Time Varying Extremes*

Ulrich K. Müller and Mark W. Watson
Department of Economics
Princeton University

This Draft: December 2023

Abstract

Standard extreme value theory implies that the distribution of the largest observations of a large cross section is well approximated by a parametric model, governed by a location, scale and shape parameter. The extremes of a panel of independent cross sections are all governed by the same parameters as long as the underlying distribution as well as the size of the cross sections are time invariant. We derive inference about these parameters, and tests of the null hypothesis of time invariance, under asymptotics that do not require the number of extremes or the number of time periods to increase. We further apply Hamiltonian Monte Carlo techniques to estimate the path of time-varying parameters. We illustrate the approach in four examples of U.S. data: damages from weather-related disasters, financial returns, city sizes and firm sizes.

Keywords: Generalized Pareto Distribution, Generalized Extreme Value Distribution, Tests of Parameter Constancy, Hamiltonian Monte Carlo
JEL: C22, C12

*Müller acknowledges financial support from the National Science Foundation grant SES-2242455.

1 Introduction

Extreme value theory characterizes the approximate distribution of the largest k observations in a sample of n i.i.d. observations. In particular, if the underlying population has a tail that is sufficiently well approximated by the Generalized Pareto (GP) distribution, then the largest observations approximately follow a (joint) Generalized Extreme Value (GEV) distribution. The location, scale and tail index parameters of this GEV approximation are functions of the sample size n, and the location, scale and shape parameter of the underlying GP distribution. Thus, the extreme observations may be used to learn about the tail properties of the underlying population. There is a well developed literature on corresponding formal inference procedures based on $k \to \infty$ asymptotics; see, for instance, Embrechts, Klüppelberg, and Mikosch (1997), Coles (2001), de Haan and Ferreira (2007) and Gomes and Guillou (2015) for reviews and references.

Approximations based on $k \to \infty$ asymptotics may be inaccurate in small to moderate sample sizes, however: On the one hand, selecting k large can lead to substantial biases, as the GP tail approximation may only hold for a small fraction of the underlying population. On the other hand, a small k invalidates the central limit theorems and similar approximations justifying the $k \to \infty$ asymptotics. As an alternative, Müller and Wang (2017) derive asymptotically valid inference for extreme quantiles and tail expectations by considering the small sample problem of observing a fixed k number of observations from the GEV distribution directly.

Now suppose we observe the largest observations from T independent samples, collected at different points in time $t = 1, \ldots, T$. If the sample sizes n_t and the underlying GP parameters are time invariant, then the resulting T vectors of the k largest observations are independent, with the same GEV parameters. We generalize Müller and Wang's (2017) generalized likelihood ratio statistic to conduct inference about these parameters, or functions thereof, based on $n_t = n \to \infty$ but fixed-k, fixed-T asymptotics. Under such asymptotics, it is not possible to consistently estimate the GEV parameters, so for null hypotheses that involve nuisance parameters, one must derive critical values that reflect this uncertainty. We also generalize these results for observations of exceedances, that is values that exceed a given (extreme) threshold.

We further develop a test for the null hypothesis whether the GEV parameter are stable

in time in the same framework. Our suggested test statistic here is that derived in Nyblom (1989), but again with a critical value that is adjusted for the small sample distribution induced by the GEV observations.

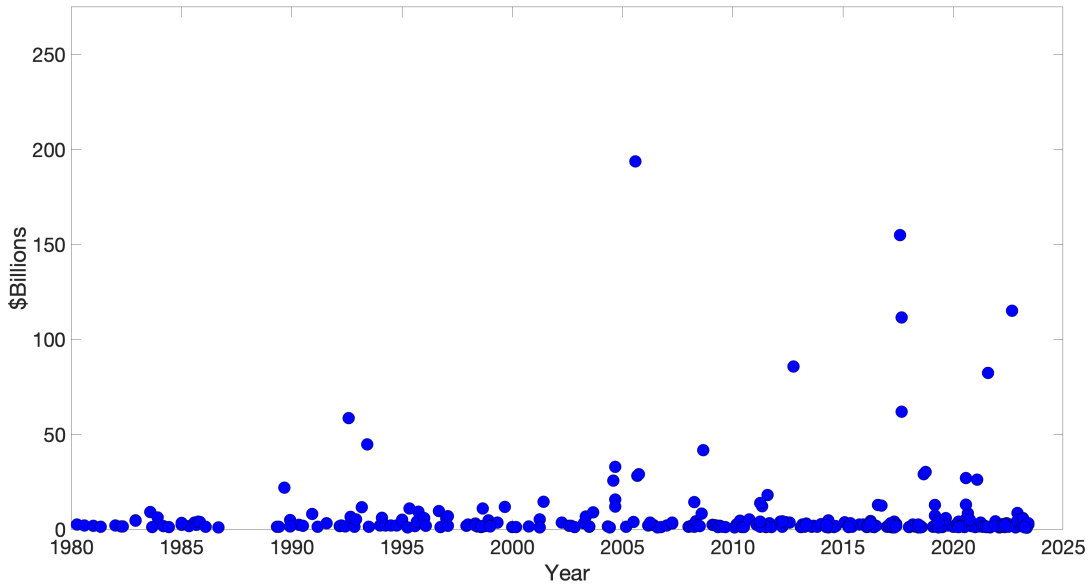
If the null hypothesis of parameter stability is rejected, it is natural to ask how the parameters evolve over time. We take a Bayesian perspective and model the evolution of the (suitably transformed) underlying parameters as Gaussian random walks, which provide a flexible characterization of the parameter paths. The computation of the posterior parameter path is a non-Gaussian filtering/smoothing problem. Gaetan and Grigoletto (2004) rely on a particle filter algorithm.[1] We instead suggest a simpler and computationally faster approach based on Hamiltonian Monte Carlo (HMC). The support of the GEV distribution depends on its parameters, so HMC is not directly applicable. We overcome this by suitably extrapolating the likelihood to the entire Euclidian space, followed by an importance sampling correction of the HMC output. We find that this approach, implemented in the Bayesian posterior package Stan, reliably yields accurate posteriors in minutes, even for problems with large T ($T = 522$ in one of our illustrations), and it can also be applied to the fast computation of Bayes factors via Meng and Wong's (1996) "simple identity."

We illustrate our approach in four empirical examples. Our first example concerns the damages from U.S. weather-related disasters that exceed one billion (2023 inflation adjusted) dollars for the months 1980:1-2023:6, as depicted in Figure 1.[2] The censoring at $\tau = \$1$ billion leads to an endogenous number k_τ of exceedances for each of the $T = 522$ months, with k_τ ranging from zero to six. Note that the sample size n_t, the number of all weather-related events in a given month, is not observed. It is hence meaningful to ask whether n_t, and/or the underlying GP parameters are time varying, as could be induced by a time varying capital stock, time-varying population density in coastal regions, or climate change (cf. Smith and Katz (2013), Katz (2015), and Pielke, Gratz, Landsea, Collins, Saunders, and Musulin (2008)). We find substantial time variation in the GEV parameters describing these data, even after normalizing the data to adjust for changes in the value of the aggregate capital stock.

Second, we consider the smallest and largest daily return of the SP500 index over non-

[1] Also see Coles (2001), Huerta and Sansó (2007), do Nascimento, Gamerman, and Lopes (2016), Nakajima et al. (2012, 2017), among others, for related approaches to introduce time variation in GEV parameters.

[2] The data are from the NCEI U.S. Billion-Dollar Weather and Climate Disasters dataset. Appendix A provides a detailed description of all data used in this paper.

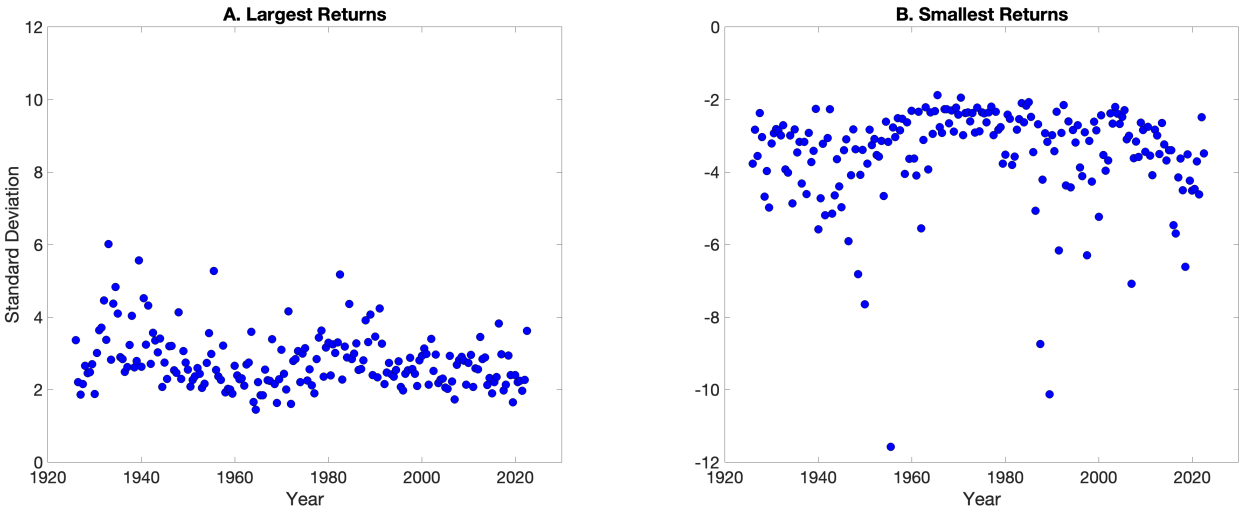


Notes: The figure shows weather-related disaster damages that exceed $1Billion for each month from 1980:1 through 2013:6. See Appendix A for a description of the data.

Figure 1: Damages from U.S. weather-related disasters that exceed one billion dollars ($2023)

overlapping six month periods from 1926 through 2022. In order to reduce the obvious time dependence in return volatility, the daily returns are standardized by a (constant parameter) GARCH(1,1) model, an approach that goes back to McNeil and Frey (2000). In this exercise there are $T = 194$ biannual observations and $k = 1$; see Figure 2. We find strong evidence of time variation, suggesting that the extreme left-tail risk of the stock market undergoes multidecadal swings with low risk in the middle of our sample relative to the GARCH(1,1) benchmark. This complements research based on time varying extreme value theory by Chavez-Demoulin, Embrechts, and Sardy (2014) and Mao and Zhang (2018) that focusses on much shorter time intervals.

Our third example considers the population of the $k = 30$ largest US cities in the $T = 4$ years 1900, 1940, 1980 and 2020 relative to total U.S. population. See panel (a) of Figure 3 which plots the logarithm of city sizes for each of the four years. Gabaix (1999, 2016) argues that city sizes follow Zipf's law, which corresponds to an underlying Pareto distribution with tail index equal to unity. We find a tail index that is significantly less than unity, reject the null hypothesis of constant GEV parameters, but find that the estimated parameter time



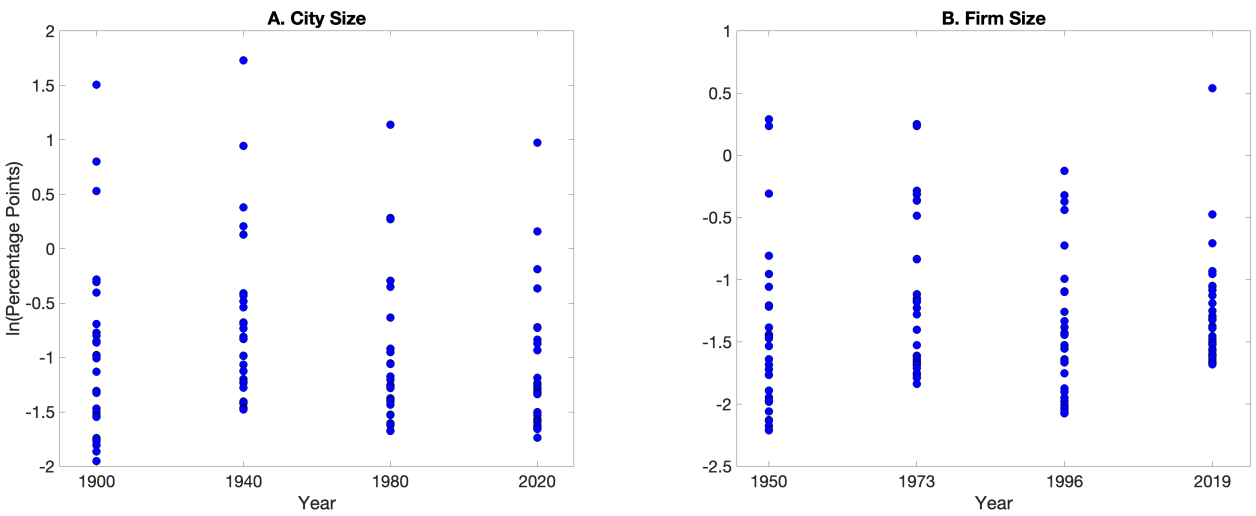
Notes: The figure shows the largest and smallest GARCH(1,1) standardized returns over non-overlapping 6-month periods. See Appendix A for a description of the data.

Figure 2: Largest and Smallest Standardized Daily Returns over 6 Month Periods

paths exhibit little variation.

The fourth and final example considers the largest $k = 30$ U.S. firms by employment in the $T = 4$ years 1950, 1973, 1996 and 2019, normalized by total U.S. employment, as depicted in panel (b) of Figure 3. An influential paper by Axtell (2001) argues that U.S. firms sizes are approximately Pareto with unit shape coefficient, and that this distribution is stable over time, but more recent evidence by Kondo, Lewis, and Stella (2023) sheds doubt on that finding. With our choice of $k = 30$, we focus exclusively on the extreme right tail of firms. We find evidence against the unit shape parameter, but little evidence against the null of constant coefficients, a finding that is inconsistent with a "rise of superstar firms" (cf. Autor, Dorn, Katz, Patterson, and Reenen (2020)) over that period.

The remainder of the paper is organized as follows. The next section reviews extreme value theory and takes a first look at the data assuming the GP parameters and sample sizes n_t are time invariant. Section 3 introduces time variation and discussed our test of the null hypothesis of time invariance. Section 4 concludes.



Notes: City size is population relative to total U.S. population. Firm size is employment relative to total private U.S. employment. See Appendix A for a description of the data.

Figure 3: Thirty Largest Cities and Firms in the United States

2 Constant Parameter Models

2.1 Extreme Value Theory for k Largest Observations

Consider an i.i.d. sample $W_1, \ldots, W_n$ from a population with c.d.f. F. Let $X_1 \geq X_2 \geq \cdots \geq X_n$ be the order statistics, with X_1 the sample maximum, and we are interested in the k largest values $\mathbf{X} = (X_1, \ldots, X_k)$. Assume initially that F is equal to the c.d.f. of a Generalized Pareto (GP) distribution, that is

$$F(w) = F_{\text{GP}}(w) = 1 - \left(1 + \xi\left(\frac{w-\nu}{\omega}\right)\right)^{-1/\xi}, \; w \in S_{\nu,\omega,\xi}$$

where (ν, ω, ξ) are the location, scale and shape parameters, expressions of the form $(1+\xi x)^{-1/\xi}$ are understood to equal e^{-x} for $\xi = 0$, and the support of F_{GP} equals $S_{\nu,\omega,\xi} = (-\infty, \nu - \omega/\xi)$ for $\xi < 0$, $S_{\nu,\omega,\xi} = \mathbb{R}$ for $\xi = 0$ and $S_{\nu,\omega,\xi} = (\nu - \omega/\xi, \infty)$ for $\xi > 0$. The quantile function $F_{\text{GP}}^{-1} : (0,1) \mapsto \mathbb{R}$ of F_{GP} is given by

$$F_{\text{GP}}^{-1}(q) = \omega\frac{(1-q)^{\xi} - 1}{\xi} + \nu.$$

A standard calculation (cf. Proposition 1.6.9 of Reiss (1989)) shows that the order statistics $U_{j:n}$ (so $U_{1:n}$ is the sample minimum) of n i.i.d. uniform random variables have joint distribution

$$\{U_{j:n}\}_{j=1}^{n} \sim \left\{ \kappa_n n^{-1} \sum_{i=1}^{j} E_i \right\}_{j=1}^{n} \tag{1}$$

where E_i are i.i.d. standard exponential and $\kappa_n^{-1} = n^{-1} \sum_{i=1}^{n+1} E_i$. By the Law of Large Numbers, $\kappa_n \xrightarrow{p} 1$ as $n \to \infty$.

By construction, the order statistics $\{X_{n-j+1}\}_{j=1}^{n}$ are related to the $\{U_{j:n}\}_{j=1}^{n}$ via the quantile function, that is

$$\{X_{n-j+1}\}_{j=1}^{n} \sim \{F_{\mathrm{GP}}^{-1}(U_{j:n})\}_{j=1}^{n} \sim \left\{ \omega \frac{(1-U_{j:n})^{-\xi} - 1}{\xi} + \nu \right\}_{j=1}^{n}$$

so, noting that $U_{j:n} \sim (1 - U_{n+1-j:n})$,

$$\begin{aligned} \{X_j\}_{j=1}^{k} &\sim \left\{ \omega \frac{U_{j:n}^{-\xi} - 1}{\xi} + \nu \right\}_{j=1}^{k} \overset{a}{\sim} \left\{ \omega n^{\xi} \frac{(\sum_{i=1}^{j} E_i)^{-\xi} - 1}{\xi} + \omega \frac{n^{\xi} - 1}{\xi} + \nu \right\}_{j=1}^{k} \qquad (2) \\ &\sim \mathrm{GEV}_k(\omega \frac{n^{\xi} - 1}{\xi} + \nu, \omega n^{\xi}, \xi) \qquad (3) \end{aligned}$$

where the approximate distributional equality "$\overset{a}{\sim}$" sets $\kappa_n = 1$, and the joint Generalized Extreme Value distribution $\mathrm{GEV}_k(\mu, \sigma, \xi)$ has density

$$f_{\mathrm{GEV}}(\mathbf{x}|\mu,\sigma,\xi) = \sigma^{-k} \exp\left[-\left(1 + \xi \frac{x_k - \mu}{\sigma}\right)^{-1/\xi} - (1 + 1/\xi) \sum_{j=1}^{k} \log(1 + \xi \frac{x_i - \mu}{\sigma}) \right] \tag{4}$$

with support characterized by the inequalities $x_1 \geq x_2 \geq \cdots \geq x_k$ and $x_j \geq \mu - \sigma/\xi$ for $j = 1, \ldots, k$. We thus have a mapping from the underlying GP parameters (ν, ω, ξ) and the sample size n to the GEV parameters $(\mu, \sigma, \xi) = (\omega \frac{n^{\xi}-1}{\xi} + \nu, \omega n^{\xi}, \xi)$ of the k largest observations, and the only approximation involved in this mapping is $\kappa_n \approx 1$. Up to location and scale, the GEV_1 distribution nests the Weibull, Gumbel and Fréchet distributions as special cases for $\xi < 0$, $\xi = 0$ and $\xi > 0$, respectively.

Now suppose that only the right tail of F is characterized by a GP distribution, that is, for some w_0 with $0 < F(w_0) < 1$, $F(w) = F_{\mathrm{GP}}(w)$ for all $w \geq w_0$. Then (2) still holds

conditional on $U_{k:n} \leq 1 - F(w_0)$, since the quantile function of F for $q \geq F(w_0)$ is just as in the above calculation. Under the approximation $\kappa_n \approx 1$, $U_{k:n}$ in (1) has mean k/n and standard deviation $\sqrt{k}/n$, so it suffices that the extreme right tail with probability mass of a little more than k/n behaves like a GP distribution.

Given that k/n converges to zero for fixed k as $n \to \infty$, it formally suffices that the asymptotic tail behavior of F is that of a GP distribution. This turns out to be a necessary and sufficient condition for extreme value theory to apply—see, for instance, Theorem 5.1.1 in Reiss (1989) for a precise statement. All "textbook" continuous distributions fall in this class, including distribution with finite support ($\xi < 0$), standard thin-tailed distributions such as the normal or log-normal distribution ($\xi = 0$), and fat-tailed distributions such as the student-t or F-distribution ($\xi > 0$). A population F with $\xi \geq 1/2$ has no second moment, and with $\xi \geq 1$ also the first moment no longer exists. In practice and for finite n, these results suggest that for many populations F, the approximation (3) is quite accurate for some choice of GP parameters (ν, ω, ξ), at least for a small enough k. Thus, with a judicious choice of k, we can learn about (μ, σ, ξ), and thus also (ω, ν, ξ), via the likelihood (4).

When $\xi > 0$, so that F has unbounded support in the right tail, the GEV location parameter $\mu = \omega\frac{n^\xi - 1}{\xi} + \nu$ is dominated for large n by the term $\omega n^\xi/\xi = \sigma/\xi$. Standard Fréchet extreme value theory correspondingly makes the additional approximation that $\mu \approx \sigma/\xi$, so that $\omega^{-1} n^{-\xi} X_1 \overset{a}{\sim} \text{GEV}_1(1/\xi, 1, \xi) \sim$Fréchet. For instance, this is implicitly applied in the derivation of the popular Hill (1975) estimator of the tail index ξ. We do not do so in the following because even if the tail of F is exactly GP, so that (3) is an excellent approximation for all moderately large n, for any n one can always choose the GP location parameter ν such that $\omega\frac{n^\xi - 1}{\xi} + \nu$ is arbitrarily different from $\omega n^\xi/\xi$. In other words, the additional approximation $\mu \approx \sigma/\xi$ is not uniformly valid over all underlying populations for which (3) yields accurate approximations. Instead, the restriction $\mu = \sigma/\xi$ is one of the null hypotheses we consider below. Note that if the GP parameters satisfy $\nu = \omega/\xi$, then $\mu = \sigma/\xi$, and the GP distribution with $\nu = \omega/\xi$ is simply the Pareto distribution. Relying on $\mu \approx \sigma/\xi$ in the GEV approximation (3) thus implicitly assumes that the tail of F is well approximable by a (non-shifted) Pareto distribution, and this another way of motivating the null hypothesis of our test below.

The generalization to a panel of independent samples is straightforward: Let $\mathbf{X}_t \in \mathbb{R}^k$ be the largest k observations of sample $W_{t,1}, \ldots, W_{t,n}$, $t = 1, \ldots, T$. Then under the same

assumptions that lead to (3), we obtain a large-n justified approximate likelihood of the form $\prod_{t=1}^{T} f_{\text{GEV}}(\mathbf{X}_t|\mu,\sigma,\xi)$, where $(\mu,\sigma,\xi) = (\omega\frac{n^{\xi}-1}{\xi}+\nu, \omega n^{\xi}, \xi)$. If the samples are weakly dependent then this approximation can still be justified — see, for instance, Leadbetter (1983) for precise conditions.

2.2 Inference Using k Largest Observations

We now discuss inference about (functions of) the GEV parameters (μ,σ,ξ) under fixed-k, fixed-T and $n \to \infty$ asymptotics. The large sample validity of this inference only requires that extreme value theory in (3) becomes an accurate approximation as $n \to \infty$ for each $\mathbf{X} = \mathbf{X}_t$, and that the $\mathbf{X}_t$ are asymptotically independent.[3] The values of (μ,σ) typically depend on n under such asymptotics; however, the inference procedures that we discuss are uniformly valid in $(\mu,\sigma) \in \mathbb{R}\times(0,\infty)$ in the small sample problem of observing

$$\mathbf{X}_t \sim \text{ i.i.d. } \text{GEV}_k(\mu,\sigma,\xi),\ t = 1,\ldots,T \tag{5}$$

so this potential dependence is innocuous.

The remaining challenge is how to conduct valid inference in the small sample problem of observing (5) for a given pair of (k,T). Suppose we want to test $H_0 : (\mu,\sigma,\xi) \in \Theta_0 \subset \mathbb{R}^3$, for some given Θ_0.[4] For instance, Θ_0 could restrict ξ to take on the value ξ_0. Let $\Theta_1 \subset \mathbb{R}^3$ be the unrestricted parameter space, which we choose to equal $\{(\mu,\sigma,\xi) : \sigma > 0,\ \xi > -0.99\}$ in our applications. We suggest using the generalized likelihood ratio statistic

$$\text{LR} = \sup_{(\mu,\sigma,\xi)\in\Theta_1} \ln\prod_{t=1}^{T} f_{\text{GEV}}(\mathbf{X}_t|\mu,\sigma,\xi) - \sup_{(\mu,\sigma,\xi)\in\Theta_0} \ln\prod_{t=1}^{T} f_{\text{GEV}}(\mathbf{X}_t|\mu,\sigma,\xi). \tag{6}$$

This generalizes the test statistic employed in Müller and Wang (2017) for $T = 1$ for inference about the quantiles and tail expectation of the underlying GP distribution to arbitrary Θ_0 and $T \geq 1$. For a test based on LR to be valid, it must control size for all values in Θ_0.[5] All of

[3]Formally, we need that the total variation distance between the distribution of $\{\mathbf{X}_t\}_{t=1}^{T}$ and T independent copies of the $\text{GEV}_k(\mu_n,\sigma_n,\xi)$ distribution converges to zero for some sequence (μ_n,σ_n) as $n \to \infty$. See Chapter 5 of Reiss (1989) for sufficient conditions.

[4]The set Θ_0 could depend on n, but we omit such dependence in our notation.

[5]The results in Bücher and Segers (2017) suggest that as $T \to \infty$, the critical values converge to their usual chi-square quantiles, but we seek inference that is valid for small T.

our choices of Θ_0 are such that the distribution of LR only depends on (μ, σ, ξ) through ξ, so ξ is the only relevant nuisance parameter. Müller and Wang (2017) restrict attention to values of ξ in $[-0.5, 0.5]$ and apply a simple sup bound, that is, their critical value cv_{sup} is chosen so that the $1-\alpha$ quantile of LR is smaller than cv_{sup} for all $\xi \in [-0.5, 0.5]$. Our empirical results below are valid for the larger range $[-0.5, 1.5]$, and for some Θ_0, a corresponding sup bound on the critical value induces tests to be heavily undersized for some ξ.

So instead, we let the logarithm of the critical value be a quadratic function of $\hat{\xi}$, the MLE of ξ only constrained to be larger than -0.99, and numerically choose the coefficients so that the resulting test is close to unbiased for the range of ξ we consider. See Appendix B for details on the determination of these coefficients. The end result is a test that rejects if $\text{LR} \geq \exp(a_0 + a_1\hat{\xi} + a_2\hat{\xi}^2)$ or, equivalently, $\text{LR}^{\hat{\xi}\text{-adjusted}} \geq 1$, where $\text{LR}^{\hat{\xi}\text{-adjusted}} = \text{LR} \,/\, \exp(a_0 + a_1\hat{\xi} + a_2\hat{\xi}^2)$ is the $\hat{\xi}$-adjusted LR statistic and the values of (a_0, a_1, a_2) are determined using the method described in Appendix B

We consider four null hypotheses. The first simply specifies ξ, $H_0 : \xi = \xi_0$.[6] We thus conduct inference about the tail index of the GEV distribution, which is also the tail index of the underlying F. In contrast to standard inference about ξ (such as based on the Hill (1975) or Pickands (1975) estimator), the validity of this inference does not require k to diverge. In practice and with k small, it thus allows testing whether the extreme right tail of F with approximate mass k/n is characterized by a tail index equal to ξ_0. Second, we test $H_0 : q_{0.9}(\mu, \sigma, \xi) = q_0$, where $q_{0.9}(\mu, \sigma, \xi) = \mu + \sigma((-\ln 0.9)^{-\xi} - 1)/\xi$ is the 90% quantile of the $\text{GEV}_1(\mu, \sigma, \xi)$ distribution. Under the approximation (3), by definition, sample maxima from the same population F exceed $q_{0.9}(\mu, \sigma, \xi)$ only 10% of the time. It also corresponds to the $1 - h/n$ quantile of F, where h solves $0.9 = e^{-h}$, so $h \approx 0.1$ (cf. Müller and Wang (2017)). In our empirical work, we invert the tests $\xi = \xi_0$ and $q_{0.9} = q_0$ to construct confidence intervals for ξ and $q_{0.9}$. For the third null hypothesis, we test $H_0 : \mu = \sigma/\xi$, $\xi \geq \varepsilon$ for some small $\varepsilon > 0$, which we choose to be $\varepsilon = 0.03$ in our applications. This corresponds to $\nu = \omega/\xi$, i.e. the validity of the Pareto (rather than shifted Pareto) approximation discussed at the end of Section 2.1. Finally, we test the constraint corresponding to Zipf's law in the extreme right tail $H_0 : \mu = \sigma/\xi$, $\xi = 1$, that is, we add $\xi = 1$ to the previous null hypothesis. Gabaix (1999), for instance, argues that for cities a tail index of $\xi = 1$ is to be expected.

[6] For this test, the distribution of the LR statistic does not depend on any nuisance parameters, so we can directly use the critical value computed by simulation.

2.3 Extreme Value Theory for Exceedances

For now we considered the behavior of the largest k observations of an i.i.d. sample $W_1, \ldots, W_n$. Suppose instead that we observe all W_i that exceed a given threshold τ, that is, we observe $\mathbf{Y} = \{W_i : W_i \geq \tau\}$ with $k_\tau \geq 0$ elements. Let $\mathbf{Y}^s$ have the same elements as $\mathbf{Y}$ sorted descendingly, and note that there are $k_\tau!$ equally likely $\mathbf{Y}$ vectors that map to a given $\mathbf{Y}^s$. The vector $\mathbf{Y}^s$ behaves just like the vector of largest observations $\mathbf{X}$, but with a sample-dependent value of $k = k_\tau$, where k_τ is such that $X_{k_\tau} \geq \tau$ and $X_{k_\tau+1} < \tau$. If we let τ depend on n such that the expected number of exceedances remains bounded, the approximation (3) therefore induces an approximate distribution for $\mathbf{Y}$. The density of $\mathbf{Y} \in \mathbb{R}^{k_\tau}$ under this approximation is given by

$$\begin{aligned} f_\tau(\mathbf{y}|\mu,\sigma,\xi) &= \frac{1}{k_\tau!}\int_{-\infty}^{\tau} f_{\text{GEV}}((\mathbf{y}^s, x_{k_\tau+1})|\mu,\sigma,\xi)dx_{k_\tau+1} \\ &= \frac{1}{k_\tau!\sigma^{k_\tau}}\exp\left[-\left(1+\xi\frac{\tau-\mu}{\sigma}\right)^{-1/\xi} - (1+1/\xi)\sum_{j=1}^{k_\tau}\log(1+\xi\frac{y_i-\mu}{\sigma})\right] \quad (7) \end{aligned}$$

where the first equality reflects the fact that all we know about $X_{k_\tau+1}$ is that it is smaller than τ, and the second equality is from a direct calculation using the form of the GEV density (4).

The density (7) can be re-expressed into a more intuitive form by introducing two new parameters $\lambda, \psi > 0$. With $\sigma = \psi\lambda^\xi$ and $\mu = \tau + \psi(\lambda^\xi - 1)/\xi$, the density in (7) can be rewritten as

$$f_\tau(\mathbf{y}|\lambda,\psi,\xi) = \frac{\lambda^{k_\tau}e^{-\lambda}}{k_\tau!}\prod_{i=1}^{k}\frac{1}{\psi}(1+\xi\frac{y_i-\tau}{\psi})^{-(1+1/\xi)}, \quad (8)$$

which is recognized as the product of a Poisson density for the number of exceedances k_τ with parameter $\lambda = (1+\xi(\tau-\mu)/\sigma))^{-1/\xi} = n(1-F_{\text{GP}}(\tau))$, and a GP density for the k_τ exceedances $Y_i - \tau$ with parameter $(0,\psi,\xi)$.

The generalization to a panel of independent exceedances with thresholds τ_t is immediate, where in the approximate model

$$\mathbf{Y}_t \sim \text{independent with density } f_{\tau_t}(\mathbf{y}|\mu,\sigma,\xi),\ t = 1,\ldots,T \quad (9)$$

with corresponding likelihood $\prod_{t=1}^{T} f_{\tau_\tau}(\mathbf{Y}_t|\mu,\sigma,\xi) = \prod_{t=1}^{T} f_{\tau_t}(\mathbf{Y}_t|\lambda_t,\psi_t,\xi)$, and the notation emphasizes that (μ,σ) are time-invariant while (λ_t,ψ_t) vary with τ_t.

2.4 Inference Using Exceedances

Inference can again be based on the generalized likelihood ratio statistic, analogous to (6). There is one complication, though: The distribution of LR for exceedances under the four null hypotheses considered in Section 2.2 not only depends on ξ, but for a constant threshold τ also on λ in the Poisson-GP parameterization (8), and for a time varying τ_t on all three parameters. It is impossible to numerically bound the relevant quantile of the null distribution of LR uniformly in $\lambda > 0$, as one would need to consider values arbitrarily close to zero, so $\mathrm{P}(k_\tau > 0) \to 0$, and arbitrarily large values of λ, so that lead to $n_t \to \infty$.

Instead, we suggest using a Bonferroni approach: First use the likelihood ratio statistic of $H_0 : (\mu, \sigma, \xi) = (\mu_0, \sigma_0, \xi_0)$ to form a 99% confidence set $\mathcal{S}_0 \subset \mathbb{R}^3$ for the GEV parameter (μ, σ, ξ)—since there are no nuisance parameters, the null distribution is known for any value of (μ_0, σ_0, ξ_0). Then for each of the four hypothesis tests of interest, numerically obtain a critical value adjustment as a function of $\hat{\xi}$ so that we obtain a 4% level test uniformly over $\mathcal{S}_0$. By construction, this yields valid 5% level tests in the small sample problem (9).

2.5 Illustrations

Table 1 summarizes results for each of the empirical examples discussed in the Introduction. We highlight a few results from the table.

Weather-Disaster Damages. The first row of the table shows the results using the damage data plotted in Figure 1. The distribution of damage extremes is characterized by a large tail index, with a 95% confidence interval spanning 0.66 to 1.06 and monthly damages that exceed \$5 billion in one out of ten months on average (that is, $q_{0.9} > 5$). The null $\mu = \sigma/\xi$ is rejected, indicating that damage extremes are not well-described by the Fréchet distribution. Moreover, as is evident in Figure 1 and as we will see more formally in Section 3, there is a marked increase over the sample period in the number of damages that exceed \$1 billion each month and in the value of damage extremes. One potential explanation for these increases is that the value of assets potentially affected by weather disasters have increased over time, so damages rise even absent changes in the severity of weather. To account for this, researchers routinely normalize damages to adjust for the value of assets at risk from weather events (e.g., Pielke and Landsea (1998) and Pielke, Gratz, Landsea, Collins, Saunders, and Musulin (2008)). With this in mind, the second row of the table repeats the empirical work using

Table 1: Constant Parameter Results

	MLE				95% Confidence Interval		$LR^{\hat{\xi}\text{-adjusted}}$ for null:	
	ξ	σ	μ	$q_{0.9}$	ξ	$q_{0.9}$	$\mu = \sigma/\xi$	$\mu = \sigma/\xi$ and $\xi = 1$
Weather Damages	0.84	0.86	0.45	6.22	0.66 to 1.06	5.11 to 7.82	1.46	1.25
Weather Damages (normalized)	1.00	0.87	0.90	8.25	0.80 to 1.22	6.68 to 10.61	0.01	0.02
Largest Returns	0.06	0.54	2.43	3.73	-0.04 to 0.15	3.54 to 3.99	2.27	64.8
Smallest Returns	0.28	0.68	-2.85	-4.96	0.17 to 0.41	-5.53 to -4.58	0.08	56.1
City Size	0.65	1.34	2.02	8.89	0.43 to 0.93	3.92 to 31.43	0.18	1.57
Firm Size	0.42	0.48	1.00	2.81	0.23 to 0.69	1.56 to 7.92	0.90	3.49

Notes: The 5% critical values are 1.0 for test statistics shown in the final two columns. For the row labeled Smallest Returs, the value shown in the $q_{0.90}$ column is the $q_{0.10}$ quantile (because smallest returns are negative).

damages that are normalized by the real quantity of the U.S. capital stock. (See Appendix A for details of the data construction.) The table suggests that extremes of normalized damages are well described by the Fréchet distribution with $\xi = 1.0$. But, to preview results in Section 3, normalized damages continue to exhibit marked increases over the sample period. (Also see Figure 5 in Appendix A.)

Returns. The data plotted in Figure 2 suggests an asymmetry in the distribution of positive and negative extremes, and this visual impression is quantified in Table 1 which reports a larger tail index for negative extremes and a more extreme 90th quantile. Table 1 also shows that the $\mu = \sigma/\xi$ constraint is rejected for large returns, suggesting a departure from the Fréchet distribution underlying commonly used estimators for the shape parameter ξ.

City and Firm Sizes. To our eyes the most notable result is that ξ is less than unity for both city and firm sizes. While the null that $\mu = \sigma/\xi$ is not rejected, the null that includes $\xi = 1.0$ is rejected, so these extremes are inconsistent with Zipf's law. These results use the largest $k = 30$ cities and firms, but the conclusions hold for cities for values of k between 20 and 100 and for firms with k between 10 and 100.

A caveat for these results is that they are predicated on independence of the samples over time, here with constant GEV parameters and in Section 3 after conditioning on the

potentially time varying GEV parameters values. For monthly weather damages and 6-month samples of daily returns, temporal dependence is arguably sufficiently weak as to have little meaningful effect on the validity of the tests and confidence intervals, as suggested by the analysis in Leadbetter (1983) discussed above. However, there is much stronger dependence for the city sizes, where 11 of the largest 30 cities in 1900 were also among the 30 largest cities in 2020. As in the case with inference about a mean parameter, this positive correlation presumably acts like a reduction in the effective sample size T, suggesting that the tests using city size data are somewhat oversized. In Section 3 we study this in more detail in the context of a test for time-varying GEV parameters. Interestingly, while there is considerable overlap in the largest 30 cities over time, there is far less overlap in firm sizes where none of the largest 30 firms in 1950 (the first year our sample) remained in the top-30 in the final year, 2019.

3 Time Variation

If the data takes the form of a panel of extremes, that is, $T > 1$, a natural question is whether the parameters that characterize the extremes are stable over time. In this section, we develop a test of that null hypothesis, and suggest a Bayesian approach to the estimation of the parameter path in an unstable model.

3.1 Testing the Null Hypothesis of Time Invariant Parameters

It is entirely straightforward to introduce time varying parameters in the approximate distribution for the extremes discussed in Sections 2.1 and 2.3: Simply subscript the GEV parameters (μ, σ, ξ) by t, so that under the approximation induced by extreme value theory, the observations are given by

$$\begin{aligned} &\mathbf{X}_t \sim \text{independent GEV}_k(\mu_t, \sigma_t, \xi_t),\ t = 1, \ldots, T \text{ and} \\ &\mathbf{Y}_t \sim \text{independent with density } f_{\tau_t}(\mathbf{y}|\mu_t, \sigma_t, \xi_t),\ t = 1, \ldots, T \end{aligned} \tag{10}$$

for the k largest observations and the exceedances over τ_t, respectively. We seek a test of parameter constancy:

$$H_0 : (\mu_t, \sigma_t, \xi_t) = (\mu_1, \sigma_1, \xi_1),\ t = 2, \ldots, T. \tag{11}$$

One idea would be to use again the generalized likelihood ratio statistic for this purpose. However, this would require estimating an unrestricted set of GEV parameters for each t, and when k (or k_τ) is small, this parameter would be informed by a very small number of observations. To obtain an approach that works for all (k, T) pairs with $T > 1$, we instead rely on the test statistic developed by Nyblom (1989). He derived the locally best test of the null hypothesis of parameter constancy against the alternative of martingale-type time variation in a parametric model. For a model with time varying parameter $\theta = \theta_t$ and likelihood $\prod_{t=1}^{T} f(\mathbf{X}_t|\theta_t)$, the feasible version of Nyblom's (1989) test is given by

$$L_T = T^{-2} \sum_{t=1}^{T} \left(\sum_{l=1}^{t} S_l(\hat{\theta}) \right)' \hat{V}^{-1} \left(\sum_{l=1}^{t} S_l(\hat{\theta}) \right)$$

where $S_t(\theta) = \partial \ln f(\mathbf{X}_t)/\partial\theta$ are the scores, $\hat{\theta}$ is the MLE satisfying $\prod_{t=1}^{T} f(\mathbf{X}_t|\hat{\theta}) = \sup_\theta \prod_{t=1}^{T} f(\mathbf{X}_t|\theta)$, and $\hat{V} = -T^{-1} \sum_{t=1}^{T} \partial S_t(\theta)/\partial\theta'|_{\theta=\hat{\theta}}$ is an estimator of the average information. Under standard regularity assumptions and $T \to \infty$ asymptotics, Nyblom shows that $L \Rightarrow \int_0^1 (W(s) - sW(1))'(W(s) - sW(1))ds$, where W is a standard Wiener process of the same dimension as θ.

We do not want to rely on $T \to \infty$ asymptotics for the validity of our inference. So instead, we again numerically determine a critical value that is small sample valid under (10) and (11). For the largest k observations, a straightforward calculation shows that the distribution of L_T only depends on the GEV parameter through ξ, just like the LR statistic of Section 2.2. So we use the same critical value construction as discussed there. For the exceedances, we again find a more complicated dependence of the distribution of L_T on the GEV parameters, so we apply the same Bonferroni approach as discussed in Section 2.4.

3.2 Estimation of the Parameter Path

In addition to testing whether the parameters are stable in time, it is also of practical interest to estimate their paths. This is most naturally done in a Bayesian framework, which we now adopt.

3.2.1 Parametrizing the Sources of Time Variation

One approach is to simply form a prior on paths of the GEV parameters directly. We found it useful, however, to instead consider potential sources of time variation that stem from the underlying population and the number of cross sectional observations.

Recall from Section 2.1 that the GEV parameters (μ, σ, ξ) are related to the GP parameters (ν, ω, ξ) via $(\mu, \sigma, \xi) = (\omega\frac{n^\xi - 1}{\xi} + \nu, \omega n^\xi, \xi)$. Thus, (μ, σ) may vary in time because the underlying population F has a time varying GP tail, or because the number of observations n is time varying. To isolate these effects, consider the following reparameterization of (μ, σ, ξ) in terms of (ξ, α, s, m): Let $n = n_0\alpha$, where n_0 is of the same order as n, so $\alpha > 0$ captures the sample size relative to the baseline n_0. We have

$$\begin{aligned} \sigma &= (\omega n_0^\xi)\alpha^\xi = s\alpha^\xi \\ \mu &= \nu + \omega\frac{(n_0\alpha)^\xi - 1}{\xi} = m + s\frac{\alpha^\xi - 1}{\xi} \end{aligned} \tag{12}$$

with $s = \omega n_0^\xi$ and $m = \nu + (\omega/\xi)(n_0^\xi - 1)$.

We model time variation in (μ_t, σ_t) for $t = 1, \ldots, T$ by independent time variation in (α_t, s_t, m_t) via the mapping (12). Changes in α_t correspond to a time varying number of underlying observations n_t—for instance, in the damages data set, one could imagine that we observe more exceedances simply because there are more weather-related events, without any change in the damage distribution F. Time variation in (s_t, m_t), in contrast, corresponds to changes of the GP tail of the underlying population F_t.

3.2.2 Choice of Prior

Our prior on the paths (ξ, α, s, m) is

$$\begin{aligned} &\xi_t = \xi_{\min} + (\xi_{\max} - \xi_{\min})\tfrac{e^{\tilde{\xi}_t}}{1+e^{\tilde{\xi}_t}},\ \tilde{\xi}_t = \tilde{\xi}_0 + \gamma_\xi\eta_{\xi,t} \\ &\ln(\alpha_t) = \ln(\alpha_0) + \gamma_\alpha\eta_{\alpha,t} \\ &\ln(s_t) = \ln(s_0) + \gamma_s\eta_{s,t} \\ &m_t = m_0 + \gamma_m\eta_{m,t} \end{aligned} \tag{13}$$

where the four η_t's are mutually independent random walks with $\mathcal{N}(0, 1/T)$ increments and $\eta_0 = 0$, and $\xi_{\min} = -0.9$ and $\xi_{\max} = 2.1$. For the initial conditions, we choose a $\mathcal{N}(0, 1.5^2)$

prior for $\tilde{\xi}_0$, which induces a nearly flat prior for ξ_0 over the $[-0.5, 1.5]$ interval, $\ln \alpha_0 = 0$ (a normalization that associates n_0 with the time $t = 0$ value of n), and improper flat priors on $\ln s_0$ and m_0. These flat priors ensure that our Bayesian inference is invariant with respect to location and scale transformations of the data, so it doesn't matter, for instance, whether damages are measured in millions or billions of dollars.

The γ parameters govern the degree of time variation in the prior: The difference between the $t = T$ and $t = 0$ value of the (transformed) parameters has standard deviation γ. We use an exponential prior for γ with mean g. In particular, $g_\alpha = g_s = 0.5$, so that the standard deviation of the change in α and s is 0.5 log points over the sample period. We set $g_\xi = 1$, which implies that a one unit increase in $\tilde{\xi}$ would raise $\xi_0 = 0.5$ to $\xi_T = 0.73$. The parameter m is measured in the same units as the data, and to preserve the invariance, we set $g_m = s_0$.

3.2.3 Computing the Posterior

With our random walk prior on the transformed parameters, estimation of the parameter path is a filtering/smoothing problem in a state-space system with linear and Gaussian evolution for the state (13), but non-Gaussian measurements given in equation (10). This makes the determination of the posterior a difficult computational problem.

To address this difficulty, we take advantage of the recent development of powerful HMC methods. HMC improves sampling over other MCMC methods by also relying on information about the derivatives for each sampled parameter point. But evaluation of the derivatives is straightforward for the prior and likelihood induced by (13) and (10). We can thus treat the entire path $\{\xi_t, \alpha_t, s_t, m_t\}_{t=1}^T$ as part of the parameter in the HMC algorithm, and still obtain fast convergence despite the potentially very large parameter space.

For our empirical work, we rely on Stan, a popular HMC implementation that includes automatic differentiation. A remaining challenge is that the support of the observations (10) depends on the parameters. Stan's posterior exploration regularly visits such values, and simply "rejecting" them leads to poor convergence. We overcome this by smoothly extrapolating the likelihood beyond the support of the GEV distribution, followed by an importance sampling correction that restores validity of the sampler for the original likelihood.

Specifically, consider the likelihood (4). Dropping time subscripts to ease notation, we

have

$$\ln f_{\text{GEV}}(\mathbf{X}|\mu,\sigma,\xi) = -k\ln\sigma - \exp(h(Z_k)) + (1+\xi)\sum_{j=1}^{k} h(Z_j)$$

where $h(z) = \ln[(1+\xi z)^{-1/\xi}]$ and $Z_j = (X_j - \mu)/\sigma$. As $1+\xi z \to 0$, $h(z)$ diverges to $+\infty$ for $\xi > 0$, and $h(z)$ diverges to $-\infty$ for $\xi < 0$. The idea is to extend h to a function $\tilde{h} : \mathbb{R} \mapsto \mathbb{R}$ with domain equal to $\mathbb{R}$ by a linear extrapolation of $h(z)$ for all arguments for which $1+\xi z < \chi$, for some small threshold $\chi > 0$. This yields

$$\tilde{h}(z) = \begin{cases} -(1/\xi)\ln\chi - (z - \frac{\chi-1}{\xi})/\chi & \text{for } 1+\xi z < \chi \\ \ln[(1+\xi z)^{-1/\xi}] \text{ otherwise.} \end{cases}$$

For numerical stability, we select the threshold χ such that $|h(\chi)| = K$ or $\chi = \exp[-K|\xi|]$ for $K = 6$. The "extended support" extrapolation of the GEV likelihood then becomes $\ln \tilde{f}_{\text{GEV}}(\mathbf{X}|\mu,\sigma,\xi) = -k\ln\sigma - \exp(\tilde{h}(Z_k)) + (1+\xi)\sum_{j=1}^{k}\tilde{h}(Z_j)$.

Sampling from the kernel with f_{GEV} replaced by $\tilde{f}_{\text{GEV}}$ does not yield the desired posterior; but to correct this, all we have to do is to reweight the draws that rely on the extrapolated likelihood by the ratio $f_{\text{GEV}}/\tilde{f}_{\text{GEV}}$. In our applications it turns out that these weights are very close to unity for essentially all draws, so the reweighting ends up being negligible for practical purposes.

The likelihood of the exceedances (7) has a very similar form, so we can apply the same approach there.

With this in place, we find that Stan reliably estimates the posterior in seconds or at most a few minutes, even for $T = 522$ where the entire parameter vector including the parameter paths has more than 2000 elements.

3.2.4 Computing Bayes Factors

It is useful to compare alternative specifications for the time varying model by the corresponding Bayes factor. This can be readily computed from Stan's posterior draws of the two competing models using Meng and Wong's (1996) "simple identity."

Table 2: $\hat{\xi}$-adjusted Nyblom test statistic for null of constant parameters

Application	Test Statistic
Weather Damages	11.6
Weather Damages (normalized)	4.2
Largest Returns	1.6
Smallest Returns	1.4
City Size	1.1
Firm Size	0.7

Notes: The 5% critical values are 1.0.

3.3 Illustrations

Nyblom Tests for Stability. Table 2 shows the Nyblom parameter stability tests for each dataset. There is overwhelming evidence for instability for the parameters describing weather damages (normalized or not), and strong evidence of instability for the parameters describing returns. The results for city and firm sizes are less clear: the $k = 30$ results shown in the table reject stability at the 5% significance level for city sizes, but not for firm sizes. Results for larger values of k (not shown) suggest some evidence of instability for firm sizes, but generally don't reject the null of stability for city sizes.[7]

Estimated Parameter Paths. Bayes estimation of the TVP-GEV model (12)-(13) produces posterior paths for (α, m, s, ξ) and the resulting GEV parameters. Figure 4 shows a selected summary of these posteriors focusing on ξ and the 90th quantile $q_{0.9}$ of the (time varying) $\text{GEV}_1(\mu_t, \sigma_t, \xi_t)$ distribution. While the results in the figure speak for themselves, we highlight the three we found most interesting. First, damages associated with weather extremes have increased markedly over the 1980-2023 period. The estimated (posterior mean) value of $q_{0.9}$ increased from \$2.7 billion in 1980 to \$13.4 billion in 2013. For normalized damages, the increase was from \$6.6 billion to \$12.9 billion (in units of the 2023 capital stock). Second, while the constant-parameter results for returns showed an asymmetry in right and left tail risk

[7]As discussed in Section 2.5, there is considerable overlap in the list of the 30 largest U.S. cities over the sample period. This temporal dependence potentially affects the size of the $\hat{\xi}$-adjusted Nyblom test. To investigate this we carried out a simulation experiment Gabaix's (1999, Proposition 1) data generating process for a Pareto distribution. The results indicated that the $\hat{\xi}$-adjusted Nyblom test was somewhat *undersized*, suggesting that the test rejections for cities (and perhaps firms) are not spuriously generated by the sampling dependence.

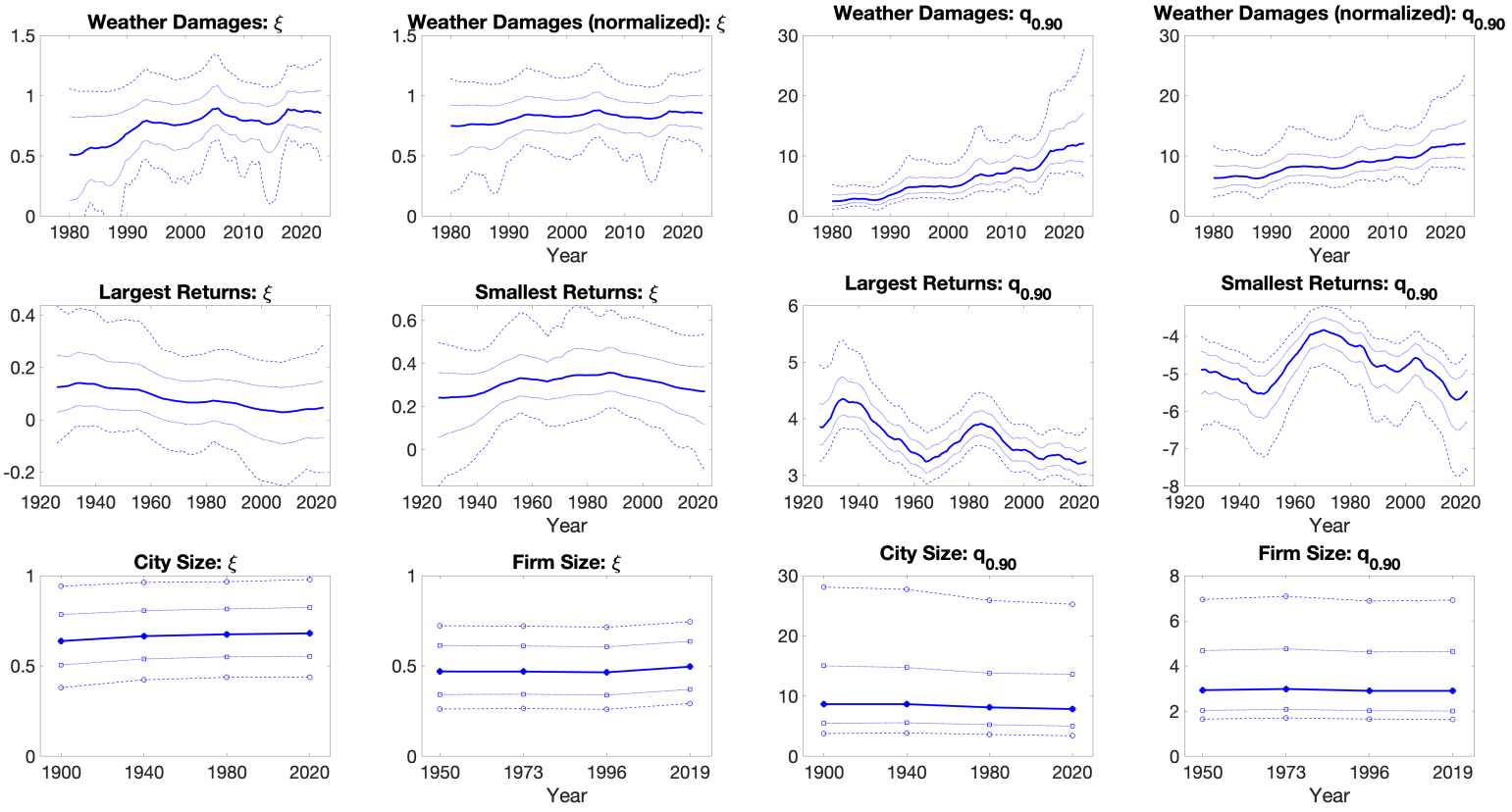
(where the magnitude of the extreme quantile for smallest returns was 1.2 standard deviations greater than for largest returns), Figure 4 shows important time variation in this asymmetry that ranges from a low of 0.4 in the mid-1970s to more than 2.5 standard deviations at the end of the sample. Third, in contrast to the results for weather damages and returns, there is little estimated time variation in the parameters for city and firm size and the time-varying parameter results in Figure 4 are similar to the constant parameter results shown in Table 1.

Sources of Time Variation. The TVP-GEV model (12) and (13) highlighted two potential sources of time variation: time variation in the underlying GP parameters (ν, ω, ξ) and time variation in the sample size n, where the former is captured by the parameters (m, s, ξ) and the latter is captured by the parameter α. To investigate the source of time variation in the GEV parameters we estimate three restricted TVP models and computed Bayes factors relative to the unrestricted TVP model. The first model imposes the restriction that all of the parameters (α, m, s, ξ) are constant, the second imposes the restriction that (m, s, ξ) are constant but allows α to vary, and in the third α was constant but (m, s, ξ) are allowed to vary.

Table 3 reports the logarithm of the Bayes factors for the three models. Three results stand out. First, as expected, the Bayes factors strongly prefer the time varying model relative to the constant parameter for weather damages and returns. Second, a potential explanation for the time variation in extreme weather damages is that the distribution of damages has not changed (that is (m, s, ξ) have remained constant) but there are simply more weather events (α has increased). The Bayes factors for models using normalized weather damages are consistent with this explanation. Finally, given the posterior results for city and firm size reported in Figure 4, it is not surprising to see that Bayes factors strongly prefer the constant parameter models to the general TVP-GEV model. But, a careful look at the results suggests that model with time varying α is preferred to the model with constant parameters for firm sizes.

4 Concluding Remarks

This paper studies inference about the distribution of extremes using panel data. The paper makes five contributions. First, it proposes tests for restrictions on the parameters of the Generalized Extreme Value (GEV) distribution characterizing a panel of extremes or ex-



Notes: 50th quantile (median) is shown a solid circle, 16.7th and 83.3th quantiles are open squares (dotted line) and 2.5th and 97.5th quantile are open circles (dashed line).

Figure 4: Posterior Quantiles for $\xi(t)$ and $q_{90}(t)$

Table 3: Log-Bayes Factors for alternative models (relative to benchmark TVP model)

Application	Time-varying parameters		
	None	α	ξ, s, m
Weather Damages	-55.2	-5.7	-0.8
Weather Damages (normalized)	-20.9	-7.6	0.1
Largest Returns	-9.9	-0.7	-0.6
Smallest Returns	-13.5	1.6	-1.7
City Size	4.1	4.0	0.1
Firm Size	2.1	2.8	0.2

Notes: Values are the logarithm of the Bayes factor for the model relative to the model that allows time variation in (α, ξ, s, m).

ceeedances. Second, it proposes a test for determining whether these parameters are constant through time. In both cases the tests are designed to be valid when the number of time periods, T and the number of extremes in each time period, k, might be small, thus avoiding commonly used large-k and large-T approximations. Third, the paper proposes a time varying parameter GEV model that focuses on two distinct sources of instability in the distribution of extremes: variation in the underlying population and variation in the sample size from which the extremes are chosen. The resulting TVP-GEV model is a hidden Markov model with a linear Gaussian state equation but a non-Gaussian measurement equation. The fourth contribution of the paper is to show how Bayes estimation of this nonlinear TVP-GEV model can be readily achieved using Hamiltonian Monte Carlo methods. The final contribution is an application of the methods to four data sets.

In the TVP-GEV model proposed here, suitably transformed versions of the GEV parameters evolve as Gaussian random walks. This random walk specification provides a simple but flexible method for tracking the evolution of parameters over the sample period. That said, alternative specifications might be preferred in some applications — for example, some parameters might evolve as functions of observables, follow alternative linear time series models, or be driven by non-Gaussian shocks. It is straightforward to modify the suggested HMC estimation approach to handle these alternatives.

The focus of this paper has been on testing for time variation and in-sample tracking of the potentially time varying parameters. Another interesting question involves forecasting:

how likely is it that an extreme of a certain magnitude or larger will be realized at some future date? Such questions are readily answered using the predictive distribution that arises as a by-product of the Bayes estimation of the model, although alternatives to the random walk evolution of parameters might be desirable in some applications.

A Data Appendix

A.1 Weather Disaster Damages

The data are from the *Billion-Dollar Weather and Climate Disaster* dataset described at https://www.ncei.noaa.gov/access/billions/. We use all of the events in the database except for droughts and wildfires (because the duration of these events can extend for many months). The normalized data are adjusted by the value of the real U.S. capital stock. We use the chain-type quantity index *Table 1.2. Chain-Type Quantity Indexes for Net Stock of Fixed Assets and Consumer Durable Goods* from the BEA Fixed Asset Tables. The data are annual observations from 1979-2022. We interpolate the annual series to monthly observations from 1980:1-2022:12 and extrapolate through 2023:6 using the value from 2022:12. Let A_t denote the value of the capital stock at time t and D_t denote the unadjusted value of damages. The normalized damages are $D_t^{norm} = D_t \times (A_t/A_{2023:6})$. The censoring threshold for the normalized damages is $\tau_t^{norm} = \tau \times (A_t/A_{2023:6})$, where τ is the censoring threshold for D_t, which is $\tau = \$1$ billion.

Figure 5 plots the damages (shown previously in Figure 1) along with the normalized damages.

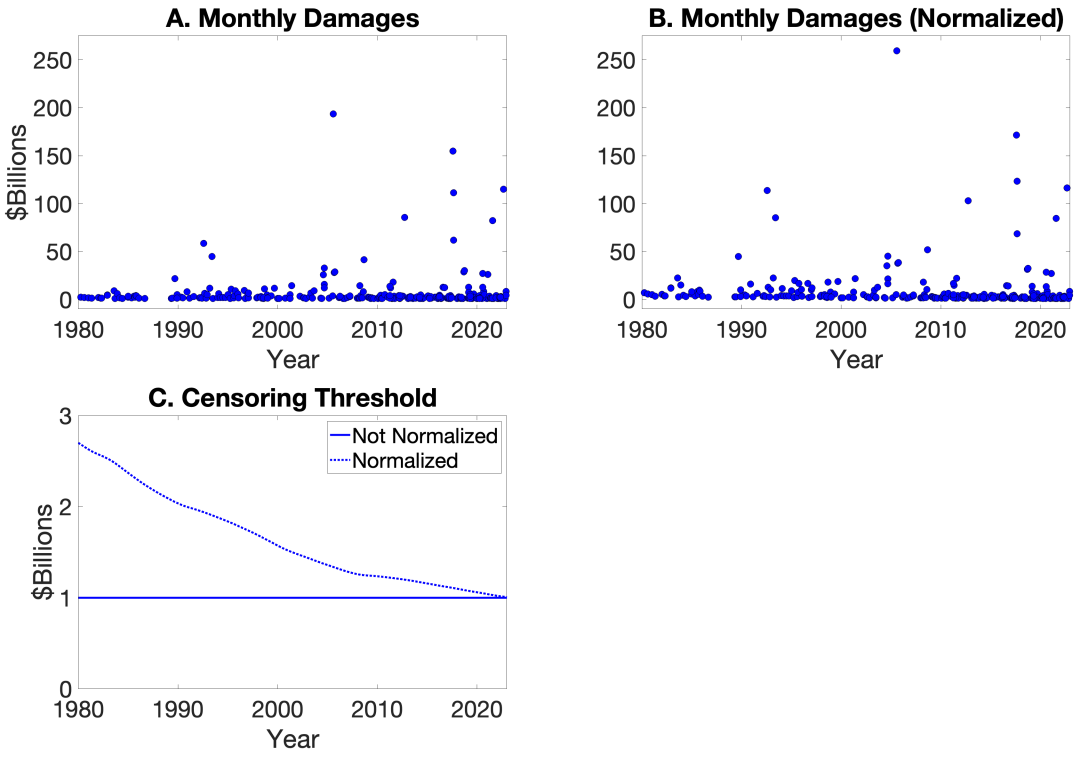


Figure 5: Damages and Normalized Damages

A.2 Returns

We begin with value-weighted daily returns from the CRSP SP500 index available from January 2, 1926 through December 30, 2022. These returns are plotted in panel A of Figure 6. We fit a GARCH (1,1) model using the full sample period and standardize the daily returns by subtracting the sample mean and dividing by the fitted GARCH standard deviation. These standardized values are plotted in panel B. Panels C and D (reported earlier in Figure 2) are the largest and smallest returns over non-overlapping 6-month periods.

A.3 Firm and City Size

Population values from the largest 100 cities for 1900, 1940, 1980 and 1920 are from https://www.biggestuscities.com/. These are divided by aggregate U.S. population. Employment levels are from Compustat for firms incorporated in the United States. These are

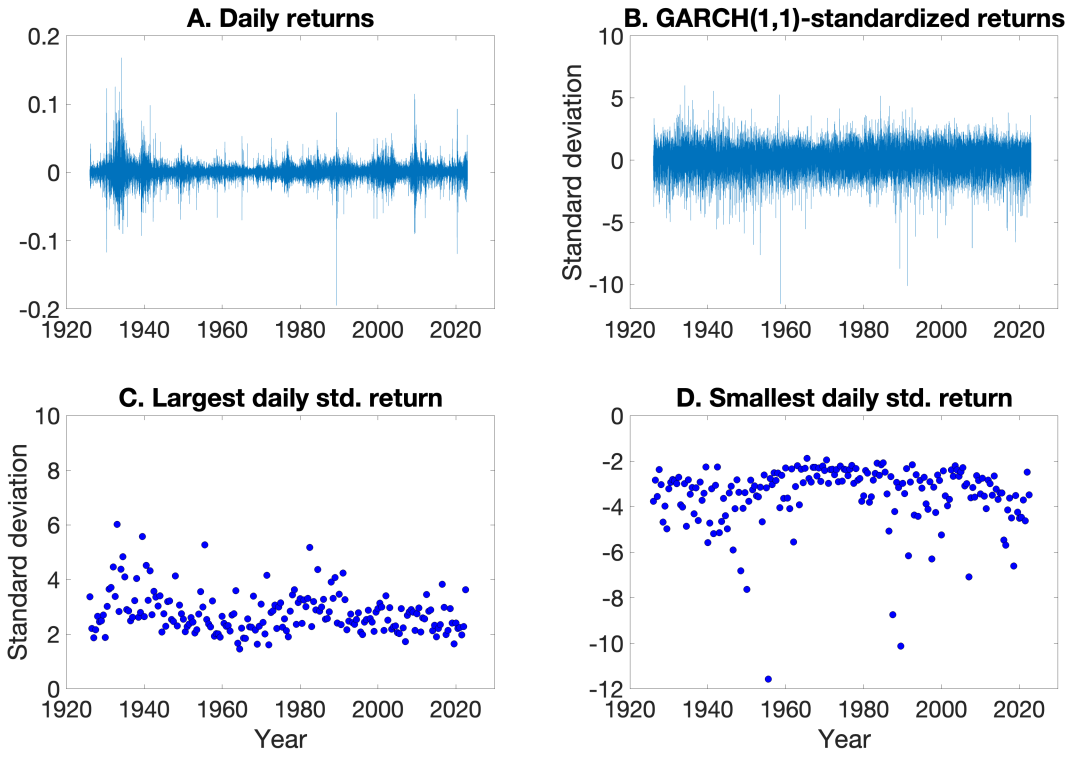


Figure 6: Daily Returns and 6-month Extremes

divided by total U.S. private employment (series USPRIV from the FRBSL FRED database).

B Computing the $\hat{\xi}$-adjusted LR Critical Values

B.1 LR Statistic for k Largest Observations

As noted in Section 2.2, the distribution of the LR statistic (6) only depends on ξ. We subdivide the possible values of $\xi \in [\xi_{\min}, \xi_{\max}]$ into an equal spaced grid with 10 values ξ_j, $j = 1, \ldots, 10$. Here $\xi_{\min} = -0.5$ for the hypothesis concerning $q_{0.9}$, $\xi_{\min} = \varepsilon = 0.03$ for the hypothesis $H_0 : \mu = \sigma/\xi$, and $\xi_{\max} = 1.5$ in both cases. We generate $N = 10,000$ independent draws $(\hat{\xi}_j^{(l)}, \mathrm{LR}_j^{(l)})$, $l = 1, \ldots, N$ of the maximum likelihood estimator $\hat{\xi}$ and the LR statistic for each value of ξ in the grid. The underlying kN independent exponential variables (cf. equation (2)) are held constant across the 10 values of ξ in these simulations. The coefficients

a_0, a_1 and a_2 are then determined by numerically solving

$$\min_{a_0,a_1,a_2} \sum_{j=1}^{10} \ell \left\{ \Psi \left[N^{-1} \sum_{l=1}^{N} \Phi \left(\frac{1 - \mathrm{LR}_j^{\hat{\xi}\text{-adjusted},(l)}}{h} \right) \right] - \Psi(0.95) \right\}$$

where $\ell(x) = 12x - \exp(12x) + 1$ is the linex-loss function with coefficient 12, Ψ is the logit function, Φ is the c.d.f. of a standard normal, and h is a bandwidth that we set equal to 0.3 times the difference between the 97th and 93th percentile of the distribution of the unadjusted LR statistic for the 5th value of ξ in the grid. The asymmetry of the linex loss function strongly penalizes overrejections. As a post-processing step, we further adjust a_0 so that the largest null rejection probability is exactly 5% on the grid.

We numerically explored potential overrejections for values of ξ that fall between grid values, and found them to be within Monte Carlo error.

B.2 LR Statistic for Exceedances

The initial level 99% confidence set for (μ, σ, ξ) is equal to

$$\mathcal{S}_0 = \{(\mu_0, \sigma_0, \xi_0) : \mathrm{LR}_{\mu_0,\sigma_0,\xi_0} < \mathrm{cv}^{0.99}_{\mu_0,\sigma_0,\xi_0}\} \subset \mathbb{R}^3$$

where $\mathrm{LR}_{\mu_0,\sigma_0,\xi_0} = \sup_{(\mu,\sigma,\xi)\in\Theta_1} \ln \prod_{t=1}^{T} f_{\tau_t}(\mathbf{Y}_t|\mu,\sigma,\xi) - \ln \prod_{t=1}^{T} f_{\tau_t}(\mathbf{Y}_t|\mu_0,\sigma_0,\xi_0)$ and $\mathrm{cv}^{0.99}_{\mu_0,\sigma_0,\xi_0}$ is the 99th percentile of the null distribution of $\mathrm{LR}_{\mu_0,\sigma_0,\xi_0}$. Our initial goal is to draw 50 points from $\mathcal{S}_0$ at random. To this end, we proceed as follows: We first compute the maximum likelihood estimator $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ and find the 99.5% level critical value $\overline{\mathrm{cv}}$ for the LR statistic of $H_0 : (\mu, \sigma, \xi) = (\mu_0, \sigma_0, \xi_0)$ where $(\mu_0, \sigma_0, \xi_0) = (\hat{\mu}, \hat{\sigma}, \hat{\xi})$. We then numerically determine $\mu_{\min}$ such that the LR statistic of $H_0 : (\mu, \sigma, \xi) = (\mu_0, \sigma_0, \xi_0)$ with $(\mu_0, \sigma_0, \xi_0) = (\mu_{\min}, \hat{\sigma}, \hat{\xi})$ is equal to $\overline{\mathrm{cv}}$. The idea here is that $\mathrm{cv}^{0.99}_{\mu_0,\sigma_0,\xi_0}$ does not vary a lot for relevant values of (μ_0, σ_0, ξ_0), so $\mathrm{cv}^{0.99}_{\mu_0,\sigma_0,\xi_0} \leq \overline{\mathrm{cv}}$, and $\mu_{\min}$ is smaller than what one would obtain had one used $\mathrm{cv}^{0.99}_{\mu_{\min},\hat{\sigma},\hat{\xi}}$. The same procedure is then applied to $\mu_{\max}$, $\sigma_{\min}$, $\sigma_{\max}$, $\xi_{\min}$ and $\xi_{\max}$, with the end result of a hypercube in $\mathbb{R}^3$ that contains $\mathcal{S}_0$. We then randomly draw values (μ_0, σ_0, ξ_0) uniformly in this hypercube, and check whether $\mathrm{LR}_{\mu_0,\sigma_0,\xi_0} < \mathrm{cv}^{0.99}_{\mu_0,\sigma_0,\xi_0}$, where $\mathrm{cv}^{0.99}_{\mu_0,\sigma_0,\xi_0}$ is obtained via simulation based on $N = 10,000$ draws, until we have collected 50 values for which we do not reject.

In the second step, we proceed just like in Section B.1, except that the "grid" now consists of these 50 values.

References

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. V. Reenen (2020): "The Fall of the Labor Share and the Rise of Superstar Firms," *Quarterly Journal of Economics*, 135, 645–709.

Axtell, R. L. (2001): "Zipf Distribution of U.S. Firm Sizes," *Science*, 293, 1818 – 1820.

Bücher, A., and J. Segers (2017): "On the maximum likelihood estimator for the Generalized Extreme-Value distribution," *Extremes*, 20, 839–872.

Chavez-Demoulin, V., P. Embrechts, and S. Sardy (2014): "Extreme-quantile tracking for financial time series," *Journal of Econometrics*, 181, 44–52.

Coles, S. (2001): *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

de Haan, L., and A. Ferreira (2007): *Extreme Value Theory: An Introduction*. Springer Science and Business Media, New York.

do Nascimento, F. F., D. Gamerman, and H. F. Lopes (2016): "Time-varying extreme pattern with dynamic models," *TEST*, 25, 131–149.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997): *Modelling extremal events for insurance and finance*. Springer, New York.

Gabaix, X. (1999): "Zipf's Law for Cities: An Explanation," *The Quarterly Journal of Economics*, 114, 739–767.

——— (2016): "Power Laws in Economics: An Introduction," *Journal of Economic Perspectives*, 30, 185–206.

Gaetan, C., and M. Grigoletto (2004): "Smoothing Sample Extremes with Dynamic Models," *Extremes*, 7, 221–236.

Gomes, M. I., and A. Guillou (2015): "Extreme Value Theory and Statistics of Univariate Extremes: a Review," *International Statistical Review*, 83, 263–292.

Hill, B. M. (1975): "A Simple General Approach to Inference about the Tail of a Distribution," *Annals of Statistics*, 3(5), 1163–1174.

Huerta, G., and B. Sansó (2007): "Time-varying models for extreme values," *Environmental and Ecological Statistics*, 14, 285–299.

Katz, R. W. (2015): "Weather and Climate Disasters," in *Extreme Value Modelling and Risk Analysis: Methods and Applications*, ed. by D. K. Dey, and J. Yan, pp. 439–469. CRC Press.

Kondo, I., L. Lewis, and A. Stella (2023): "Heavy tailed but not Zipf: Firm and establishment size in the United States," *Journal of Applied Econometrics*, 38, 767 – 785.

Leadbetter, M. R. (1983): "Extremes and local dependence in stationary sequences," *Probability Theory and Related Fields*, 65(2), 291–306.

Mao, G., and Z. Zhang (2018): "Stochastic tail index model for high frequency financial data with Bayesian analysis," *Journal of Econometrics*, 205, 470–487.

McNeil, A., and R. Frey (2000): "Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach," *Journal of Emprical Finance*, 7, 271–300.

Meng, X. L., and W. H. Wong (1996): "Simulating Ratios of Normalizing Constants via a Simple Identity: a Theoretical Exploration," *Stastistica Sinica*, 6, 831–860.

Müller, U. K., and Y. Wang (2017): "Fixed-k Asymptotic Inference about Tail Properties," *Journal of the American Statistical Association*, 112, 1334–1343.

Nakajima, J., T. Kunihama, and Y. Omori (2017): "Bayesian modeling of dynamic extreme values: extension of generalized extreme value distributions with latent stochastic processes," *Journal of Applied Statistics*, 44(7), 1248 – 1268.

Nakajima, J., T. Kunihama, Y. Omori, and S. Frühwirth-Schnatter (2012): "Generalized extreme value distribution with time-dependence using the AR and MA models in state space form," *Comput. Stat. Data Anal.*, 56, 3241–3259.

Nyblom, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

Pickands, III, J. (1975): "Statistical inference using extreme order statistics," *Annals of Statistics*, 3(1), 119–131.

Pielke, R., J. Gratz, C. W. Landsea, D. J. Collins, M. A. Saunders, and R. T. Musulin (2008): "Normalized Hurricane Damage in the United States: 1900-2005," *Natural Hazards Review*, 9, 29–42.

Pielke, R., and C. Landsea (1998): "Normalized Hurricane Damages in the United States: 1925-95," *Weather and Forecasting*, 13(3), 621–631.

Reiss, R.-D. (1989): *Approximate distributions of order statistics: with applications to non-parametric statistics*. Springer Verlag, New York.

Smith, A. B., and R. W. Katz (2013): "US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases," *Natural Hazards*, 67, 387–410.