funds. *Journal of Financial Economics*, 110(2):300–323, 2013.
Liu, X. and Mello, A.S. The fragile capital structure of hedge funds and the limits to arbitrage. *Journal of Financial Economics*, 102(3):491–506, 2011.
MacDonald, A.; Scarrott, C.J.; Lee, D.; Darlow, B.; Reale, M., and Russell, G. A flexible extreme value mixture model. *Computational Statistics Data Analysis*, 55(6):2137–2157, 2011.
Mason, D. Laws of large numbers for sums of extreme values. *Annals of Probability*, 10:754–764, 1982.
Massacci, D. Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. *Management Science*, 63(9):3072–3089, 2017.
McNeil, A.J.; Frey, R., and Embrechts, P. *Quantitative Risk Management: Concepts, Techniques and Tools - Revised Edition*. Princeton University Press, 2015.
Mhalla, L.; Hambuckers, J., and Lambert, M. Extremal connectedness of hedge funds. *Journal of Applied Econometrics*, 37 (5):987–1009, 2022.
Moskowitz, T.J.; Ooi, Y.H., and Pedersen, L.H. Time series momentum. *Journal of Financial Economics*, 104(2):228–250, 2012.
Naveau, P.; Huser, R.; Ribereau, P., and Hannart, A. Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52(4):2753–2769, 2016.
Pastor, L. and Stambaugh, R.F. Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3):642–685, 2003.
Patton, A.J. and Ramadorai, T. On the High-Frequency Dynamics of Hedge Fund Risk Exposures. *The Journal of Finance*, 68(2):597–635, 2013.
Pedersen, L.H. *Efficiently Inefficient: How Smart Money Invests and Market Prices Are Determined*. Princeton University Press, 2019.
Pickands, J. Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3(1):119 – 131, 1975.
Reynkens, T.; Verbelen, R.; Beirlant, J., and Antonio, K. Modelling censored losses using splicing: A global fit strategy with mixed erlang and extreme value distributions. *Insurance: Mathematics and Economics*, 77:65–77, 2017.
Sadka, R. Liquidity risk and the cross-section of hedge-fund returns. *Journal of Financial Economics*, 98(1):54–71, 2010.
Shleifer, A. and Vishny, R.W. The limits of arbitrage. 52(1):35–55, 1997.
Stulz, R.M. Hedge funds: Past, present, and future. *Journal of Economic Perspectives*, 21(2):175–194, 2007.
Tancredi, A.; Anderson, C., and O'Hagan, A. Accounting for threshold uncertainty in extreme value estimation. *Extremes*, 9:87–106, 2006.
Van der Vaart, A.W. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.
Wang, X.G. and Zidek, J.V. Selecting likelihood weights by cross-validation. *The Annals of Statistics*, 33(2):463 – 500, 2005.