Journal of Econometrics 235 (2023) 2266–2284

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Tail index estimation in the presence of covariates: Stock returns' tail risk dynamics☆

João Nicolau [a,b], Paulo M.M. Rodrigues [c,d], Marian Z. Stoykov [d,*]

[a] ISEG, Universidade de Lisboa, Portugal
[b] CEMAPRE, Portugal
[c] Banco de Portugal, Portugal
[d] Nova School of Business and Economics, Portugal

ARTICLE INFO

Article history:
Received 11 October 2021
Received in revised form 23 March 2023
Accepted 28 April 2023
Available online 2 June 2023

JEL classification:
C22
C58
G12

Keywords:
Extreme value theory
Pareto-type distributions
Tail index
Covariates information

ABSTRACT

This paper provides novel theoretical results for the estimation of the conditional tail index of Pareto and Pareto-type distributions in a time series context. We show that both the estimators and relevant test statistics are normally distributed in the limit, when independent and identically distributed or dependent data are considered. Simulation results provide support for the theoretical findings and highlight the good finite sample properties of the approach in a time series context. The proposed methodology is then used to analyse stock returns' tail risk dynamics. Two empirical applications are provided. The first consists in testing whether the time-varying tail exponents across firms follow Kelly and Jiang's (2014) assumption of common firm level tail dynamics. The results obtained from our sample seem not to favour this hypothesis. The second application, consists of the evaluation of the impact of two market risk indicators, *VIX* and *Expected Shortfall* (ES) and two firm specific covariates, *capitalization* and *market-to-book* on stocks tail risk dynamics. Although all variables seem important drivers of firms' tail risk dynamics, it is found that ES and firms' *capitalization* seem to have overall wider impact.

1. Introduction

Since the seminal works of Mandelbrot (1963) and Fama (1963) considerable evidence that unconditional return distributions are heavy tailed and suitably described by a power law has emerged. Heavy-tailed distributions place higher probability mass in their tails than the Gaussian distribution, and have been successfully used in fields, such as economics, finance, insurance, hydrology and many others.

The driving force behind the occurrence of extreme events in heavy tailed distributions is typically characterized by the tail index; see, for instance, Beirlant et al. (2004), Novak (2011) and Resnick (2007) for overviews. The tail index has been a useful indicator of the concentration of the probability mass in the tails, and the number of finite integer moments of a distribution. As such, over the years, considerable attention has been devoted to the tail index both in applied

☆ The authors thank two anonymous referees, an Associate Editor, and the Co-Editor (Torben Andersen) for their helpful and constructive feedback. Financial support from the Portuguese Foundation for Science and Technology (FCT) through projects CEMAPRE/REM - UIDB/05069/2020, PTDC/EGE-ECO/28924/2017, and (UID/ECO/00124/2013 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209) is also gratefully acknowledged.
* Corresponding author.
E-mail address: paulo.m.m.rodrigues@novasbe.pt (M.Z. Stoykov).

https://doi.org/10.1016/j.jeconom.2023.04.002

and theoretical research, and different procedures have been developed for its estimation. Such approaches include, among others, conditional maximum likelihood estimators (Hill, 1975), quantile–quantile plots (Kratz and Resnick, 1996), ordinary least squares-based log–log rank–size regressions (Rosen and Resnick, 1980, Gabaix, 1999, and Gabaix and Ibragimov, 2012), weighted least-squares (Beirlant et al., 1996) and kernel estimators (Csorgo et al., 1985), with the literature branching thereof, seeking to provide improvements over existing procedures. Challenging problems in the area still remain as discussed by Gabaix and Ibragimov (2012), with the works of Einmahl et al. (2016) and Nicolau and Rodrigues (2019) mitigating some of them by proposing different estimators with better finite-sample properties.

Estimation of the tail index in the presence of covariate information has been discussed by Beirlant and Goegebeur (2003). They model the tail index as a function of explanatory variables, which translates into measuring the heaviness of the conditional distribution of the dependent variable given covariate information. They obtain parameter estimates by profile maximum likelihood estimation and the asymptotic properties of their regression estimators are discussed by Wang and Tsai (2009). The latter authors establish the consistency of the estimators and show that, in the limit they follow a multivariate normal distribution with a non-zero mean vector. The aforementioned authors only consider the case in which the pair $(y_t, \mathbf{X}_t)$ is independently and identically distributed (i.i.d.). This is, however, a restrictive assumption which hinders application to time-series data which is typically characterized by strong dependence. One of the contributions of this paper is to relax the i.i.d. assumption in order to validate the application of the approach in a time series context. We study a setup similar to the one considered by Wang and Tsai (2009), but consider i.i.d. and dependent samples where the dependence structure is assumed to be of the strong mixing type. In both cases, we demonstrate multivariate asymptotic normality with a zero-mean vector. This allows for the construction of hypothesis tests with critical values taken from the standard normal distribution. Results of an in depth Monte Carlo analysis support the theoretical findings in this paper.

From an empirical perspective, the new results proposed in this paper are useful tools for the analysis of a multitude of research questions. One important topic relates to the dynamic nature of firms' tail risk. For instance, Gabaix (2012) and Wachter (2013) showed that allowing for time-varying tail risk in standard equilibrium based asset pricing models may help explain the apparent excess volatility of aggregate equity index returns. Bollerslev and Todorov (2014) introduced flexible estimation approaches that explicitly allow for the possibility of time-varying tails for large jump moves. They find that the magnitude of the left jump tail associated with extreme market declines far exceeds that of the right jump tail corresponding to large market appreciations. They further find non-trivial predictable temporal dependencies in the tail index parameters characterizing the decay in both tails. Kelly and Jiang (2014) reinforce the idea of time varying tail risks, and further argue that temporal variation in the tail parameters may help understand aggregate market returns as well as cross-sectional differences in average returns.

However, one of the main difficulties of this type of analysis is finding suitable measures of tail risk (see e.g. Andersen et al., 2015a,b, and Christoffersen et al., 2020). Since computing a measure of aggregate tail risk dynamics from a single time series is infeasible because of the infrequent nature of extreme events, Kelly and Jiang (2014) propose a panel estimation approach that captures common variation in the tail risks of individual firms.[1] Their argument for the use of this approach is that if firm-level tail distributions possess common dynamics, then cross-sections of crash events can be used to estimate the common component of their stocks tail risk at each point in time.

We also provide two empirical applications that focus on the tail risk dynamics of stock returns as an additional contribution of this paper. The first consists in using the approach developed in this paper to test Kelly and Jiang's (2014) assumption that the time-varying tail exponents across firms have common firm level tail dynamics. The results obtained from the sample of daily data used in our analysis seem not to concur with this hypothesis. Although market risk is an important driver of firms' tail risk its impact is heterogeneous across firms and sectors. The second application, focuses on the evaluation of the impact of two market risk indicators, *VIX* and *Expected Shortfall* (ES), and of two firm specific covariates, *capitalization* and *market-to-book*, on stocks tail risk dynamics. It is found that, overall, all four covariates used are important, but that ES and firms' *capitalization* seem to display statistical significance for a larger group of firms.

The remainder of the paper is organized as follows. Section 2 describes the setup in detail under a pure Pareto distribution and provides the theoretical properties of the maximum likelihood estimator (MLE). Section 3 extends the setup to Pareto-type distributions and an approximate MLE is obtained. This section also discusses the asymptotic properties of the estimators derived. Section 4 provides an in depth Monte Carlo study of the finite sample performance of the procedures. Section 5 discusses the results of the two empirical applications on the tail risk dynamics of firms' stock returns, and Section 6 concludes the paper. Finally, an Appendix collects detailed proofs of the results presented throughout the paper as well as additional Monte Carlo results.

The following notation has been used throughout the paper: letters in bold are used to denote vectors and matrices; and the symbols $\xrightarrow{a.s.}$ and $\xrightarrow{d}$ denote almost sure convergence and convergence in distribution, respectively.

[1] Literature on the analysis of tail features with panel data sets is still limited. Interesting contributions include Hill and Shneyerov (2013), Lee et al. (2021) and Sasaki and Wang (2022).

2. The Pareto model

Consider the time series $(Y_t, \mathbf{X}_t)$, $t = 1, \ldots, n$, where $Y_t \in \mathbb{R}$ is the response variable and $\mathbf{X}_t = (X_{1t}, \ldots, X_{pt}) \in \mathbb{R}^p$ is a p-dimensional vector of explanatory variables. Let $F_{Y_t|\mathbf{X}_t, Y_t > w_n}(y|\mathbf{x}_t, Y_t > w_n) \equiv F(y|\mathbf{x}_t, w_n) = P(Y_t \leq y|\mathbf{X}_t = \mathbf{x}_t, Y_t > w_n)$ be the cumulative distribution function (CDF) of Y_t conditional on $\mathbf{X}_t = \mathbf{x}_t$ and $Y_t > w_n$, where $w_n \in \mathbb{R}$ and possibly depends on n. The threshold w_n is typically positive. If interest lies in analysing extreme events in the left-tail or in both tails, the response variable can simply be multiplied by minus one or its absolute value can be considered, respectively.

The threshold w_n plays an important role in Pareto-type random variables. Given a specific assumption on the probability density function of the variable under analysis which we discuss in the next section, for values greater than w_n, a large class of useful distributions have densities that become similar to each other. In the limit, as w_n grows large, they collapse to the pure Pareto case, which we analyse in this section. In the pure Pareto case, w_n is taken as a constant. In fact, in this case, the threshold w_n can either be constant or increase with the sample size but that is not a problem — the results in this case are always exact due to the nature of the pure Pareto density. The choice of w_n is, however, of importance in the Pareto-type case and in Section 4 where numerical simulations are conducted, we demonstrate that a constant threshold performs adequately. The benefit of this construction is that it allows us to analyse a general class of distributions with the same approach. This is also well-suited to handle financial and economic data which are often realizations of random variables with densities that are skewed and leptokurtic (heavy-tailed); see e.g. Nicolau and Rodrigues (2019) and many of the references therein.

Going back to the model, assume that the survival function of Y_t conditional on $\mathbf{X}_t = \mathbf{x}_t$ and $Y_t > w_n$, $\bar{F}(y_t|\mathbf{x}_t, w_n)$, is governed by a Pareto distribution, *viz.*,

$$\bar{F}(y|\mathbf{x}_t, w_n) = 1 - F(y|\mathbf{x}_t, w_n) = (y/w_n)^{-\alpha(\mathbf{x}_t, \boldsymbol{\beta})}, \qquad y \geq w_n > 1, \tag{2.1}$$

where $\ln(\alpha(\mathbf{X}_t, \boldsymbol{\beta})) = \mathbf{X}_t\boldsymbol{\beta}$, with $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)' \in \mathbb{B} \subset \mathbb{R}^p$, $\mathbb{B}$ is compact and the true parameter vector $\boldsymbol{\beta}_0$ is an interior point of $\mathbb{B}$. The probability density of Y_t conditional on $\mathbf{X}_t = \mathbf{x}_t$ and $Y_t > w_n$ is,

$$f(y|\mathbf{x}_t, w_n, \boldsymbol{\beta}) = \alpha(\mathbf{x}_t, \boldsymbol{\beta})(y/w_n)^{-\alpha(\mathbf{x}_t, \boldsymbol{\beta})-1} w_n^{-1}, \qquad y \geq w_n > 1. \tag{2.2}$$

The assumed functional form of the tail index ensures that it is positive, which is a requirement for the Pareto distribution to be defined. The exponential function is smooth in its arguments which helps with optimization. A similar form of the link function is, for instance, used in Poisson regressions where the conditional expectation of the dependent variable is assumed to be of the same form to ensure that it is positive. Implicitly, a vector of ones is included as part of the explanatory variables in the specification to capture any constant effects.

2.1. The iid case

The interest of our analysis resides in the estimation of the unknown parameter vector $\boldsymbol{\beta}_0$, and on inference and hypotheses testing. We start our asymptotic characterization by considering the i.i.d. case first. The assumption that follows is sufficient for the estimation of $\boldsymbol{\beta}_0$.

Assumption 1. We assume that:

1. The sequence $\{(Y_t, \mathbf{X}_t)\}$ is i.i.d;
2. The joint support of $\mathbf{X}_t$ is finite;
3. The variance–covariance matrix $\mathrm{E}(\mathbf{X}_t'\mathbf{X}_t)$ is positive definite.

Under Assumption 1, we are able to consistently estimate the vector of parameters and obtain a non-degenerate asymptotic distribution, as demonstrated in the next theorem. This is the simplest case in which we are assuming that the observations of the marginal distributions are i.i.d. This condition will be extended later to accommodate weakly dependent sequences. The requirement on the support is not restrictive and allows us to demonstrate the existence of certain expectations of functions of the variables, which are needed for the theory to hold. The last requirement is the usual assumption of non-singularity which rules out full linear dependence between covariates.

Let $n_0 := \sum_{t=1}^{n} \mathbb{I}(Y_t > w_n)$ and $\ell_n(\boldsymbol{\beta})$ denote the average of the log-likelihood function of Y_t conditional on $\boldsymbol{X}_t$ and $Y_t > w_n$ ignoring, without loss of generality, the $-\ln(w_n)$ term. Thus, conditional on $Y_t > w_n$, and under Assumption 1, it follows that,

$$\ell_n(\boldsymbol{\beta}) = \frac{1}{n_0}\sum_{t=1}^{n_0}(\mathbf{X}_t\boldsymbol{\beta} - (\exp(\mathbf{X}_t\boldsymbol{\beta}) + 1)\ln(Y_t/w_n)). \tag{2.3}$$

The gradient, $G_n(\boldsymbol{\beta})$, of the log-likelihood function in (2.3) is,

$$G_n(\boldsymbol{\beta}) \equiv \frac{\partial \ell_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{1}{n_0}\sum_{t=1}^{n_0}\mathbf{X}_t'(1 - \exp(\mathbf{X}_t\boldsymbol{\beta})\ln(Y_t/w_n)), \quad \text{for } Y_t > w_n \tag{2.4}$$

and the Hessian $H_n(\boldsymbol{\beta})$,

$$H_n(\boldsymbol{\beta}) \equiv \frac{\partial^2 \ell_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\frac{1}{n_0} \sum_{t=1}^{n_0} \mathbf{X}_t' \mathbf{X}_t \exp(\mathbf{X}_t \boldsymbol{\beta}) \ln(Y_t / w_n), \quad \text{for } Y_t > w_n. \tag{2.5}$$

It is clear that $H_n(\boldsymbol{\beta})$ is negative definite. Therefore, the solution to $G_n(\boldsymbol{b}) = 0$, where $\boldsymbol{b}$ is a vector of parameter estimates (if they exist), corresponds to a global maximum. In general, a closed-form solution to the maximization problem does not exist unless $\mathbf{X}_t$ is one dimensional and constant, which leads to the Hill estimator (Hill, 1975). Hence, obtaining $\boldsymbol{b}$ is possible only via numerical methods. Nevertheless, the gradient vector and Hessian matrix are easy to derive and have simple forms, as shown above. Both of them can be utilized in numerical methods, which take advantage of them, to reduce computational time in the optimization process.

The above specification is slightly different from the one in Wang and Tsai (2009). Their log-likelihood function is[2]

$$\mathcal{K}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ \exp(\mathbf{X}_i^\top \boldsymbol{\theta} \log(Y_i/\omega_n) - \mathbf{X}_i^\top \boldsymbol{\theta}) \right\} \boldsymbol{I}(Y_i > \omega_n), \tag{2.6}$$

where $\boldsymbol{\theta}_0$ is the true parameter vector to be estimated. In their proofs, they work with n observations which leads to some technical complications arising with respect to the ratio n_0/n, which they demonstrate convergence in probability properties for, but treat it as non-random when taking certain expectations. In our case we condition on being in the tail. This has the advantage of simplifying some of the analysis by avoiding some of these complications.

We are now ready to state the first result. Considering (2.3), (2.4) and (2.5) the following Theorem summarizes the asymptotic behaviour for the i.i.d. case.

Theorem 1. *Given the survival function $\bar{F}(y_t|\mathbf{x}_t, w_n)$ in (2.1), under Assumption 1 as $n \to \infty$ (and $n_0 \to \infty$) it follows that,*

(a) $n_0^{1/2} \boldsymbol{\Sigma}^{-1/2} G_n(\boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_p)$;
(b) $\boldsymbol{\Sigma}^{-1/2} H_n(\boldsymbol{\beta}_0) \boldsymbol{\Sigma}^{-1/2} \xrightarrow{a.s.} -\mathbf{I}_p$;
(c) $\boldsymbol{\Sigma}^{1/2} n_0^{1/2} (\boldsymbol{b} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I}_p)$,

where $\boldsymbol{\Sigma} = \mathrm{E}(\mathbf{X}_t' \mathbf{X}_t | Y_t > w_n)$.

There are a few things to note regarding these results. Firstly, the effective convergence rate is being reduced — it is given by n_0, the number of observations in the tail, which is never greater than n (the full sample size). Even if the ratio $n_0/n \to c \in (0, 1)$, the convergence rate is reduced by a constant factor.[3] Secondly, in result *b*), we demonstrate almost sure convergence which implies converge in probability. Furthermore, the almost sure convergence is to a constant matrix. This is possible only under identically distributed variables. This will not be the case in the weakly dependent case with heterogeneously distributed data. The result in (c) establishes the consistency and joint normality of the MLE in the i.i.d. case. Similarly to (b), since we have assumed i.i.d data, the variance of the gradient is identical to the negative of the expectation of the Hessian. As a result, we can estimate $\boldsymbol{\Sigma}$ by substituting $\boldsymbol{b}$ for $\boldsymbol{\beta}$ in either $G_n(\boldsymbol{\beta})$ or $H_n(\boldsymbol{\beta})$. In the first instance, we would attempt to estimate the variance of the score, and in the second the negative of the expectation of the Hessian. Both procedures lead to the same result asymptotically. Thus, it follows that hypotheses testing can be conducted via standard normal critical values and we can test whether a particular explanatory variable is statistically significant in explaining the tail index. With this, we can construct the time-varying tail index as $\hat{\alpha}_t = \exp(\mathbf{X}_t \boldsymbol{b})$ and study its behaviour over time.

2.2. The weakly dependence case

We now focus on the asymptotic properties of $\boldsymbol{b}$ under weak dependence. The following assumption is sufficient.

Assumption 2. We assume that:

1. The sequence $\{(Y_t, \mathbf{X}_t)\}$ is strongly mixing of size $-r/(r-2)$, $r > 2$;
2. The joint support of $\mathbf{X}_t$ is finite;
3. The long-run variance–covariance matrix $Var\left(n_0^{-1/2} \sum_{t=1}^{n_0} G_n(\boldsymbol{\beta}_0)\right)$ is uniformly positive definite.

Assumption 2 differs from Assumption 1 in that we allow for sequences that are non-i.i.d which widens the applicability of the setup to dependent data. The size of the mixing process is standard in the literature. We still require that the joint support of $\mathbf{X}_t$ is finite for the purpose of demonstrating the existence of certain moments, but this is not restrictive. Assumption 2 permits the usage of (non-stationary) time-varying explanatory variables in the analysis. We also require

[2] In their notation.

[3] We thank a referee for pointing this out.

the technical condition on the uniform positivity of the long-run variance, without which the asymptotic results that follow would not hold.

Let us consider the setup in economic terms. If we take extreme events to mean times of economic crisis, we can argue that variables would not be realizations from distributions that are identical to those during non-crisis periods, or that the crisis periods are homogeneous. Hence, we need to allow for such behaviour, and Assumption 2 allows us to do that. We only require that over time they become "relatively" independent as governed by the mixing condition.

The following theorem summarizes the asymptotic behaviour of $\boldsymbol{b}$ under weak dependence.

Theorem 2. *Let Assumption 2 hold. Then, as $n \to \infty$ (and $n_0 \to \infty$) it follows that,*

(a) $\mathbb{G}_n(\boldsymbol{\beta}_0)^{-1/2} n_0^{1/2} G_n(\boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_p)$;
(b) $H_n(\boldsymbol{\beta}_0) - \mathbb{H}_n(\boldsymbol{\beta}_0) \xrightarrow{a.s.} 0$;
(c) $\left(\mathbb{H}_n(\boldsymbol{\beta}_0)^{-1}\mathbb{G}_n(\boldsymbol{\beta}_0)\mathbb{H}_n(\boldsymbol{\beta}_0)^{-1}\right)^{-1/2} n_0^{1/2}(\boldsymbol{b} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_p)$,

where $\mathbb{G}_n(\boldsymbol{\beta}_0) = \lim_{n\to\infty} Var(n_0^{-1/2} \sum_{t=1}^{n_0} G_n(\boldsymbol{\beta}_0))$, $\mathbb{H}_n = \lim_{n\to\infty} n_0^{-1} \mathrm{E}(H_n(\boldsymbol{\beta}_0))$, $H_n(\boldsymbol{\beta}_0)$ is the Hessian matrix and $\boldsymbol{b}$ the vector of maximum likelihood parameter estimates.

There are a few points to note. Firstly, misspecifying the likelihood function in terms of the dependence structure is not a problem — we are still able to consistently estimate the parameter of interest. Secondly, as per the discussion following Theorem 1, the result in (b) demonstrates a non-constant convergence in the almost sure sense. This is due to the fact that the variables are assumed to be heterogeneous. Thirdly, in (c), we observe the usual "sandwich" normalization in MLE in misspecified models. The matrices $\mathbb{H}_n(\boldsymbol{\beta}_0)$ and $\mathbb{G}_n(\boldsymbol{\beta}_0)$ are the expectation of the Hessian and the long-run variance of the score, respectively. Since $\boldsymbol{b}$ is a consistent estimator, we can replace it for $\boldsymbol{\beta}_0$ in the Hessian to obtain an estimate of $\mathbb{H}_n(\boldsymbol{\beta}_0)$. Since the reciprocal is a continuous function, we have that the difference between $H_n(\boldsymbol{b})^{-1}$ and $\mathbb{H}_n(\boldsymbol{\beta}_0)^{-1}$ converges to zero by virtue of the continuous mapping theorem. Dealing with the long-run variance–covariance matrix requires more effort. There are two main approaches to the problem. The first one is the VAR heteroskedasticity and autocorrelation consistent procedure proposed by Haan and Levin (1996). The idea behind this approach is to fit a finite-order VAR to the series and then construct the long-run variance–covariance matrix implied by the estimated VAR. It consists of two steps. In the first step, the lag length is selected for each VAR equation, and in the second step, the long-run variance is estimated by the implied VAR (see e.g. Hayashi, 2000, pp. 410–412 for details). The second approach, which is the one taken in this paper, is based on using a kernel. The idea is to estimate a number of covariances, which possibly grows large, and then weigh those appropriately via a kernel, as in e.g. Newey and West (1987). The estimator of $\mathbb{G}_n(\boldsymbol{\beta}_0)$ is then given by

$$\hat{\mathbb{G}}_n(\boldsymbol{\beta}_0) = \hat{\Gamma}_0 + \sum_{j=1}^{q(n_0)-1} k\left(\frac{j}{q(n_0)}\right)\left(\hat{\Gamma}_j + \hat{\Gamma}_j'\right); \tag{2.7}$$

$$k(x) = \begin{cases} 1 - |x|, & \text{for } |x| \le 1, \\ 0, & \text{for } |x| > 1; \end{cases} \tag{2.8}$$

$$\hat{\Gamma}_j = \frac{1}{n_0 - j} \sum_{t=j+1}^{n_0} \hat{\mathbf{g}}_t \hat{\mathbf{g}}_{t-j}'; \tag{2.9}$$

$$\hat{\mathbf{g}}_t = \mathbf{X}_t'(1 - \exp(\mathbf{X}_t\boldsymbol{\beta})\ln(Y_t/w_n)), \quad \text{for } Y_t > w_n, \tag{2.10}$$

where $k(x)$, in this case, corresponds to the Bartlett kernel. The choice of the bandwidth $q(n_0)$ is set at $q(n_0) = \lfloor n_0^{1/3} \rfloor$. There are other more sophisticated choices of kernels and bandwidths, such as, for instance, the ones suggested by Andrews and Monahan (1992) (see, *inter alia*, also Kiefer and Vogelsang, 2005, Müller, 2007, Lazarus et al., 2018 and references therein for further interesting contributions and discussions on HAC estimation). Since in the numerical section the procedure described in (2.7)–(2.10) displayed adequate performance we also used it in the empirical analyses in Section 5.

3. Pareto-type model

The assumption that the response variable follows an exact Pareto distribution may be restrictive in empirical applications. Therefore, the framework considered in this section, will allow for a broader class of heavy tailed distributions. Specifically, consider the survival function,

$$\bar{F}(y|\mathbf{x}_t, w_n) = 1 - F(y|\mathbf{x}_t, w_n) = (y/w_n)^{-\alpha(\mathbf{x}_t,\boldsymbol{\beta})}\mathcal{L}(y_t|\mathbf{x}_t), \tag{3.1}$$

where $\mathcal{L}(y|\mathbf{x}_t)$ is a slowly varying function at infinity, i.e., for any $w_n > 0$ we have that $\mathcal{L}(yw_n|\mathbf{x}_t)/\mathcal{L}(y|\mathbf{x}_t) \to 1$ as $y \to \infty$. In particular, we assume that,

$$\mathcal{L}(y|\mathbf{x}_t) = \gamma_0(\mathbf{x}_t) + \gamma_1(\mathbf{x}_t)y^{-\zeta(\mathbf{x}_t)} + o\left(y^{-\zeta(\mathbf{x}_t)}\right), \tag{3.2}$$

where $\gamma_0(\mathbf{x}_t)$ and $\gamma_1(\mathbf{x}_t)$ are functions in $\mathbf{x}_t$, with $\gamma_0(\mathbf{x}_t) > 0$, $\zeta(\cdot)$ is a positive function and the remainder term $o\left(y^{-\delta(\mathbf{x}_t)}\right)$ is of lower order than $y^{-\delta(\mathbf{x}_t)}$ (see e.g. Hall, 1982). It then follows that $\mathcal{L}(y|\mathbf{x}_t) \to \gamma_0(\mathbf{x}_t)$ and $\partial\mathcal{L}(y|\mathbf{x}_t)/\partial y \to 0$ as $y \to \infty$.

Specification (3.1) allows for a number of different distributions (up to a scaling factor) as special cases, such as, e.g. the Burr, the Student-t and the α-stable distribution. Importantly, as $y \to \infty$, specification (3.1) collapses to (2.1), the exact Pareto distribution up to a scaling factor. By introducing the tuning parameter w_n again and considering situations in which y grows large, the conditional probability density of Y_t can be approximated by $\alpha(\mathbf{x}_t, \boldsymbol{\beta})(y/w_n)^{-\alpha(\mathbf{x}_t,\boldsymbol{\beta})}y^{-1}$, up to a scaling factor, in view of the slowly varying function's properties. This implies that we can use the approximate log-likelihood function $\ell_n(\boldsymbol{\beta})$ up to an additive constant that is bounded and, thus, can be omitted. However, we also need to allow $w_n \to \infty$. Since we have omitted the $\ln(w_n)$ term from the log-likelihood function (as, for instance, in Wang and Tsai, 2009), we do not need to worry about identification issues. Furthermore, w_n needs to grow at a rate sufficient to guarantee that enough observations are available for estimation. For example, setting w_n equal to an empirical quantile that grows large would be sufficient. We note that since the analysis considered here is for large y, the second and above terms from the slowly varying function in (3.2) are of lower order and can be ignored.[4] From here, the following assumption is sufficient to demonstrate an asymptotic law.

Assumption 3. For all $1 \leq i, j \leq n_0$, let

$$\sigma_{ij} = \mathrm{E}\{(1 - \exp(\mathbf{X}_i\boldsymbol{\beta})\ln(Y_i/w_n))(1 - \exp(\mathbf{X}_j\boldsymbol{\beta})\ln(Y_j/w_n))|\mathbf{X}_i', \mathbf{X}_j, Y_i > w_n, Y_j > w_n\}.$$

We assume that the covariances σ_{ij} are independent of w_n, when w_n grows large, and that the expectations $\mathrm{E}\{|(1 - \exp(\mathbf{X}_i\boldsymbol{\beta})\ln(Y_i/w_n))(1 - \exp(\mathbf{X}_j\boldsymbol{\beta})\ln(Y_j/w_n))|^{r+\delta}|Y_i > w_n, Y_j > w_n\}$ exist.

Assumption 3 is needed to ensure that $\mathbb{G}_n$ and $\mathbb{H}_n$ are not degenerate as $n \to \infty$ (and $n_0, w_n \to \infty$) and that we can estimate them consistently. One could perhaps make an assumption on the joint behaviour of Y_i and Y_j that would imply the above condition, but that would be a stronger condition. For example, at the marginal level $\ln(Y_i/w_n)$, conditional on $Y_i > w_n$, is approximately exponentially distributed and, as such, its moments are independent of w_n. We require that to happen with the covariances of $\ln(Y_i/w_n)$ and $\ln(Y_j/w_n)$.

With this, we are ready to state the following theorem.

Theorem 3. *Let the data be generated as in (3.1). Then as $n \to \infty$ (and $n_0, w_n \to \infty$) it follows that,*

(a) under Assumption 1 the conclusions of Theorem 1 remain unchanged;
(b) under Assumptions 2 and 3 the conclusions of Theorem 2 remain unchanged.

Theorem 3 generalizes our previous results to processes which are Pareto-type. Thus, the approximate MLE is asymptotically normally distributed with a variance–covariance matrix which can be consistently estimated. It follows that one can construct confidence intervals which have the standard normal coverage. The generality of the results permits analysis of time series which are weakly dependent and are generated from Pareto-type distributions.

We conclude this section with a Corollary to Theorem 3 which follows from the continuous mapping theorem.

Corollary 1. *Let the data be generated as in (3.1). Then as $n \to \infty$ (and $n_0, w_n \to \infty$), it follows that,*

(a) under Assumption 1

$$n_0(\boldsymbol{b} - \boldsymbol{\beta}_0)'\boldsymbol{\Sigma}(\boldsymbol{b} - \boldsymbol{\beta}_0) \xrightarrow{d} \chi^2(p), \tag{3.3}$$

where $\boldsymbol{\Sigma}$ is defined in Theorem 1, and;

(b) under Assumptions 2 and 3

$$n_0(\boldsymbol{b} - \boldsymbol{\beta}_0)'\left\{\mathbb{H}_n(\boldsymbol{\beta}_0)^{-1}\mathbb{G}_n(\boldsymbol{\beta}_0)\mathbb{H}_n(\boldsymbol{\beta}_0)^{-1}\right\}^{-1}(\boldsymbol{b} - \boldsymbol{\beta}_0) \xrightarrow{d} \chi^2(p), \tag{3.4}$$

where $\mathbb{H}_n(\boldsymbol{\beta}_0)$ and $\mathbb{G}_n(\boldsymbol{\beta}_0)$ are defined in Theorem 2.

$\chi^2(p)$ denotes a chi-squared distribution with p degrees of freedom.

Corollary 1 provides a way to test joint significance. The respective unknown matrices can be estimated consistently with the procedures outlined in the discussions surrounding Theorems 1 and 2.

4. Numerical results

In this section we provide an in depth Monte Carlo analysis to assess the finite sample properties of the estimators and corresponding test statistics. To generate the data, we consider several heavy-tailed distributions, such as the Pareto as well as other distributions which satisfy (3.1), e.g. the Burr, the Student-t, and the α-stable[5] distribution, which are

[4] In the empirical illustration of Section 5.2 we detail how to empirically determine w_n. Specifically, we adopt a distance metric as in Wang and Tsai (2009), which is based on the minimum distance between the empirical distribution and a Pareto.

[5] Data from an alpha-stable distribution with index α is generated as $X_t = \frac{\sin(\alpha\eta_t)}{[\cos(\eta_t)]^{1/\alpha}}\left(\frac{\cos[(1-\alpha)\eta_t]}{z_t}\right)^{\frac{(1-\alpha)}{\alpha}}$ where η_t is uniform on $(-\pi/2, \pi/2)$ and z_t is an exponential variate with mean 1 (see Samorodnitsky and Taqqu, 2004).

frequently used in the extreme-value literature (see Beirlant et al., 2004). We begin by generating the tail index as $\ln(\alpha(\mathbf{X}_t, \boldsymbol{\beta}_0)) = \beta_1 + \beta_2 x_t$ from three different scenarios:

$$\text{(a)} \quad \xi_t = \rho\xi_{t-1} + \sigma\epsilon_t \tag{4.1}$$

$$\text{(b)} \quad \xi_t = \rho\xi_{t-1} + \cos(t+U)\sigma\epsilon_t \tag{4.2}$$

$$\text{(c)} \quad \xi_t = \sqrt{1 + 0.5\zeta_{t-1}^2}\sigma\epsilon_t \tag{4.3}$$

and applying the transformation

$$x_t = \sqrt{12}\left(\Phi\left(\xi_t\right) - 1/2\right), \tag{4.4}$$

where $\{\epsilon_t\}$ is a $N(0, 1)$ white noise process, the initial condition ξ_0 is a $N(0, 1)$ random variable, U is uniformly distributed on $[0, 1)$, and Φ refers to the standard normal distribution function. The transformation in (4.4) is considered in Wang and Tsai (2009) and it ensures that x_t has finite support. The three specifications in (4.1)–(4.3) aim to capture different dynamics in the explanatory variables. The first one in (4.1) is a simple weakly stationary autoregressive process. The second one in (4.2) is an autoregressive process with a non-constant variance. The uniform distribution has been added to mitigate cyclical behaviour that could occur due to the periodicity of the cosine function. Finally, the third specification in (4.3) is an ARCH(1) process. In what follows, we only report the outcomes for the specification in (4.1) - case (a). The results obtained when (4.2) is considered are qualitatively the same as those obtained when the dynamics of ξ_t is as in (4.1). The results for case (c) in (4.3) are also similar to those obtained with (4.1) in terms of the empirical size of the test, but power is approximately half. Hence, for the sake of space, the results for cases (b) and (c) are provided in the supplementary appendix (see Tables C.1–C.6). To construct the dependent variable we invoke the inverse transformation method. Let $\{u_t\}$ be a sequence of independent draws from a uniform distribution over (0, 1). Then

$$y_t|x_t = F^{-1}(u_t|x_t; \alpha(x_t, \boldsymbol{\beta}_0)), \tag{4.5}$$

where F is the desired cumulative distribution function. For example, to simulate from a conditional Pareto we generate $y_t|x_t = u_t^{-1/\alpha(x_t,\boldsymbol{\beta}_0)}$.

For the power analysis we set $\beta_1 = 0.5$, $\rho \in \{0, 0.1, \ldots, 0.9\}$, $\beta_2 \in \{0, 0.1, \ldots, 1\}$, and $\sigma \in \{0.1, 0.2\}$, and for the finite sample size we use $\beta_1 = 0.5$, $\beta_2 = 0$, $\rho \in \{0, 0.3, 0.5, 0.7, 0.9\}$ and $\sigma \in \{0.05, 0.1, 0.2\}$. For all experiments we also compute the estimation bias $\mathrm{E}(\hat{\beta}_2 - \beta_2)$ of $\hat{\beta}_2$. We generate samples of size n, with $n \in \{1000, 2000, 5000\}$, from heavy-tailed distributions (Pareto, Burr, Student-t and α-stable). In this section we follow Quintos et al. (2001), who have found that letting $\kappa = 0.1$ and choosing the upper $\lfloor \kappa n \rfloor$ ordered statistics for the application of the Hill estimator provides adequate performance. In our simulations we consider $\kappa = 0.1$ and $\kappa = 0.2$. Based on the generated data we estimate β_1 and β_2 (and consequently the conditional tail index $\alpha(\mathbf{X}_t, \boldsymbol{\beta})$) as described in Sections 2 and 3, from sub-samples of size $\lfloor \kappa n \rfloor$. In specific, we evaluate the performance of the one-sided, $t_{\beta_2}^+$, and the two-sided, t_{β_2}, t-tests on β_2, which consider $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 > 0$ and $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$, respectively. All results provided are based on 5000 replications.

4.1. Empirical size

Table 1 below and Tables B.1–B.3 in the supplementary appendix provide the finite sample size results for $t_{\beta_2}^+$ and t_{β_2} when data is generated from a Pareto, Burr, Student-t and α-stabled distribution, respectively. The results in Table 1 correspond to our benchmark case, as these refer to applications of the estimators and tests to data generated from a Pareto distribution. The results will also prove useful in corroborating the theory presented in Section 2. Tables B.1 - B.3 refer to results from the Pareto type distributions, which will be useful for validating the theoretical results of Section 3. From Table 1 we observe that the empirical rejection frequencies of the right-sided, $t_{\beta_2}^+$, and two-sided, t_{β_2}, statistics, are very close to the 5% nominal size considered, regardless of the values of σ and κ. The latter result, is expected, as the tail cut off point (i.e., κ) is of no importance in the Pareto case. We also observe from this table that the estimation bias of β_2 is quite small regardless of the value of $\rho \in \{0, \ldots, 0.9\}$ considered in the simulations. As expected, the bias also decreases as the sample size increases.

Regarding the results in Tables B.1–B.3 for the Pareto type distributions, when data is generated from a Burr distribution (Table B.1), the empirical size of the t-ratio is close to the nominal significance level of 5% for $\kappa = 0.1$, regardless of the values of $\rho \in \{0, \ldots, 0.9\}$ and $\sigma \in \{0.05, 0.1, 0.2\}$. However, for $\kappa = 0.2$ size decreases slightly (i.e., the t-tests become more conservative). This type of behaviour is however more marked for the Student-t distribution (Table B.2) and for the α-stable distribution (Table B.3). Note that for these latter two distributions, the empirical size is basically half of the nominal size when $\kappa = 0.2$, hence highlighting the importance of correctly choosing the tail cut off point in the case of Pareto type distributions. This observation has led us in the empirical application to use a data driven approach as in Wang and Tsai (2009).

Table 1
Empirical rejection frequencies of the right-sided, $t^+_{\beta_2}$, and two-sided, t_{β_2}, t-tests when $\beta_2 = 0$. DGP is case a). The tail index is generated as: $\ln(\alpha(\mathbf{X}_t, \boldsymbol{\beta}_0)) = \beta_1 + \beta_2 x_t$, where $x_t = \sqrt{12}\,(\Phi(\xi_t) - 1/2)$ and $\xi_t = \rho\xi_{t-1} + \sigma\epsilon_t$. $\{\epsilon_t\}$ is a $N(0,1)$ white noise process, the initial condition ξ_0 is a $N(0,1)$ random variable, U is uniformly distributed on $[0,1)$, and Φ refers to the standard normal distribution function. The response variable of interest is then generated as: $y_t|x_t = F^{-1}(u_t|x_t; \alpha(x_t, \boldsymbol{\beta}_0))$, where $\{u_t\}$ is a sequence of independent draws from a uniform distribution over $(0,1)$ and F is the Pareto distribution i.e. $y_t|x_t = u_t^{-1/\alpha(x_t,\boldsymbol{\beta}_0)}$.

ρ	σ	$n = 1000$			$n = 2000$			$n = 5000$		
		$E(\hat{\beta}_2 - \beta_2)$	$t^+_{\beta_2}$	t_{β_2}	$E(\hat{\beta}_2 - \beta_2)$	$t^+_{\beta_2}$	t_{β_2}	$E(\hat{\beta}_2 - \beta_2)$	$t^+_{\beta_2}$	t_{β_2}
$\kappa = 0.1$										
0	0.05	−0.016	0.053	0.056	−0.002	0.052	0.052	−0.002	0.050	0.050
0.3	0.05	0.006	0.054	0.056	−0.008	0.051	0.052	−0.001	0.052	0.051
0.5	0.05	−0.001	0.052	0.053	0.001	0.051	0.052	0.001	0.051	0.051
0.7	0.05	−0.007	0.052	0.053	0.007	0.052	0.051	−0.001	0.051	0.050
0.9	0.05	0.004	0.051	0.052	0.002	0.053	0.053	0.001	0.051	0.052
0	0.1	0.000	0.052	0.052	−0.001	0.050	0.051	0.000	0.051	0.052
0.3	0.1	0.002	0.053	0.055	0.001	0.052	0.052	0.000	0.050	0.050
0.5	0.1	0.002	0.053	0.054	0.001	0.051	0.049	0.000	0.050	0.050
0.7	0.1	0.002	0.052	0.052	−0.001	0.051	0.051	0.001	0.050	0.050
0.9	0.1	−0.002	0.052	0.053	0.000	0.051	0.050	0.001	0.051	0.052
0	0.2	0.002	0.052	0.054	0.001	0.052	0.052	0.000	0.048	0.048
0.3	0.2	0.003	0.052	0.051	−0.001	0.051	0.053	0.001	0.051	0.049
0.5	0.2	−0.002	0.051	0.054	−0.001	0.050	0.051	0.000	0.051	0.050
0.7	0.2	0.000	0.051	0.052	0.001	0.053	0.052	0.001	0.050	0.050
0.9	0.2	−0.001	0.051	0.052	0.000	0.050	0.052	0.000	0.051	0.051
$\kappa = 0.2$										
0	0.05	−0.001	0.051	0.052	−0.002	0.050	0.051	0.000	0.050	0.050
0.3	0.05	−0.001	0.050	0.052	−0.006	0.050	0.051	0.001	0.051	0.049
0.5	0.05	0.006	0.051	0.052	−0.003	0.048	0.050	0.002	0.051	0.050
0.7	0.05	−0.002	0.051	0.053	0.001	0.050	0.050	0.000	0.051	0.051
0.9	0.05	0.002	0.050	0.050	0.001	0.050	0.050	0.001	0.050	0.050
0	0.1	0.004	0.051	0.051	0.003	0.051	0.050	−0.001	0.050	0.050
0.3	0.1	0.001	0.051	0.052	0.000	0.048	0.050	0.000	0.049	0.051
0.5	0.1	0.000	0.051	0.051	0.000	0.052	0.051	−0.001	0.049	0.051
0.7	0.1	0.001	0.050	0.051	−0.001	0.049	0.050	−0.001	0.050	0.049
0.9	0.1	0.000	0.053	0.055	0.000	0.052	0.051	−0.001	0.049	0.050
0	0.2	0.000	0.051	0.052	0.000	0.051	0.050	0.000	0.050	0.050
0.3	0.2	−0.002	0.049	0.051	0.000	0.051	0.050	−0.001	0.049	0.050
0.5	0.2	0.001	0.052	0.050	−0.001	0.049	0.051	0.001	0.050	0.049
0.7	0.2	0.000	0.052	0.053	0.000	0.051	0.050	0.000	0.049	0.050
0.9	0.2	0.001	0.051	0.052	0.000	0.049	0.050	0.000	0.050	0.050

Note: The column labelled $E(\hat{\beta}_2 - \beta_2)$ presents the results of the estimation bias which is computed as $1/Q\sum_{s=1}^{Q}(\hat{\beta}_2 - \beta_2)$, where $Q = 5000$ is the number of Monte Carlo replications. κ is used to determine the tail cut-off point.

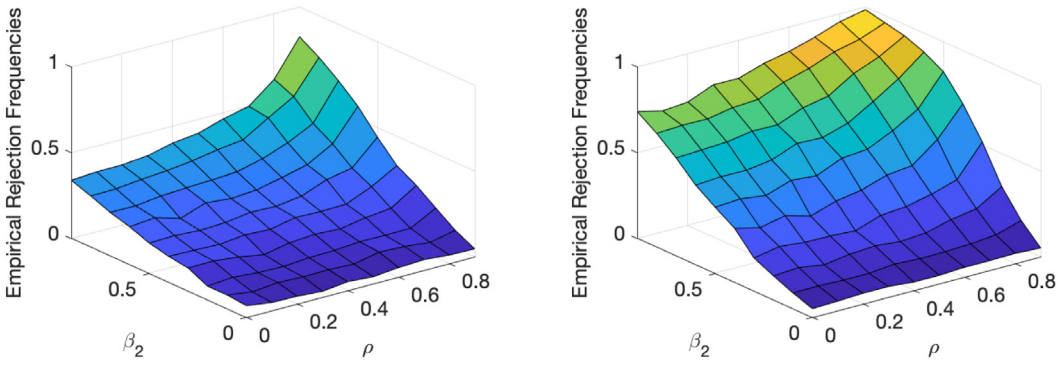
4.2. Empirical power

Fig. 1 and Figures B.1–B.3 in the supplementary appendix display the power surfaces of the $t^+_{\beta_2}$ test results when applied to data generated from a Pareto, Burr, Student-t and α-stable distribution, respectively, and $\kappa = 0.1$ (results for $\kappa = 0.2$ are provided in Figures C.1–C.3 of the supplementary appendix). The power surfaces are very similar across all four cases, highlighting the consistency of the tests regardless of whether the underlying distribution is Pareto or Pareto type. These Figures show that power increases as the sample size, β_2 and the persistence of x_t increases (i.e. as ρ increases).[6]

4.3. Joint testing

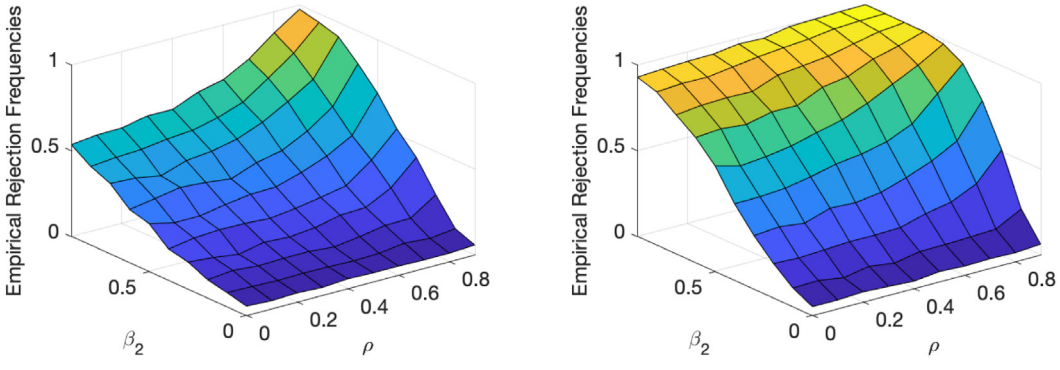
In Corollary 1 we derived a test for joint significance of the parameters. To assess its performance, we again use (4.4) but this time generate the tail index as $\ln(\alpha(\mathbf{X}_t, \boldsymbol{\beta}_0)) = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1}$, with $\beta_2 = 0$, $\kappa = 0.1$ and the rest of the specification as in the previous sub-section, but excluding the α-stable distribution. The choice of $\kappa = 0.1$ is motivated by the outcome of the study in sub- Section 4.1. The null hypothesis is $H_0 : \beta_2 = \beta_3 = 0$. In addition to providing insights into the performance of the test, this framework also allows studying a construction in a predictive context. The outcome is depicted in Figures B.4–B.5. Results are virtually identical to the ones from the single-variable testing (see Figure 2).

[6] Results for the two-sided test are not reported as the conclusions (power Figures) are qualitatively the same as for the one-sided tests, but these can be obtained from the authors.



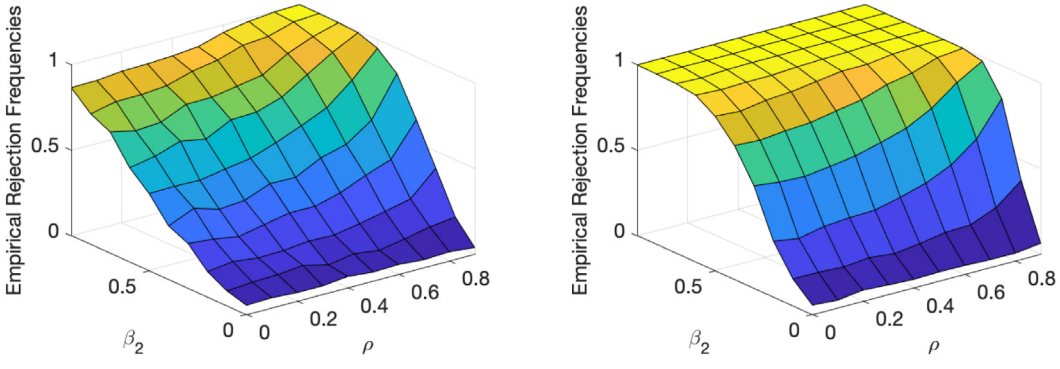
(a) $\sigma = 0.1$. (b) $\sigma = 0.2$.

$\kappa = 0.1,\ n = 1000$



(c) $\sigma = 0.1$. (d) $\sigma = 0.2$.

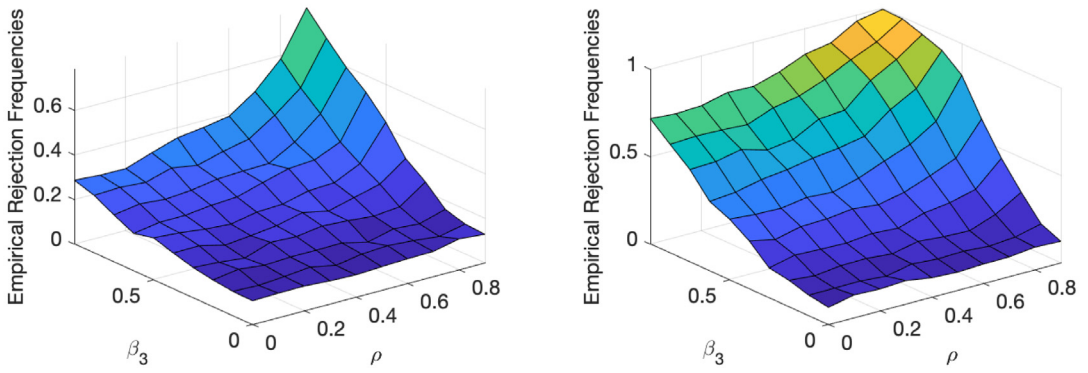
$\kappa = 0.1,\ n = 2000$



(e) $\sigma = 0.1$. (f) $\sigma = 0.2$.

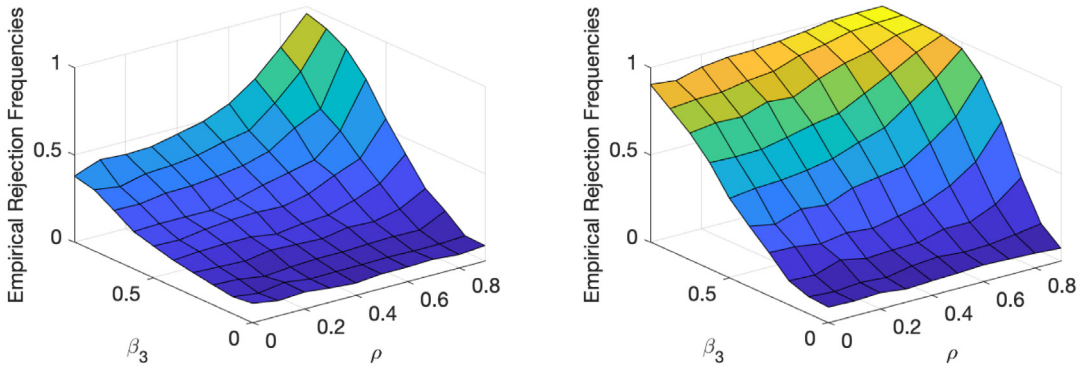
$\kappa = 0.1,\ n = 5000$

Fig. 1. Empirical rejection frequencies of the one-sided test, $t^{+}_{\beta_2}$. **DGP is case (a)**. The tail index is generated as: $\ln(\alpha(\mathbf{X}_t, \boldsymbol{\beta}_0)) = \beta_1 + \beta_2 x_t$, where $x_t = \sqrt{12}\,(\Phi(\xi_t) - 1/2)$ and $\xi_t = \rho\xi_{t-1} + \sigma\epsilon_t$. $\{\epsilon_t\}$ is a $N(0,1)$ white noise process, the initial condition ξ_0 is a $N(0,1)$ random variable, U is uniformly distributed on $[0,1)$, and Φ refers to the standard normal distribution function. The response variable of interest is then generated as: $y_t|x_t = F^{-1}(u_t|x_t; \alpha(x_t, \boldsymbol{\beta}_0))$, where $\{u_t\}$ is a sequence of independent draws from a uniform distribution over $(0,1)$ and F is the **Pareto distribution** i.e. $y_t|x_t = u_t^{-1/\alpha(x_t,\boldsymbol{\beta}_0)}$.



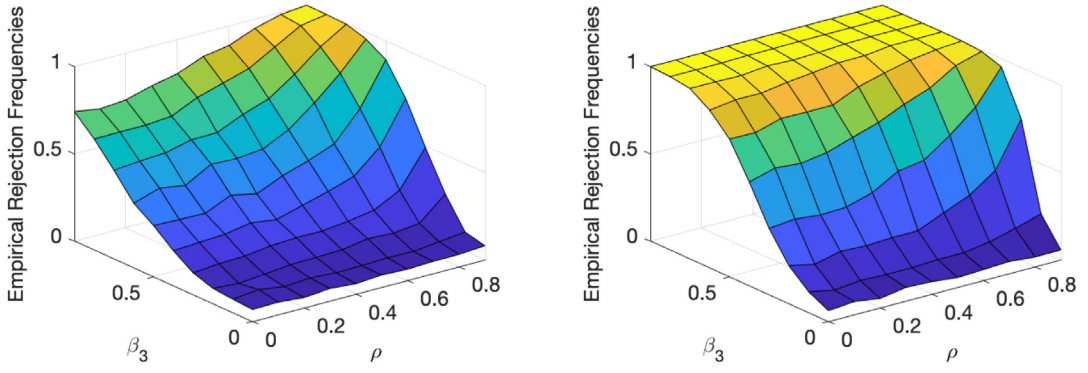
(a) $\sigma = 0.1$. (b) $\sigma = 0.2$.

$\kappa = 0.1,\ n = 1000$



(c) $\sigma = 0.1$. (d) $\sigma = 0.2$.

$\kappa = 0.1,\ n = 2000$



(e) $\sigma = 0.1$. (f) $\sigma = 0.2$.

$\kappa = 0.1,\ n = 5000$

Fig. 2. Empirical rejection frequencies of the joint test described in Corollary 1. **DGP is case (a)**. The response variable of interest is then generated as: $y_t|x_t = F^{-1}(u_t|x_t; \alpha(x_t, \boldsymbol{\beta}_0))$, where $\{u_t\}$ is a sequence of independent draws from a uniform distribution over $(0, 1)$ and F is the **Pareto distribution** i.e. $y_t|x_t = u_t^{-1/\alpha(x_t,\boldsymbol{\beta}_0)}$.

5. Empirical analysis

This section provides two empirical applications that focus on stock returns' tail risk dynamics in the US market. The data analysed correspond to stock return series of domestic US firms collected from the Refinitiv Eikon platform. Daily information on closing prices and respective sectors of activity on all available firms from 03/12/1999 to 04/08/2022 (for some cases the sample starts after 03/12/1999) is obtained. The final sample consists of $n = 4389$ firms. The firms were then classified using the Global Industry Classification Standard into eleven sectors: Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Communication Services, Utilities, and Real Estate.[7]

5.1. Testing Kelly and Jiang's hypothesis

Following Kelly and Jiang (2014), the time t left-tail distribution of stock i is defined as the set of return events falling below some extreme negative threshold w_t, i.e.,

$$P(r_{i,t+1} < r | r_{i,t+1} < w_t \text{ and } \mathcal{F}_t) = \left(\frac{r}{w_t}\right)^{-a_i \alpha_{Mt}}, \tag{5.1}$$

where $r < w_t < 0$ and $\mathcal{F}_t$ is the information set available up to time t. Hence, stock i's left-tail index in this framework is $\alpha_{it} = a_i \alpha_{Mt}$.[8] Since $r < w_t < 0$, $r/w_t > 1$ and $a_i \alpha_{Mt} > 0$, this ensures that $0 \leq \left(\frac{r}{w_t}\right)^{-\alpha_{Mt}} \leq 1$. Kelly and Jiang (2014) refer to $1/\alpha_{Mt}$ as the "tail risk" at time t. Note that (5.1) corresponds to a pure-Pareto setup, however in light of the results presented above (Sections 2 and 3) it can also be used without being altered in Pareto-type processes.

The identifying assumption for the estimation of α_{Mt} is that the tail risk of individual assets includes a common component which, according to Kelly and Jiang (2014) can be computed at each point in time from a cross-section of extreme events for individual assets. α_{Mt} is estimated based on the tail index estimator of Hill (1975) applied to the time t cross-section of firms' returns. For estimation of α_{Mt} we consider, as in Kelly and Jiang (2014), that w_t is the fifth percentile of the cross-section in each period.

One interesting hypothesis that we can test with our approach is whether the time-varying tail exponents across firms follow Kelly and Jiang's (2014) assumption of common firm level tail dynamics, i.e., $\alpha_{it} = a_i \alpha_{Mt}$. Applying a logarithmic transformation to $\alpha_{it} = a_i \alpha_{Mt}$ results in $\ln\alpha_{it} = \ln a_i + \ln\alpha_{Mt} = c_i + \ln\alpha_{Mt}$. The resulting linear representation can be seen as a particular case of,

$$\ln\alpha_{it} = c_i + \beta_i \ln\alpha_{Mt}, \tag{5.2}$$

in which $\beta_i = 1$. Hence, we will use our approach to estimate (5.2) and test the null hypothesis $H_0 : \beta_i = 1$ against the alternative $H_a : \beta_i \neq 1$.

The first step of our analysis consists of the determination of the tail threshold w_{in}, for which we use a discrepancy measure (a detailed discussion and illustration of this measure is provided in the next section). The analysis will be based on firms' excess returns and risk-adjusted returns (the latter are adjusted for the Fama–French three factors; see, e.g., Jiang et al., 2020).

To test the null hypothesis we use the fact that $\left\{\hat{\beta}_i; i = 1, 2, \ldots, n\right\}$ is a sequence of independently heterogeneously distributed observations and consider that,

$$\bar{\beta}_n := E\left(\frac{\sum_{i=1}^n \hat{\beta}_i}{n}\right)$$
$$\frac{\bar{\sigma}_n^2}{n} := Var\left(\frac{\sum_{i=1}^n \hat{\beta}_i}{n}\right).$$

Since $\hat{\beta}_i$ are consistently estimated, the null hypothesis implies that, $\bar{\beta}_n \to 1$. Based on the CLT for independently heterogeneously distributed observations in White (2001, p.117) it then follows that,

$$\frac{\sqrt{n}\left(\bar{\hat{\beta}} - \bar{\beta}_n\right)}{\bar{\sigma}_n} \xrightarrow{d} N(0, 1).$$

Thus, under the null hypothesis,

$$\frac{\sqrt{n}\left(\bar{\hat{\beta}} - 1\right)}{\bar{\sigma}_n} \xrightarrow{d} N(0, 1).$$

[7] For details see https://www.spglobal.com/marketintelligence/en/documents/112727-gics-mapbook_2018_v3_letter_digitalspreads.pdf.

[8] Note that in the notation of Kelly and Jiang (2014) $\alpha_{Mt} = 1/\lambda_t$.

Moreover,

$$\bar{\sigma}_n^2 = nVar\left(\bar{\hat{\beta}}\right) \Rightarrow \bar{\sigma}_n^2 = Var\left(\hat{\beta}_i\right),$$

and therefore the variance s^2 of estimated $\hat{\beta}_i$ can be used to set up the test statistic,

$$Z = \frac{\bar{\hat{\beta}} - 1}{s/\sqrt{n}} \xrightarrow{d} N(0, 1) \tag{5.3}$$

where $\bar{\hat{\beta}} = n^{-1}\sum_{i=1}^{n}\hat{\beta}_i$ and s is the standard deviation of $\hat{\beta}_i$. Applying (5.3) on the beta estimates, $\hat{\beta}_i, i = 1, \ldots, N$, obtained from all firms in our sample, produces a Z-test result of -69.7 (p-value $= 0.000$) corresponding to a rejection of the null. We also tested the same hypothesis in each sector (Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Communication Services, Utilities, and Real Estate) and, for all cases, the null hypothesis was rejected.

It is important to note that in Kelly and Jiang's notation, the Hill estimator of $\alpha_{Mt}^{-1} = \lambda_t$ converges to,

$$E_{t-1}\left[\frac{1}{K_t}\sum_{i=1}^{K_t} ln\frac{R_{i,t}}{u_t}|\lambda_t, R_{i,t} < u_t\right] = \frac{1}{K_t}\sum_{i=1}^{K_t}\frac{1}{a_i\alpha_{Mt}^{\beta_i}} \tag{5.4}$$

or more precisely, the difference between the two converges to zero due to the heterogeneity in i. Now, if $\beta_i = 1$, then we can estimate α_{Mt} times some constant. But we find that $\beta_i \neq 1$ for more than 5% of the firms when tested at 0.05 significance level. In fact, we are never estimating the true β_i because the βs are not equal to one, which implies that we are never actually estimating α_{Mt} so our βs are unidentified because there is no way to disentangle (5.4) above.

Although most of the $\hat{\beta}_i$ are statistically different from zero, which means that the tail risks of individual firms may actually be driven by a common firm-level process, the relationship between the tail risks of individual firms and the common process α_{Mt} is likely more complex than what Kelly and Jiang's hypothesis may suggest. The parameter β_i in equation $\ln\alpha_{it} = c_i + \beta_i \ln\alpha_{Mt}$ is an elasticity, and what these results show is that the elasticity of the tail risk of individual assets is not necessarily equal to one. It may happen that certain securities react more aggressively to a change in market risk ($\beta_i > 1$), while others react less aggressively ($\beta_i < 1$). For some specific cases, individual tail risk may not react at all to market risk ($\beta_i = 0$), or display counter cyclical effects ($\beta_i < 0$).[9]

To provide further evidence of the sensitivity of the various sectors to $\ln\alpha_{Mt}$ we compute boxplots for the estimates of β_i in each sector (see Fig. 3). The boxplot is a useful tool as it corresponds to a simple and summarized way of displaying the distribution of the sectorial set of β_i estimates based on their minimum value, the first quartile (Q1), the median, the third quartile (Q3), and their maximum value. It is also informative with respect to outliers and, whether estimates are symmetric, and how tightly they are grouped. The whiskers (the two lines outside the box) extend to the highest and lowest observations.

From Fig. 3 we observe that Financials, Health Care and Information Technology seem to be the most sensitive sectors to $\ln\alpha_{Mt}$. For Real Estate and Utilities we observe that $\hat{\beta}_i < 0$ for a large number of stocks, which suggests that $\ln\alpha_{Mt}$ has a counter-cyclical effect on these stocks. The Energy, Financials, Real-estate and Utilities sector's median $\hat{\beta}_i$ is lower than the corresponding overall median (the median of all stocks in the sample). Hence, for stocks in these sectors, the likelihood of observing extreme negative returns is considerably higher when market risk increases i.e., when $\ln\alpha_{Mt}$ decreases. In the opposite direction, we observe stocks in the Health Care and Information Technology sectors. Furthermore, the impact of $\ln\alpha_{Mt}$ is more homogeneous in the Energy sector than in the Financials sector. Note that although the inter-quartile range in the Financials sector is relatively narrow the plot does display a large number of outliers. Hence, although market tail risk is an important variable of stocks tail risk, other variables may also play a relevant role in firms' tail risk dynamics.

5.2. Time-varying tail risk dynamics

In this section, we analyse the impact that market and firm specific variables have on the tail risk of firms' stock returns. The resulting risk measure exploits the impact of the CBOE Volatility Index (*VIX*),[10] Expected Shortfall (*ES*), capitalization (*cap*) and the market-to-book ratio (*mtb*) on firms returns' tail risk.[11] These covariates were chosen because they convey specific information related to market (*VIX* and *ES*) and firm dynamics (*cap* and *mtb*). Specifically, the *VIX* is a real-time market index that represents the market's expectations for volatility over the coming 30 days. This indicator is chosen because volatility is often seen as a way to gauge market sentiment, i.e. it is an "investor fear gauge" as indicated by Bollerslev et al. (2015, p.113). In our analysis below, we consider the log of the *VIX*, vix_t, as our regressor.

[9] It might be interesting to relate c_i to the "idiosyncratic tail risk" of the individual asset and β_i to the systematic risk component. However, this analysis will be left for future work.

[10] See https://markets.cboe.com/tradable_products/vix/.

[11] Other variables, such as, the CBOE Skew Index (https://www.cboe.com/) and the Treasury Term Premia provided by the Federal Reserve Bank of New York (https://www.newyorkfed.org/research/data_indicators/term-premia-tabs), have also been considered, but the VIX and ES displayed more widespread significance across firms.

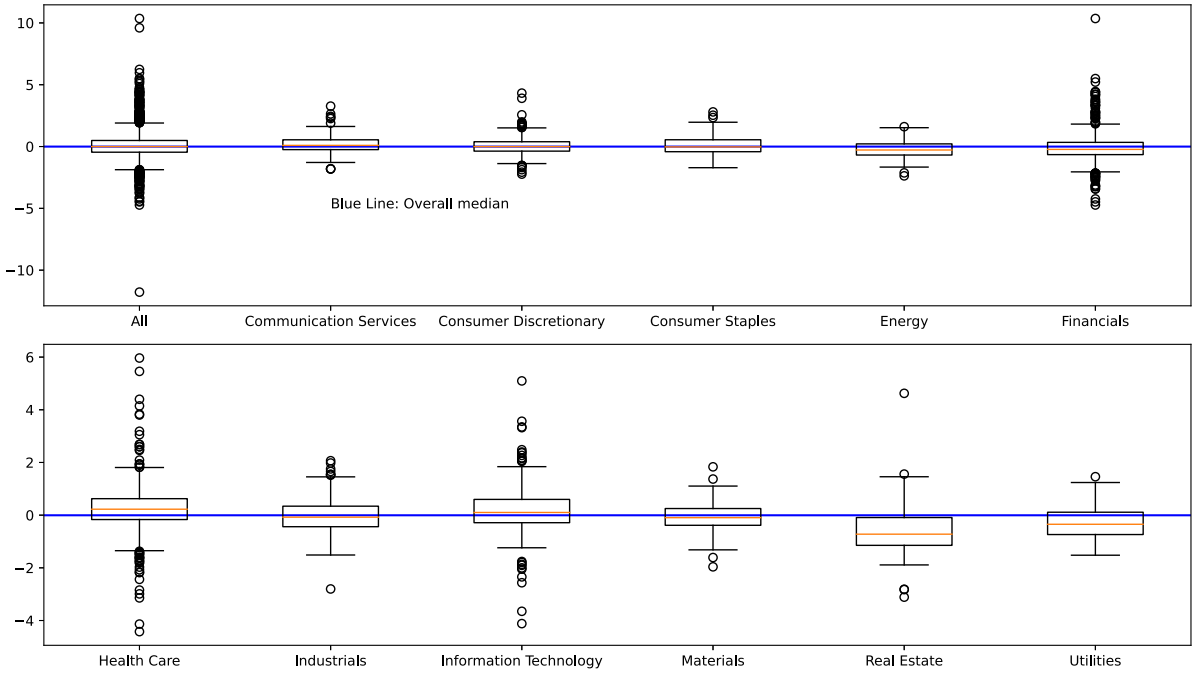


Fig. 3. Impact of $\ln\alpha_{Mt}$ on firms' tail risk (β_i estimates in the various sectors).

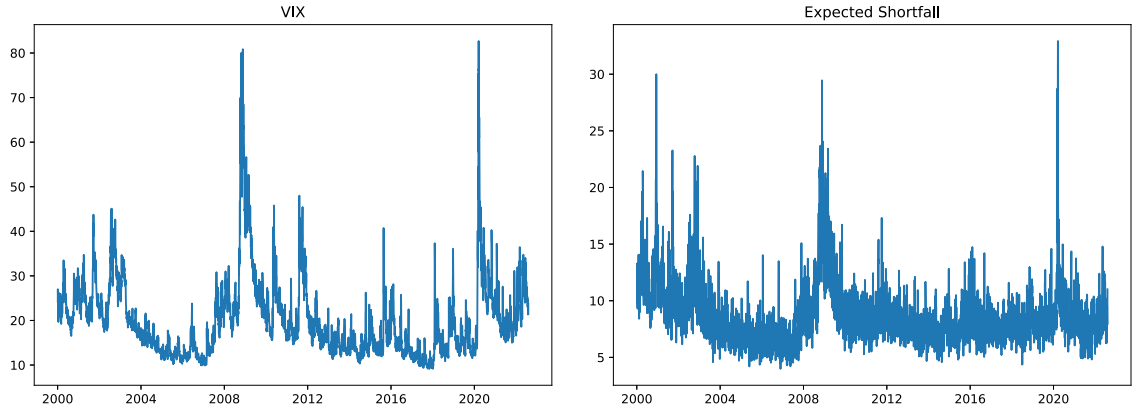


Fig. 4. Plots of the VIX_t and ES_t.

ES is the other market risk variable used, which is sensitive to the shape of the tail of the distribution of returns, unlike the more commonly used value-at-risk (*VaR*). *ES* is calculated at a specific point in time by averaging all cross-sectional returns smaller than the 5th empirical quantile of the distribution of returns. Thus, *ES* is computed as the average of the worst 5% returns and is multiplied by -1 to represent a loss.[12] In Fig. 4 we plot these two market risk indicator variables.

Regarding the firm specific covariates, capitalization, cap_{it}, is of interest as a lot of empirical evidence has shown that firms with small market capitalizations are riskier than firms with higher market capitalizations (Banz, 1981; Reinganum, 1981). Evidence suggests that small stocks are more prone to certain risk factors than large stocks, which leads to return differentials between small and large market capitalization stocks, compensating investors who bear those risks (Fama and French, 1993, 1996). According to Aboura and Arisoy (2019) the justifications for this differential are: (1) higher transaction costs associated with small stocks (Amihud and Mendelson, 1986); (2) a consequence of investors wrongly valuing stocks (Lakonishok et al., 1994; Porta et al., 1997); and (3) small stocks being more prone to business downturns, due to their limited capacity to cope with deterioration in investment opportunities (Lettau and Ludvigson, 2001; Petkova and Zhang,

12 The portfolio or financial position is 100 dollars or euros.

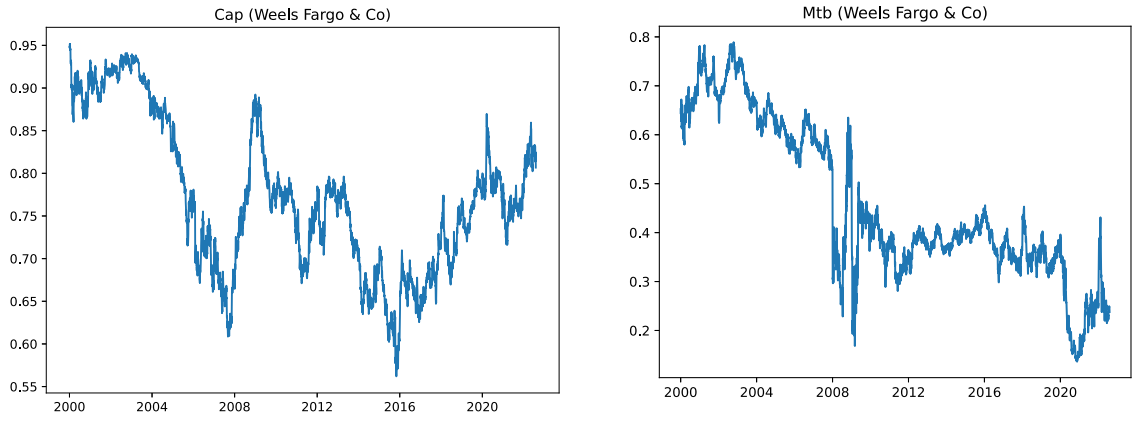


Fig. 5. Plots of *cap* and *mtb* for WELLS FARGO & CO.

2005). In our application, cap_{it} is the cross-sectional quantile order of the empirical distribution of capitalization of firm i at time t. For instance, a $cap_{it} = 0.80$ indicates that 80% of the firms at time t have a smaller capitalization than firm i.

Finally, mtb_{it} is the cross-sectional quantile order of the current closing price of the stock divided by the most current quarter's book value per share. mtb_{it} is used in the literature either to indicate the value that the market places on the common equity or net assets of a firm (Ceccagnoli, 2009; Lee and Makhija, 2009), or as a firm specific risk indicator (Griffin and Lemmon, 2002; Liew and Vassalou, 2000). It is well known that in the cross-section of stock returns firms with a low *mtb* are, on average, more prone to risk than firms with a high *mtb*. Considering the value premium, defined as the difference in average returns between low and high *mtb* firms, Fama and French (1992, 1996) argue that this difference (the value premium) is a proxy for distress risk. Chen et al. (2001) also show that higher (lower) *mtb* is associated with more negative (positive) skewness. Different asymmetries between the return distributions of low and high *mtb* firms could also imply different exposure to aggregate tail risk. Given that crashes tend to occur in times of market stress (Daniel and Moskowitz, 2016), and that the value premium is countercyclical, low *mtb* firms could thus be more negatively exposed to tail risk than high *mtb* firms (see Aboura and Arisoy, 2019 for a detailed discussion of this variable). As an illustration, Fig. 5 plots the two firm specific covariates, cap_{it} and mtb_{it}, for WELLS FARGO & CO.

Since our aim is to analyse the tail risk of firms we will focus on the left-tail of the corresponding returns' distribution. The left-tail is of importance, as according to Bollerslev et al. (2015), most of the variance risk premium predictability is attributed to the premium for bearing jump tail risk, and it is the negative tail risk that seems generally to be priced (see also Kelly and Jiang, 2014, Chow et al., 2019 and Andersen et al., 2020).

To estimate the left-tail index of stock i's returns we consider a survival function as in (3.1), i.e.,

$$\bar{F}(r_{it}|\mathbf{X}_{it}, w_{in}) = 1 - F(r_{it}|\mathbf{X}_{it}, w_{in}) = (r/w_{in})^{-\alpha(\mathbf{X}_{it},\boldsymbol{\beta})}\mathcal{L}(r_{it}|\mathbf{X}_{it}), \tag{5.5}$$

where $r < w_{in} < 0$ and $\mathcal{L}(r_{it}|\mathbf{X}_{it})$ is a slowly varying function at infinity. Thus, small values of $\alpha_{it}(\mathbf{X}_{it}, \boldsymbol{\beta})$ correspond to heavy tails and high probabilities of extreme returns. In our analysis we focus on excess and risk-adjusted returns (the latter are adjusted for the Fama–French three factors; as e.g. in Jiang et al., 2020). Specifically, r_{it} follows a Pareto-type distribution, which is characterized by a tail index parameter that is a function of the covariates discussed above, i.e., $\mathbf{X}_{it} := (vix_t,\ ES_t,\ cap_{it},\ mtb_{it})$, i.e.,

$$\alpha_{it}\left(\mathbf{X}_{it}, \boldsymbol{\beta}\right) = \exp\left(\beta_{i0} + \beta_{i1} vix_t + \beta_{i2} ES_t + \beta_{i3} cap_{it} + \beta_{i4} mtb_{it}\right). \tag{5.6}$$

Expression (5.6) considers that the sensitivity of a particular stock to the market risk indicators (vix_t and ES_t) and the firm-specific features (cap_{it} and mtb_{it}) is captured by the magnitude of β_{ik}, $k = 1, \ldots, 4$. Moreover, from (5.6) and based on the previous discussion on the nature of the chosen covariates, one expects vix_t and ES_t to, generally, impact negatively firms' tail risk ($\beta_{ik} < 0$, $k = 1, 2$) and for cap_{it} and mtb_{it} to have a positive impact, $\beta_{ik} > 0$, $k = 3, 4$.

One important first step in estimating the conditional tail index $\alpha_{it}(\mathbf{X}_{it}, \boldsymbol{\beta})$ is to select a suitable tail threshold w_{in}, which indicates the beginning of the tail of the distribution of the returns of firm i (see discussion in Sections 2 and 3). To determine w_{in} empirically, we use the discrepancy measure proposed by Wang and Tsai (2009) in a regression context. This measure looks to ensure that the sample fraction chosen produces the smallest discrepancy between the empirical distribution of $\{\hat{U}_{it}(\mathbf{X}_{it}) : r_{it} < w_{in}\}$ and $U[0, 1]$, where $\hat{U}_{it}(\mathbf{X}_{it}) \equiv \hat{U}_{it} = \exp\left(-\exp(\mathbf{X}_t\hat{\boldsymbol{\beta}})\ln\left(\frac{r_{it}}{w_{in}}\right)\right)$, conditional on $r_{it} < w_{in}$, is approximately $U[0, 1]$, that is, uniformly distributed on $[0, 1]$, if w_{in} is well defined. Hence, the discrepancy measure is,

$$\hat{D}(\omega_{in}, \mathbf{X}_{it}) = \frac{1}{n_0}\sum_{t=1}^{n}\{\hat{U}_{it} - \hat{F}_n(\hat{U}_{in})\}^2 I(r_{it} < \omega_{in}), \tag{5.7}$$

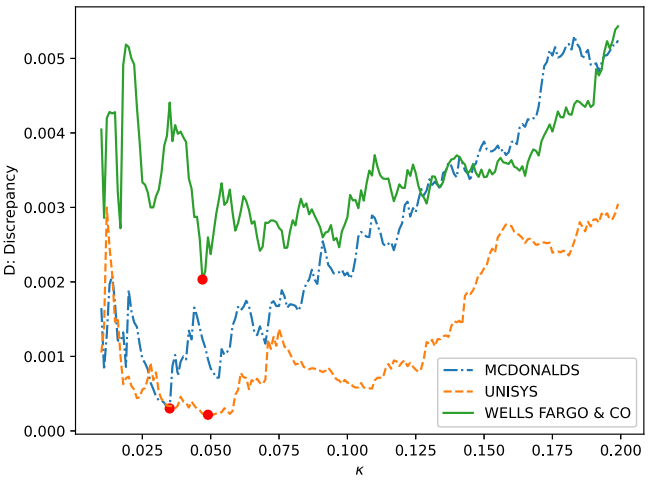


Fig. 6. Plots of $\hat{D}(\omega_{in}, \mathbf{X}_{it})$ for the determination of the left-tail cut off point, ω_{in}.

Table 2
Illustration of heterogeneous impacts of covariates on stocks' tail risk.

	β_{i1}	β_{i2}	β_{i3}	β_{i4}
BANK OF AMERICA	−0.18***	−0.05**	42.46**	0.71**
MCDONALDS	−0.34***	−0.04**	14.03***	0.06
NVIDIA	0.04	−0.03**	2.64**	0.23*
UNISYS	0.27*	−0.06*	0.45*	0.30***
WALMART	−0.30***	−0.06*	28.59**	1.39**
WELLS FARGO & CO	−0.71***	−0.02**	110.31***	1.14**

Note: For inference purposes HAC standard errors, computed as detailed in are considered.
*Significance at the 10% level.
**Significance at the 5% level.
***Significance at the 1% level.

where $\hat{F}_n(.)$ is the empirical distribution of $\{\hat{U}_{it}\}$. If $\hat{U}_{it}$ is indeed uniformly distributed on [0, 1], then $\hat{F}_n(u) \approx u$ for every $u \in [0, 1]$. Accordingly, the value of $\hat{D}(\omega_{in})$ should be small, which suggests that ω_{in} can be selected as

$$\omega_{in}^* = \arg\min_{\omega_{in}} \hat{D}(\omega_{in}, \mathbf{X}_{it}).$$

Since $\hat{U}_{it}$ is a function of the covariates, the optimal value ω_{in}^* is also a function of the covariates $\mathbf{X}_{it}$. This means that if we add or remove one of more covariates, the optimal value of the threshold, ω_{in}^*, may change accordingly.

Fig. 6 provides an illustration of the application of $\hat{D}(\omega_{in}, \mathbf{X}_{it})$ in (5.7) to the returns series of three firms (to MCDONALDS, UNISYS and WELLS FARGO & CO) to determine the left-tail cut off point for the corresponding estimation of (5.6). From Fig. 6 we establish that $\kappa < 0.05$ for the three returns series considered. w_{in} corresponds, in each case, to the minimum of the upper $\lfloor \kappa n \rfloor$ ordered statistics. Based on ω_{in} the sample of returns that compose stock i's returns distribution's left-tail are determined and (5.6) is then used to compute the estimates of β_{ik}, $i = 1, \ldots, N$, where N is the number os stocks used in the analysis, and $k = 0, 1, 2, 3, 4, 5$. The estimates of β_{ik} from (5.6) reveal heterogeneous impacts of the different covariates on stocks' tail risk.

As an illustration, Table 2 provides the conditional tail index regression estimation results for six firms in our sample (BANK OF AMERICA, MCDONALDS, NVIDIA, UNISYS, WALMART and WELLS FARGO & CO). From Table 2 we observe, regarding the market risk indexes (vix_t and ES_t), that the estimates of β_{i1} and β_{i2} generally display the expected sign, i.e., $\beta_{ik} < 0$, $k = 1, 2$ (except for NVIDIA and UNISYS, for which the impact of the vix_t seems to be positive, although only at a 10% significance level for UNISYS and with no statistical significance for NVIDIA). This suggests that, *ceteris paribus*, an increase (decrease) in any of these two variables i.e., an increase (decrease) in market risk as measured through vix_t or ES_t generally contributes to increase (decrease) firm i's stock returns tail risk. The results in Table 2 also show that the firm specific variables, cap_{it} and mtb_{it}, display the expected signs ($\beta_{ik} > 0$, $k = 3, 4$), which suggest that as these indicators increase (decrease) they contribute to a reduction (increase) of the tail risk of these firms.

Overall, the analysis of the tail index estimates for all stocks in our sample reveals that the percentage of coefficient estimates with the expected sign and statistical significance were 47.6% for $\hat{\beta}_{i1}$, 68.3% for $\hat{\beta}_{i2}$, 64.6% for $\hat{\beta}_{i3}$ and 51.9% for $\hat{\beta}_{i4}$ (see results for *All* in Fig. 7). These results highlight the importance of ES_t and cap_{it} for the tail risk dynamics of stocks in general, although vix_t and mtb_{it} also display an important role.

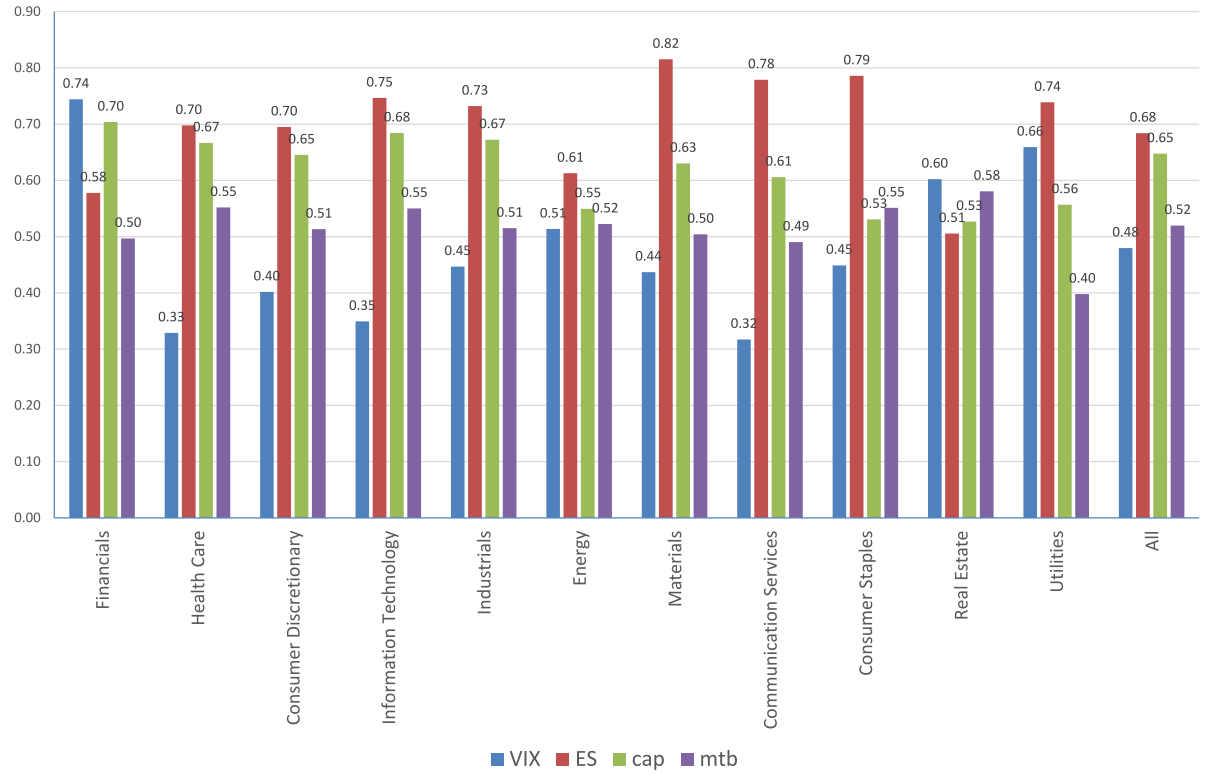


Fig. 7. Percentage of statistically significant coefficient estimates with the expected sign.

Fig. 7 also illustrates the heterogeneous impact of the variables considered in the different sectors of activity. Specifically, of the market risk indicators considered (vix_t and ES_t), ES_t is a very relevant variable in all sectors. The sectors that display the smallest percentage of firms, for which ES_t is statistically significant are Financials and Real Estate (58% and 51%, respectively). For Energy, ES_t is significant for 61% of the firms, and for all other sectors the percentage is above 70%. Moreover, with the exception of Financials and Real Estate, ES_t also seems statistically significant for a larger percentage of firms in all other sectors than vix_t. The more cyclical sectors in terms of tail risk are those in which market risk is the strongest, i.e., those with a higher percentage of significant vix_t and ES_t, such as Financials, Materials and Consumer Staples. The less cyclical are Information, Technology and Health Care.

Regarding the firm specific variables (cap_{it} and mtb_{it}), with the exception of Real Estate, cap_{it} presents a larger percentage of firms for which it is statistically significant than mtb_{it} (see Fig. 7). However, the difference in the percentage of firms for which these two variables are significant is not as marked as between the two market risk indicators considered.

As an illustration, Fig. 8 plots the estimates of α_{it} ($\mathbf{X}_{it}$, $\boldsymbol{\beta}$), for the six firms considered in Table 2. The purpose of this figure is to illustrate the time-varying dynamics of α_{it} ($\mathbf{X}_{it}$, $\boldsymbol{\beta}$). This figure shows that the probabilities of observing extreme values in the left tail critically increased during the crises of 2000,2008–2009 and 2020 for the six firms considered in this illustration. The crises of 2000, 2008–2009 and 2020 were a generalized phenomena which impacted all sectors. However, from Fig. 8 we also observe firm specific episodes of increased tail risk behaviour. For instance, the plot for WALMART presents two episodes of increased tail risk, one in 2012 and the other in 2016, which seem specific to this firm. In fact, in 2012 WALMART workers went on strike on Black Friday at many US stores nationwide, which impacted the companies stock prices. In 2016 due to increased competition (e.g. from Amazon.com) and the implementation of strategies that did not prove effective enough, WALMART announced the closure of 269 stores across the globe. The other example is the BANK OF AMERICA. The plot for this firm shows an increase in tail risk in 2012. In this year, following the bank's involvement in several law suits after the financial crises and declining revenues due to new regulations and a slow economy, the bank was forced to cut around 16,000 jobs by the end of 2012.

6. Conclusion

This paper has studied pure Pareto and Pareto-type distributions in the presence of covariate information. We link a set of explanatory variables to the tail index such that the log of the tail index is linear in the regressors. We demonstrate

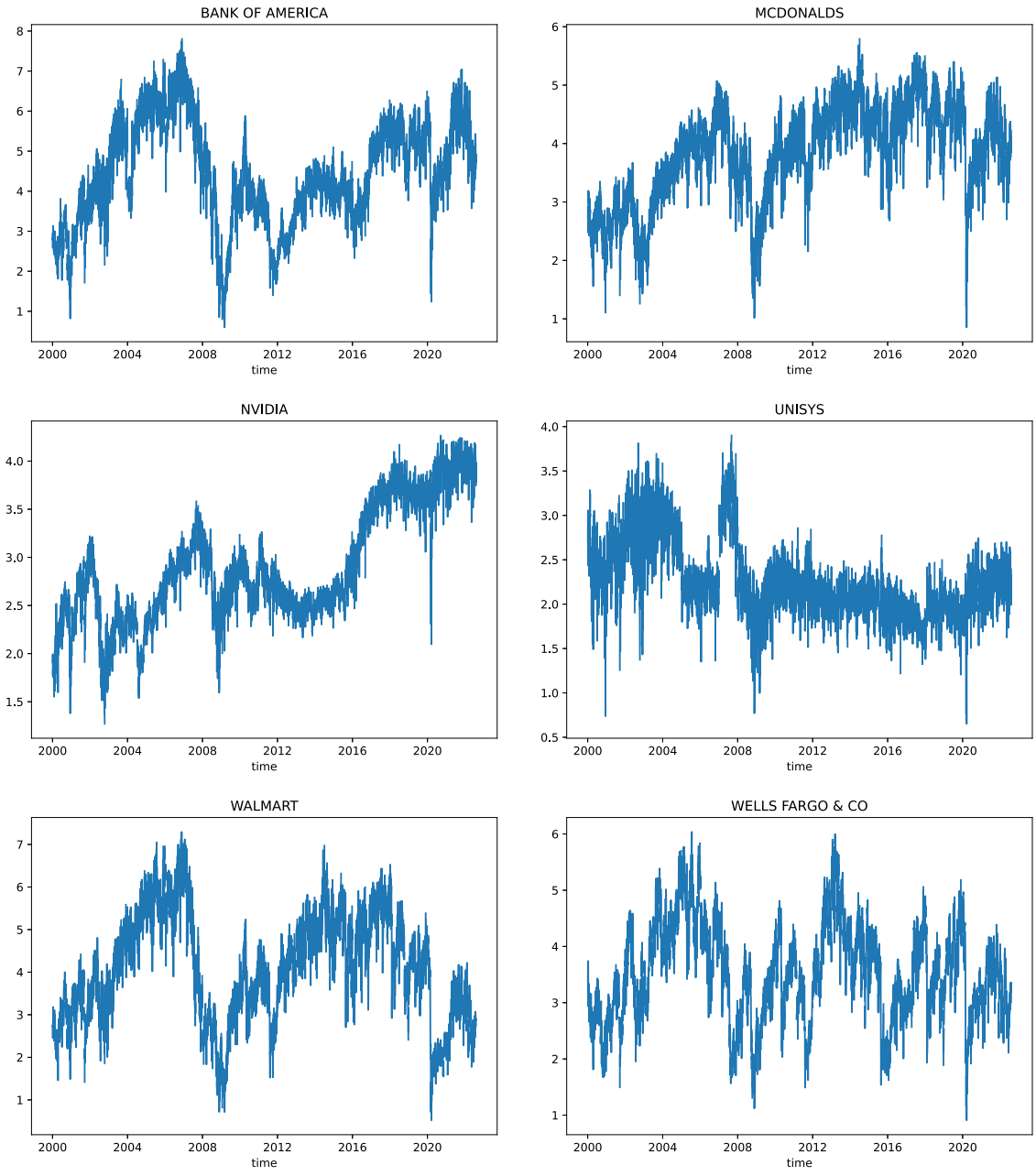


Fig. 8. Tail Index Estimates - $\hat{\alpha}_{it}(\mathbf{X}_{it}, \hat{\boldsymbol{\beta}})$ (A lower value of the tail index indicates a higher probability of extreme values occurring, which in turn translates into a greater level of tail risk.).

that, under appropriate normalization, the Maximum Likelihood estimate of the parameter vector is consistent and asymptotically normal with a zero mean vector and an identity variance–covariance matrix. The results hold under both independently distributed or weakly dependent data. This allows for hypotheses testing with approximate critical values coming from a standard normal distribution. The simulation results provide support for the theoretical findings in the paper and demonstrate the excellent finite sample properties of the estimator.

Our empirical section provides two applications to firms' tail risk dynamics. The first focuses specifically on a market risk measure computed from a cross-section of firms returns, in line with Kelly and Jiang (2014). Interestingly, this analysis shows that this market risk variable is indeed a potential important driver of firm level tail dynamics, but with heterogeneous impacts across firms. We observe that sectors such as Financials, Health Care and Information Technology seem to be the most sensitive to this market risk indicator, whereas Real Estate and Utilities seem to display counter-cyclical behaviour. Moreover, our analysis also reveals that for stocks in the Energy, Financials, Real-estate and Utilities

sectors the likelihood of observing extreme negative returns is higher when market risk increases, whereas for stocks in the Health Care and Information Technology sectors it decreases.

In the second application the impact of market risk on firms' tail risk is considered through the VIX and ES, and firm specific information, such as *capitalization* and *market-to-book* is also included. It is observed that the estimated firm specific tail index provides an interesting indicator of firms tail risk dynamics. Overall, our analysis highlights the importance of ES and *capitalization* for the analysis of firms' tail risk dynamics in general, although the VIX and ES also display an important role. In line with the first application, it is also important to highlight the heterogeneous impact of the variables considered on the stocks from the different sectors of activity. The sectors that display the smallest percentage of firms, for which ES displays the expected sign and is statistically significant are Financials and Real Estate (58% and 51%, respectively). For Energy, ES is significant for 61% of the firms, and for all other sectors, the percentage is above 70%. With the exception of Financials and Real Estate, ES also seems statistically significant for a larger percentage of firms in all other sectors than VIX. The more cyclical sectors in terms of tail risk are those in which market risk is strongest, i.e., those with a higher percentage of significant VIX and ES, such as Financials, Materials and Consumer Staples. The less cyclical are Information, Technology and Health Care. Regarding the firm specific variables, with the exception of Real Estate, *capitalization* presents a larger percentage of firms for which it is statistically significant than *market-to-book*. However, the difference in the percentages is not as marked as between the VIX and ES.

From a theoretical point of view, the literature can further be extended in the following directions. Firstly, extend the type of explanatory variables that can be used. In this paper, we have assumed that we can estimate the variance–covariance structure of the regressors under certain conditions. It would be interesting to develop models that accommodate heavier tails of the independent variables or regressors which have a unit root. Secondly, we have assumed that the explanatory variables are linked to the logarithm of the tail index in a linear fashion. Further contributions could include introducing non-linearities or semi-parametric and non-parametric setups. The analysis could also be extended to cover the case where the conditional variance is determined independently of the tail index. From an empirical point of view, further important steps consist of the analysis of risk pricing, using the firm-specific tail risk measure developed in this paper, as well as its potential from a predictive point of view.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2023.04.002.

References

Aboura, S., Arisoy, Y.E., 2019. Can tail risk explain size, book-to-market, momentum, and idiosyncratic volatility anomalies? J. Bus. Finance Account. 46 (9–10), 1263–1298.

Amihud, Y., Mendelson, H., 1986. Liquidity and stock returns. Financ. Anal. J. 42 (3), 43–48.

Andersen, T.G., Fusari, N., Todorov, V., 2015a. Parametric inference and dynamic state recovery from option panels. Econometrica 83 (3), 1081–1145.

Andersen, T.G., Fusari, N., Todorov, V., 2015b. The risk premia embedded in index options. J. Financ. Econ. 117 (3), 558–584.

Andersen, T., Fusari, N., Todorov, V., 2020. The pricing of tail risk and the equity premium: Evidence from international option markets. J. Bus. Econom. Statist. 38 (3), 662–678.

Andrews, D.W.K., Monahan, J.C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60 (4), 953–966.

Banz, R.W., 1981. The relationship between return and market value of common stocks. J. Financ. Econ. 9 (1), 3–18.

Beirlant, J., Goegebeur, Y., 2003. Regression with response distribution of Pareto-type. Comput. Statist. Data Anal.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., Waal, D., Ferro, C., 2004. Statistics of Extremes: Theory and Applications. Wiley, Series in Probability and Statistics. John Wiley & Sons.

Beirlant, J., Vynckier, P., Teugels, J., 1996. Tail index estimation, Pareto quantile plots, and regression diagnostics. J. Amer. Statist. Assoc. 91, 1659–1667.

Bollerslev, T., Todorov, V., 2014. Time-varying jump tails. J. Econometrics 183 (2), 168–180, Analysis of Financial Data.

Bollerslev, T., Todorov, V., Xu, L., 2015. Tail risk premia and return predictability. J. Financ. Econ. 118 (1), 113–134.

Ceccagnoli, M., 2009. Appropriability, preemption, and firm performance. Strateg. Manag. J. 30 (1), 81–98.

Chen, J., Hong, H., Stein, J.C., 2001. Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices. J. Financ. Econ. 61 (3), 345–381.

Chow, K., Li, J., Sopranzetti, B., 2019. The predictive power of tail risk premia on individual stock returns. In: Working Paper.

Christoffersen, P., Feunou, B., Jeon, Y., Ornthanalai, C., 2020. Time-varying crash risk embedded in index options: The role of stock market liquidity*. Rev. Finance 25 (4), 1261–1298.

Csorgo, S., Deheuvels, P., Mason, D., 1985. Kernel estimates of the tail index of a distribution. Ann. Statist. 13, 1050–1077.

Daniel, K., Moskowitz, T.J., 2016. Momentum crashes. J. Financ. Econ. 122 (2), 221–247.

Einmahl, J.H.J., de Haan, L., Zhou, C., 2016. Statistics of heteroscedastic extremes. J. R. Stat. Soc. Ser. B Stat. Methodol. 78 (1), 31–51.

Fama, E.F., 1963. Mandelbrot and the stable paretian hypothesis. J. Bus. 36 (4), 420–429.

Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. J. Finance 47 (2), 427–465.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33 (1), 3–56.

Fama, E.F., French, K.R., 1996. Multifactor explanations of asset pricing anomalies. J. Finance 51 (1), 55–84.

Gabaix, X., 1999. Zipf's law for cities: An explanation. Q. J. Econ. 114, 739–767.

Gabaix, X., 2012. Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance *. Q. J. Econ. 127 (2), 645–700.

Gabaix, X., Ibragimov, R., 2012. Log(rank-1/2): A simple way to improve the OLS estimation of tail exponents. J. Bus. Econ. Stat. 29 (1), 24–39.

Griffin, J.M., Lemmon, M.L., 2002. Book-to-market equity, distress risk, and stock returns. J. Finance 57 (5), 2317–2336.

Haan, W.J.D., Levin, A., 1996. Inferences from Parametric and Non-Parametric Covariance Matrix Estimation Procedures. Technical Working Paper Series 195, National Bureau of Economic Research.

Hall, P., 1982. On some simple estimates of an exponent of regular variation. J. R. Stat. Soc. Ser. B Stat. Methodol. 44 (1), 37–42.

Hayashi, F., 2000. Econometrics. Princeton University Press.
Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. Ann. Statist. 3, 1163–1174.
Hill, J.B., Shneyerov, A., 2013. Are there common values in first-price auctions? A tail-index nonparametric test. J. Econometrics 174 (2), 144–164.
Jiang, L., Wu, K., Zhou, G., Zhu, Y., 2020. Stock return asymmetry: Beyond skewness. J. Financ. Quant. Anal. 55 (2), 357–386.
Kelly, B., Jiang, H., 2014. Tail risk and asset prices. Rev. Financ. Stud. 27 (10), 2841–2871.
Kiefer, N.M., Vogelsang, T.J., 2005. A new asymptotic theory for Heteroskedasticity-autocorrelation robust tests. Econom. Theory 21 (6), 1130–1164.
Kratz, M., Resnick, S.I., 1996. The qq-estimator and heavy tails. Stoch. Models 12, 699–724.
Lakonishok, J., Shleifer, A., Vishny, R.W., 1994. Contrarian investment, extrapolation, and risk. J. Finance 49 (5), 1541–1578.
Lazarus, E., Lewis, D.J., Stock, J.H., Watson, M.W., 2018. HAR inference: Recommendations for practice. J. Bus. Econom. Statist. 36 (4), 541–559.
Lee, S.-H., Makhija, M., 2009. Flexibility in internationalization: is it valuable during an economic crisis? Strateg. Manag. J. 30 (5), 537–555.
Lee, J.H., Sasaki, Y., Toda, A.A., Wang, Y., 2021. Fixed-k tail regression: New evidence on tax and wealth inequality from forbes 400. arXiv.org 2105.10007.
Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. J. Finance 56 (3), 815–849.
Liew, J., Vassalou, M., 2000. Can book-to-market, size and momentum be risk factors that predict economic growth? J. Financ. Econ. 57 (2), 221–245.
Mandelbrot, B., 1963. The variation of certain speculative prices. J. Bus. 36.
Müller, U.K., 2007. A theory of robust long-run variance estimation. J. Econometrics 141 (2), 1331–1352.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.
Nicolau, J., Rodrigues, P.M.M., 2019. A new regression-based tail index estimator. Rev. Econ. Stat. 101 (4), 1–14.
Novak, S.Y., 2011. Extreme Value Methods with Applications To Finance. Chapman & Hall, CRC Monographs on Statistics and Applied Probability.
Petkova, R., Zhang, L., 2005. Is value riskier than growth? J. Financ. Econ. 78 (1), 187–202.
Porta, R.L., Lakonishok, J., Shleifer, A., Vishny, R.W., 1997. Good news for value stocks: Further evidence on market efficiency. J. Finance 52 (2), 859–874.
Quintos, C., Fan, Z., Phillips, P., 2001. Structural change tests in tail behaviour and the Asian crisis. Rev. Econom. Stud. 68 (3), 633–663.
Reinganum, M.R., 1981. Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values. J. Financ. Econ. 9 (1), 19–46.
Resnick, S.I., 2007. Heavy-Tail Phenomena: Probability and Statistical Modelling. Springer, New York.
Rosen, K.T., Resnick, M., 1980. The size distribution of cities: an explanation of the Pareto law and primacy. J. Urban Econ. 8, 165–186.
Samorodnitsky, G., Taqqu, M., 2004. Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance: Stochastic Modeling, first ed. Taylor & Francis.
Sasaki, Y., Wang, Y., 2022. Fixed-k inference for conditional extremal quantiles. J. Bus. Econom. Statist. 40 (2), 829–837.
Wachter, J.A., 2013. Can time-varying risk of rare disasters explain aggregate stock market volatility? J. Finance 68 (3), 987–1035.
Wang, H., Tsai, C., 2009. Tail index regression. J. Amer. Statist. Assoc. 104 (487), 1233–1240.
White, H., 2001. Asymptotic theory for econometricians. Emerald Group Publishing Limited.