Modeling panels of extremes

Submitted to the Annals of Applied Statistics

MODELING PANELS OF EXTREMES

BY DEBBIE J. DUPUIS[1,a], SEBASTIAN ENGELKE[2,b] AND LUCA TRAPIN[3,c]

[1]*Department of Decision Sciences, HEC Montréal, 3000, chemin de la Côte-Sainte-Catherine, Montréal (Québec) H3T 2A7, Canada* [a]*debbie.dupuis@hec.ca*

[2]*Research Center for Statistics, University of Geneva, Boulevard du Pont d'Arve 40, 1205 Geneva, Switzerland,* [b]*sebastian.engelke@unige.ch*

[3]*Department of Statistics, University of Bologna, Via delle Belle Arti 41, 40126 Bologna, Italy* [c]*luca.trapin@unibo.it*

Extreme value applications commonly employ regression techniques to capture cross-sectional heterogeneity or time-variation in the data. Estimation of the parameters of an extreme value regression model is notoriously challenging due to the small number of observations that are usually available in applications. When repeated extreme measurements are collected on the same individuals, i.e., a panel of extremes is available, pooling the observations in groups can improve the statistical inference. We study three data sets related to risk assessment in finance, climate science, and hydrology. In all three cases, the problem can be formulated as an extreme value panel regression model with a latent group structure and group-specific parameters. We propose a new algorithm that jointly assigns the individuals to the latent groups and estimates the parameters of the regression model inside each group. Our method efficiently recovers the underlying group structure without prior information, and for the three data sets it provides improved return level estimates and helps answer important domain-specific questions.

1. Introduction. Extreme value theory provides the basis for the estimation of probabilities and quantiles associated with extreme events (Coles, 2001). These quantities are used as inputs for managerial and policy decisions to prevent or limit the damage caused by extreme negative outcomes. Applications of extreme value theory commonly employ regression techniques to capture distributional variations across individuals and over time in datasets where covariates are available. For example, spatial applications model the parameters of the extreme value model as a function of geographical characteristics to explain heterogeneous behavior across stations (Davison, Padoan and Ribatet, 2012) and climate applications typically include time as a covariate in the extreme value model to capture the effect of a changing climate (Katz, 2013). When repeated extreme measurements are available for several individuals, we may speak of a panel of extremes. In particular, we will consider random variables $Y_{i,t}$ for individuals $i = 1, \dots, N$ and time points $t = 1, \dots, T$ that follow generalized extreme value distributions (Coles, 2001) with parameters depending on a covariate vector $\mathbf{X}_{i,t} \in \mathbb{R}^K$ in a parametric way. In this case, practitioners would like to pool all observations to estimate the extreme value regression model. This idea of borrowing strength is very attractive in extreme value applications as the number of observations for the analysis is inherently small. It does present a practical challenge however as it requires observations to be homogeneous with respect to the model parameters.

The naive approach pools all available individuals into one group, regardless of the underlying structure. This may hide large differences in the model parameters and lead to misguided inference and sub-optimal quantile estimates. More often, empirical analyses rely on *ad hoc* strategies to partition the individuals into homogeneous groups. For example, in spatial applications one may select groups of geographically close stations as homogeneous

Keywords and phrases: clustering, extreme value regression, panel data, risk assessment.

regions (Asadi, Davison and Engelke, 2015; Overeem, Buishand and Holleman, 2008); meteorological applications partition the data based on climatological similarity (Alila, 1999) or physiographical similarity and meteorological factors (Cheng, 1998); while financial applications partition the data based on economic sectors, business lines (Mhalla, Hambuckers and Lambert, 2020) or type of losses (Hambuckers and Kneib, 2021). While this approach introduces flexibility in the extreme panel regression model, it requires *a priori* domain knowledge. The latter might not be available, or may be inconsistent with the data generating mechanism.

In this paper, we study three data sets where identification of the correct group structures is crucial for accurate statistical inference. The first data set is for a financial study involving 48 assets where an extreme value regression with time-varying asset-specific covariates is used to tie financial losses to risk factors relevant to generate stress test scenarios. This is a critical tool to track firms' exposures to adverse extreme market events. Pooling across all assets allows estimating the regression model with few observations per asset, but yields poor estimates of the relationship between financial losses and risk factors. Standard practice in finance would group assets based on the Standard Industrial Classification (SIC), but this might not fully explain the heterogeneity (Oh and Patton, 2020). The second study investigates the effect of climate change on extreme temperatures in the U.S. Midwest. Data are for 127 locations where an extreme value regression (Zwiers and Kharin, 1998; Wang et al., 2016) models the impact of the global temperature anomaly (a common time-varying variable) on minimum temperatures, controlling for time-invariant location-specific covariates. The sought-after global mean/local extreme relation can be inferred at a single location, but short series yield large standard errors. An analysis based on homogeneous groups of locations could reduce standard errors. The third data set on flood risk assessment is based on 31 gauging stations in the Danube river basin where an extreme value regression with time-invariant station-specific characteristics is used to model spatial heterogeneity. Risk analysis based only on individual stations is insufficient since short record length would lead to large uncertainty of return level estimates. There are many methods to construct homogenous groups of stations (e.g., Merz and Blöschl, 2005; Asadi, Engelke and Davison, 2018), but many of them are heuristic or require domain knowledge.

While the three data sets come from different areas and exhibit domain specific challenges, the general structure of the data can be cast as a panel regression problem for extremes. Throughout the paper, we assume that the individuals are partitioned according to an unknown latent group structure and that the extreme value regression model presents group-specific parameters. Our aim is to make joint inference on the group structure and the model parameters. This requires finding the correct number of groups and the right assignment of the individuals. A full-likelihood approach would require estimating the model parameters and the group assignment concurrently, but this problem is computationally infeasible. To solve this issue, we develop a novel expectation-maximization (EM) algorithm that is able to estimate the model parameters while uncovering the latent group structure in a data-driven way. In this perspective, we add to a growing literature on identifying latent group structures in panel data (Su, Shi and Phillips, 2016; Gu and Volgushev, 2019; Oh and Patton, 2020; Wang and Su, 2021). We also contribute to recent literature on extreme value clustering methods for spatial data (Carreau, Naveau and Neppel, 2017; Reich and Shaby, 2019; Rohrbeck and Tawn, 2021; Vignotto, Engelke and Zscheischler, 2021).

Our new methodology offers a solution to the challenges encountered in the three data sets. Comparing our data-driven group structure with groupings based on domain knowledge, we find that our approach always yields superior predictions and offers more reliable inference. In the first study, the SIC grouping only partially explains the heterogeneity in the assets, and our model with data-driven group assignments produces superior risk estimates. In the

second study, our grouped panel model yields more precise estimates, identifying locations where extreme temperatures are more severely impacted by the global temperature anomaly and crop yields are more threatened. In the third study, our optimal panel model discovers coherent spatial similarity without using domain knowledge, yields smaller BIC values than grouping based on domain knowledge, and provides a good fit even for stations with a lot of missing data.

The remainder of the paper is structured as follows. In Section 2, after a brief introduction to extreme value theory, we present our panel model for extremes. Section 3 studies the finite sample properties of the proposed EM algorithm. In Sections 4 to 6, we use our panel model for extremes in our three studies to provide superior estimates and more reliable inference. Section 7 concludes. The Appendix contains additional results.

2. A panel model for extremes.

2.1. *Extreme Value Theory.* Let $Z_1, \dots, Z_s$ be a sample of independent observations of a distribution F and define the sample block maximum $Y^{(s)} = \max\{Z_1, \dots, Z_s\}$, where s is called the block size. The Fisher–Tippett–Gnedenko theorem (e.g., Embrechts, Klüppelberg and Mikosch, 1997, Theorem 3.2.3) states that if there exist sequences of normalizing constants a_s and b_s such that the normalized $Y^{(s)}$ converges in distribution to a non-degenerate limit distribution G,

$$\lim_{s\to\infty} \mathbb{P}\left(\frac{Y^{(s)} - a_s}{b_s} \le y\right) = G(y), \tag{1}$$

then G must be the generalized extreme value (GEV) distribution

$$H(y \mid \mu, \sigma, \xi) = \begin{cases} \exp\left\{-\left(1+\xi\frac{y-\mu}{\sigma}\right)_+^{-1/\xi}\right\}, & \xi \neq 0, \\ \exp\left\{-\exp\left(-\frac{y-\mu}{\sigma}\right)\right\}, & \xi = 0, \end{cases} \quad y \in \mathbb{R}, \tag{2}$$

where $x_+ = \max(0, x)$. In this case F is said to be in the max-domain of attraction of the GEV distribution H. The parameters $\mu \in \mathbb{R}$, $\sigma > 0$ and $\xi \in \mathbb{R}$ are the location, scale and shape parameters, respectively. The shape is the most important parameter since it characterizes the heaviness of the tail of the distribution F: if $\xi > 0$ then F is heavy tailed (e.g., Cauchy, Pareto) and the GEV is a Fréchet distribution; if $\xi = 0$, then F is light-tailed (e.g., Gaussian, exponential) and the GEV is a Gumbel distribution; if $\xi < 0$ then F has a finite upper endpoint (e.g., uniform, beta) and the GEV is a Weibull distribution; see Coles (2001) for details.

The limiting result (1) holds under a very mild assumption on the tail of F, which is satisfied for almost all relevant continuous distributions. Since the GEV distribution is the only possible limit for sample maxima, it is an asymptotically motivated model for the distribution of $Y^{(s)}$ for finite values of s. Suppose we have N independent observations $Y_1^{(s)}, \dots, Y_N^{(s)}$ of the block maximum $Y^{(s)}$. We use maximum likelihood estimation to obtain the parameters of the GEV distribution that best approximate the distribution of $Y^{(s)}$,

$$(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = \arg\max_{\mu,\sigma,\xi} \sum_{i=1}^{N} \log h(Y_i^{(s)} \mid \mu, \sigma, \xi), \tag{3}$$

where $\log h(\cdot \mid \mu, \sigma, \xi)$ is the log-likelihood of the GEV distribution

$$\log h(y \mid \mu, \sigma, \xi) = -\log(\sigma) - \left(1 + \frac{1}{\xi}\right)\log\left(1 + \xi\frac{y-\mu}{\sigma}\right) - \left(1 + \xi\frac{y-\mu}{\sigma}\right)^{-1/\xi}. \tag{4}$$

The maximum likelihood estimator is consistent and asymptotically normal if $\xi > -1/2$ (Smith, 1985; Bücher and Segers, 2017). Under mild conditions on the dependence structure

of a stationary time series, the GEV also emerges as the only possible non-degenerate limiting distribution for normalized maxima of blocks of observations from this series (Leadbetter, Lindgren and Rootzén, 1983). Asymptotic properties of the maximum likelihood estimator continue to hold in this setting (Bücher and Segers, 2018).

In applications, the GEV distribution is often fitted to observations of $Y^{(s)}$ to estimate return levels. The p-quantile of the GEV distribution is

$$Q^p(\mu,\sigma,\xi) = \begin{cases} \mu - \dfrac{\sigma}{\xi}\left[1 - \{-\log(p)\}^{-\xi}\right], & \xi \neq 0, \\ \mu - \sigma\log\{-\log(p)\}, & \xi = 0, \end{cases} \tag{5}$$

and the Sth return level, for $S > 1$, is defined as the $(1 - 1/S)$-quantile. It represents the level that is expected to be exceeded on average once in S time periods, where the time period corresponds to the block length s. For instance, if $Y^{(s)}$ represents a yearly maximum, then $RL^S(\mu,\sigma,\xi) = Q^{(1-\frac{1}{S})}(\mu,\sigma,\xi)$ is the S-year return level.

2.2. *A panel GEV regression model.* Panel data studies have flourished over the last years (Hsiao, 2007). Some of the advantages of panel data compared to cross-sectional and time series data are: more efficient estimates of the model parameters; the possibility to test more complicated models; controlling the impact of omitted variables (fixed effects); generating more accurate predictions by pooling information across individuals. Nowadays, static and dynamic panel models are available for linear regression (Hsiao, 2014), count data regression (Cameron and Trivedi, 2015), discrete choice models (Greene, 2009), and volatility models (Pakel, Shephard and Sheppard, 2011). We consider a panel of maxima $Y_{i,t}$, with $i = 1,\dots,N$ and $t = 1,\dots,T$, extracted for N individuals in T blocks. We omit the superscript for the block size s, but still assume that $Y_{i,t}$ is a block maximum and can be reasonably approximated by a GEV distribution. We further let the GEV model parameters depend on a covariate vector $\mathbf{X}_{i,t} \in \mathbb{R}^K$ measured for each ith individual in each tth block. The panel GEV regression model is defined as

$$Y_{i,t} \mid \mathbf{X}_{i,t} \sim H(y \mid \mu_{i,t}, \sigma_{i,t}, \xi_{i,t}), \tag{6}$$

where the model parameters depend on the covariates as

$$\begin{aligned} \mu_{i,t} &= e_\mu\left(\boldsymbol{\kappa}^\top \mathbf{X}_{i,t}\right), \\ \sigma_{i,t} &= e_\sigma\left(\boldsymbol{\gamma}^\top \mathbf{X}_{i,t}\right), \\ \xi_{i,t} &= e_\xi\left(\boldsymbol{\delta}^\top \mathbf{X}_{i,t}\right), \end{aligned}$$

where $\boldsymbol{\theta} = (\boldsymbol{\kappa},\boldsymbol{\gamma},\boldsymbol{\delta}) \in \boldsymbol{\Theta} \subset \mathbb{R}^P$ is a vector of regression parameters, and e_μ, e_σ, and e_ξ are suitable link functions that can be adapted to the requirements in applications. As we are not interested in the cross-sectional dependence structure in $\mathbf{Y}_t = (Y_{1,t},\dots,Y_{N,t})$, we can make inference on the marginals by estimating the model parameters by quasi maximum likelihood (QML), pooling information across individuals, i.e.,

$$\widehat{\boldsymbol{\theta}}_{QML} = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{N}\sum_{t=1}^{T} \log h(Y_{i,t} \mid \mu_{i,t}, \sigma_{i,t}, \xi_{i,t}) \tag{7}$$

where the log-likelihood of the GEV distribution is defined in (4). Under the usual regularity conditions and as $T \to \infty$ (see Chandler and Bate, 2007) we have

$$\widehat{\boldsymbol{\theta}}_{QML} \xrightarrow{d} N(\boldsymbol{\theta}, \mathbf{H}^{-1}\mathbf{V}\mathbf{H}^{-1})$$

where $\mathbf{H}$ is the expected Hessian of the log-likelihood and $\mathbf{V}$ is the covariance matrix of the score evaluated at the true parameter. Consistent estimates of these two quantities can be obtained as

$$\widehat{\mathbf{H}} = \sum_{t=1}^{T}\sum_{i=1}^{N} \frac{\partial \mathbf{s}_{it}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\widehat{\boldsymbol{\theta}}_{QML}},$$

$$\widehat{\mathbf{V}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{N}\mathbf{s}_{it}(\widehat{\boldsymbol{\theta}}_{QML})\right)\left(\sum_{i=1}^{N}\mathbf{s}_{it}(\widehat{\boldsymbol{\theta}}_{QML})\right)^{\top},$$

where $\mathbf{s}_{it}(\boldsymbol{\theta}) = \partial/\partial\boldsymbol{\theta}\log h(Y_{i,t} \mid \mu_{i,t}, \sigma_{i,t}, \xi_{i,t})$.

2.3. *A grouped panel GEV regression model.* To introduce flexibility in panel models, a fast-growing literature explores the existence of latent group structures: see Su, Shi and Phillips (2016) and Wang and Su (2021) for examples in linear and non-linear panel regression models; Gu and Volgushev (2019) for an example in panel quantile models; Oh and Patton (2020) for an example with dynamic copula models. We assume that each of the N individuals is a member of one of $G \geq 1$ groups. We let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_N)$ be the N-dimensional vector whose ith entry $\tau_i \in \{1, \ldots, G\}$ denotes the group membership of the ith individual. The grouped panel GEV model is defined as

$$Y_{i,t} \mid \mathbf{X}_{i,t} \sim H\left(y \mid \mu_{i,t}(\tau_i), \sigma_{i,t}(\tau_i), \xi_{i,t}(\tau_i)\right), \tag{8}$$

with

$$\begin{aligned} \mu_{i,t}(\tau_i) &= e_\mu\left(\boldsymbol{\kappa}_{(\tau_i)}^{\top}\mathbf{X}_{i,t}\right) \\ \sigma_{i,t}(\tau_i) &= e_\sigma\left(\boldsymbol{\gamma}_{(\tau_i)}^{\top}\mathbf{X}_{i,t}\right) \\ \xi_{i,t}(\tau_i) &= e_\xi\left(\boldsymbol{\delta}_{(\tau_i)}^{\top}\mathbf{X}_{i,t}\right) \end{aligned} \tag{9}$$

where $\boldsymbol{\theta}_{(g)} = \left(\boldsymbol{\kappa}_{(g)}, \boldsymbol{\gamma}_{(g)}, \boldsymbol{\delta}_{(g)}\right) \in \boldsymbol{\Theta} \subset \mathbb{R}^P$ is the vector of regression parameters for the gth group, $g \in \{1, \ldots, G\}$. We denote the full parameter vector by $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_{(1)}, \ldots, \boldsymbol{\theta}_{(G)}\right)$. In the existing literature, the individuals are typically grouped according to observed characteristics, such as physical distance in spatial applications and business lines in economic applications (Asadi, Davison and Engelke, 2015; Mhalla, Hambuckers and Lambert, 2020). Given the group assignments, the parameters of the grouped panel GEV model can be estimated group-wise using the QML estimator in (7). However, *a priori* assignments may not provide the best fit to the data; see the applications in Sections 4 and 6.

We therefore propose a data-driven method that jointly estimates the group assignments $\boldsymbol{\tau}$ and the vector of model parameters $\boldsymbol{\theta}$ based on an EM algorithm. This algorithm iterates between estimating the regression parameters given the group assignments and estimating the group assignments given the regression parameters. Disentangling the estimation of the regression parameters and the group assignments drastically simplifies the estimation problem.

Let $(Y_{i,t}, \mathbf{X}_{i,t})$ be the observations from the grouped panel GEV model, where $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Our algorithm has the following steps.

0. **Initialization**. For a fixed value of G, select an initial group assignment $\widehat{\boldsymbol{\tau}}^{(0)}$ at random.
1. **Iteration**. We iterate between the following two steps until convergence. At the jth iteration, we have

1.1 **Maximization**. Conditioning on the group structure $\widehat{\boldsymbol{\tau}}^{(j-1)}$ identified at the previous iteration, we apply the QML estimator over each of the G groups. The EM estimator is thus defined as $\widehat{\boldsymbol{\theta}}^{(j)} = \left(\widehat{\boldsymbol{\theta}}^{(j)}_{(1)}(\widehat{\boldsymbol{\tau}}^{(j-1)}), \ldots, \widehat{\boldsymbol{\theta}}^{(j)}_{(G)}(\widehat{\boldsymbol{\tau}}^{(j-1)})\right)$ with

$$\widehat{\boldsymbol{\theta}}^{(j)}_{(g)}(\boldsymbol{\tau}) = \arg\max_{\boldsymbol{\theta}} \sum_{i:g=\tau_i} \sum_{t=1}^{T} \log h\left(Y_{i,t} \mid \mu_{i,t}(\tau_i), \sigma_{i,t}(\tau_i), \xi_{i,t}(\tau_i)\right), \quad g = 1, \ldots, G.$$

1.2 **Expectation**. Given the estimated parameters $\widehat{\boldsymbol{\theta}}^{(j)}$, we can find the best group assignments $\widehat{\boldsymbol{\tau}}^{(j)}$ maximizing the individual contribution to the likelihood, i.e.

$$\widehat{\tau}_i^{(j)} = \arg\max_{g \in \{1,\ldots,G\}} \sum_{t=1}^{T} \log h(Y_{i,t} \mid \widehat{\mu}_{i,t}(g), \widehat{\sigma}_{i,t}(g), \widehat{\xi}_{i,t}(g)), \quad i = 1, \ldots, N,$$

where $\widehat{\mu}_{i,t}(g) = e_\mu\left(\widehat{\boldsymbol{\kappa}}_{(g)}^\top \mathbf{X}_{i,t}\right)$, $\widehat{\sigma}_{i,t}(g) = e_\sigma\left(\widehat{\boldsymbol{\gamma}}_{(g)}^\top \mathbf{X}_{i,t}\right)$, $\widehat{\xi}_{i,t}(g) = e_\xi\left(\widehat{\boldsymbol{\delta}}_{(g)}^\top \mathbf{X}_{i,t}\right)$. This step only requires the comparison of G constants for each individual.

The output of the EM algorithm is the estimated group assignments $\widehat{\boldsymbol{\tau}}_T$ and the estimated group parameters $\widehat{\boldsymbol{\theta}}_T$, where we sometimes drop the subscript T for the number of samples in time. We call Step 1.2 an expectation step for familiarity even though we do not compute an expectation. We rather use the alternative approach of assigning each observation to the likelier group (McLachlan and Krishnan, 2008) as per Oh and Patton (2020).

Theorem 1 in Appendix A provides a consistency result for this EM algorithm proving that for a fixed number of individuals N and a growing number of samples per individual $T \to \infty$, the group assignments and estimated group parameters converge to their true counterparts, that is, we have the convergences in probability

$$\widehat{\boldsymbol{\tau}}_T \xrightarrow{p} \boldsymbol{\tau}_0, \qquad \widehat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0, \qquad T \to \infty.$$

Our assumption on the sequence of samples $(Y_{i,t}, \mathbf{X}_{i,t})$, $i = 1, \ldots, N$, $t = 1, \ldots, T$, for growing T is that they form a stationary and erdogic sequence. We note that this is fairly general since despite the stationarity of the joint sequence, the conditional distributions $Y_{i,t} \mid \mathbf{X}_{i,t}$ vary across individuals and over time with the covariate vector $\mathbf{X}_{i,t}$ according to the panel model in (8). We also require a set of regularity conditions on the log-likelihood of the GEV distribution as a function of the model parameters $\boldsymbol{\theta}$. We note here that such regularity conditions are not always easy to verify for the GEV distribution because of the changing support, and, in general, it will depend on the link functions used in the grouped panel GEV model whether these conditions are met. The asymptotic theory of maximum likelihood estimation for the GEV under the most general conditions remains an active area of research even in the i.i.d. case, see Dombry (2015), Bücher and Segers (2017) and Dombry and Ferreira (2019) for recent results.

Inference on the parameters of the grouped panel is performed as in the single group setting in Section 2.2. Standard errors for the parameters of each group are computed assuming independence among observations in different groups. An important topic in the clustering literature is the selection of the number of groups (Su, Shi and Phillips, 2016; Oh and Patton, 2020), and we face an analogous challenge. As the number of groups is unknown, we repeat our procedure for different values of G and rely on the BIC to select the optimal number of groups G^*, i.e.

$$G^* = \arg\min_{G \in \mathbb{N}} \text{BIC}(G)$$

where

$$\text{BIC}(G) = -2\sum_{g=1}^{G}\sum_{i:\widehat{\tau}_i=g}\sum_{t=1}^{T}\log h(Y_{i,t} \mid \widehat{\mu}_{i,t}(\widehat{\tau}_i), \widehat{\sigma}_{i,t}(\widehat{\tau}_i), \widehat{\xi}_{i,t}(\widehat{\tau}_i)) + \log(NT) \times PG,$$

and P is the dimension of the parameter vector $\boldsymbol{\Theta}$ in (8) in each group.

3. Simulation study. We design a comprehensive simulation study in order to assess the finite sample properties of our EM algorithm for the grouped panel GEV model in typical settings. We generate sample maxima $\mathbf{Y}_t = (Y_{1,t}, \ldots, Y_{N,t})$ for $t = 1, \ldots, T$ according to the following model

$$\begin{aligned}\mathbf{U}_t &= (U_{1,t}, \ldots, U_{N,t}) \sim \mathbf{C}_\alpha,\\ \mathbf{Y}_t &= \left(H_{1,t}^{-1}(U_{1,t}), \ldots, H_{N,t}^{-1}(U_{N,t})\right),\end{aligned}$$

where $\mathbf{C}_\alpha$ is a copula characterizing the cross-sectional dependence structure with dependence parameter α, and $H_{i,t}(y) = H\left(y \mid \mu_{i,t}(\tau_i), \sigma_{i,t}(\tau_i), \xi_{i,t}(\tau_i)\right)$ is the marginal GEV distribution. We assume constant shape parameters through time and across individuals in the same group, i.e., $\xi_{i,t}(g) = \delta_{0,(g)}$, and time-varying location and scale parameters as functions of the vector of covariates $\mathbf{X}_{i,t} = \left(X_{i,t}^{(1)}, X_i^{(2)}\right)$. In particular,

$$\begin{aligned}\mu_{i,t}(\tau_i) &= \kappa_{0,(\tau_i)} + \kappa_{1,(\tau_i)}X_{i,t}^{(1)} + \kappa_{2,(\tau_i)}X_i^{(2)},\\ \sigma_{i,t}(\tau_i) &= \exp\left(\gamma_{0,(\tau_i)} + \gamma_{1,(\tau_i)}X_{i,t}^{(1)} + \gamma_{2,(\tau_i)}X_i^{(2)}\right),\end{aligned}$$

with $\tau_i \in \{1, \ldots, G\}$. We let $X_{i,t}^{(1)}$ evolve according to the factor model

$$X_{i,t}^{(1)} = \omega + \lambda t + \beta f_t + \epsilon_{i,t},$$

where $f_t \sim N(0, \nu_f)$ and $\epsilon_{i,t} \sim N(0, \nu_i)$. These dynamics are designed to characterize a time-varying individual-specific covariate as in the financial risk application of Section 4. We let $X_i^{(2)}$ be uniformly distributed within the interval $(\underline{u}, \overline{u})$ to characterize a time-invariant individual-specific covariate like the spatial characteristics used in the climatological and hydrological applications of Sections 5 and 6, respectively.

For 100 repetitions, we generate samples $\{Y_{i,t}, \mathbf{X}_{i,t}\}_{i=1,t=1}^{N,T}$ with $N = 24$ and $T \in \{10, 20, 50\}$. The performance of the EM algorithm should improve as T increases. We consider three copula functions to assess how the algorithm behaves under different dependence structures: the independence copula ($\mathbf{C}^{Ind}$) imposing zero cross-sectional dependence; the Gaussian copula ($\mathbf{C}_{0.5}^{Gauss}$) with constant correlation coefficient $\alpha = 0.5$ across the individuals, implying a moderate cross-sectional dependence (Kendall's $\tau = 0.33$) but zero tail dependence; the Gumbel copula ($\mathbf{C}_2^{Gum}$) with parameter $\alpha = 2$ implying similar moderate cross-section dependence (Kendall's $\tau = 0.5$), but positive tail dependence. We assume that the true number of groups is $G_0 = 4$ and fix the model parameters as in Table 1. We assign an equal number of individuals $N/G_0 = 6$ to each group. We estimate the grouped panel GEV model considering $G \in \{1, \ldots, 6\}$ and select the optimal number of groups using the BIC. Table 2 reports the performance of this approach. The ability of the BIC to recover the correct number of groups is already very good with only 20 time points and is perfect when $T = 50$. The quality of the assignment when $G = 4$, as measured by the Rand index (Rand, 1971), a measure of similarity between two partitions, is good for small T and it is almost perfect for $T = 50$. An interesting aspect emerging from Table 2 is that the ability of the BIC to recover the correct number of groups grows with increasing cross-sectional dependence.

Stronger dependence helps the group identification because it reduces the variance among the individuals in the same group. Consequently, as this variance is lower, the variance of the estimator must increase, as expected by QML estimation under dependence.

As applications are often concerned with the estimation of high quantiles of the GEV distribution, we also assess the performance of the EM algorithm from this perspective. First, we assess the quality of the quantile estimates for the different group structures obtained with the algorithm. We compute the mean relative absolute error (MRAE) between the true 99th quantiles $Q_{it}^{0.99}$ (see equation (5)) and those estimated with $G \in \{1, \ldots, 6\}$. Figure 1 shows that the quality of the estimates is poor when the panel size is very small ($T = 10$). As T increases, $G = G_0 = 4$ provides the best quantile estimates. Interestingly, even though stronger dependence improves the group assignments (Table 2), the quality of the quantile estimates deteriorates as the dependence increases. This is coherent with the higher uncertainty of the QML estimator under dependence. Second, we assess the quality of the quantile estimates under the BIC-selected number of groups. We compute the MRAE between $Q_{it}^{0.99}$ and the quantiles estimated with the selected model. Figure 1 shows that the quantiles estimated with the BIC-selected model attain almost optimal performance in terms of MRAE compared to the pre-selected number of groups for T as small as 20.

TABLE 1

True value of parameters used in simulations to study the finite sample properties of the EM algorithm.

	Group parameters				Covariate parameters	
	$g=1$	$g=2$	$g=3$	$g=4$		
$\kappa_{0,(g)}$	3.10	3.40	3.20	3.10	ω	-0.8
$\kappa_{1,(g)}$	2.40	1.40	1.10	1.70	λ	0.4/T
$\kappa_{2,(g)}$	2.00	1.00	0.50	1.50	β	0.8
$\gamma_{0,(g)}$	-0.05	-0.15	-0.20	-0.10	ν_f	0.5
$\gamma_{1,(g)}$	0.10	0.06	0.04	0.08	ν_I	0.5
$\gamma_{2,(g)}$	0.17	0.07	0.02	0.12	$\underline{u}$	2
$\delta_{0,(g)}$	0.30	0.27	0.24	0.20	$\overline{u}$	6

TABLE 2

Performance of the BIC in selecting the correct number of groups and average Rand index computed between the true and estimated group structures when $G_0 = 4$ over 100 replications: independence ($\mathbf{C}^{Ind}$), Gaussian ($\mathbf{C}_{0.5}^{Gauss}$) and Gumbel ($\mathbf{C}_{2}^{Gum}$) dependence.

Size	$\mathbf{C}^{Ind}$		$\mathbf{C}_{0.5}^{Gauss}$		$\mathbf{C}_{2}^{Gum}$	
	BIC	Rand	BIC	Rand	BIC	Rand
$T = 10$	24%	88%	34%	91%	42%	93%
$T = 20$	78%	94%	92%	97%	88%	98%
$T = 50$	100%	99%	99%	99%	100%	99%

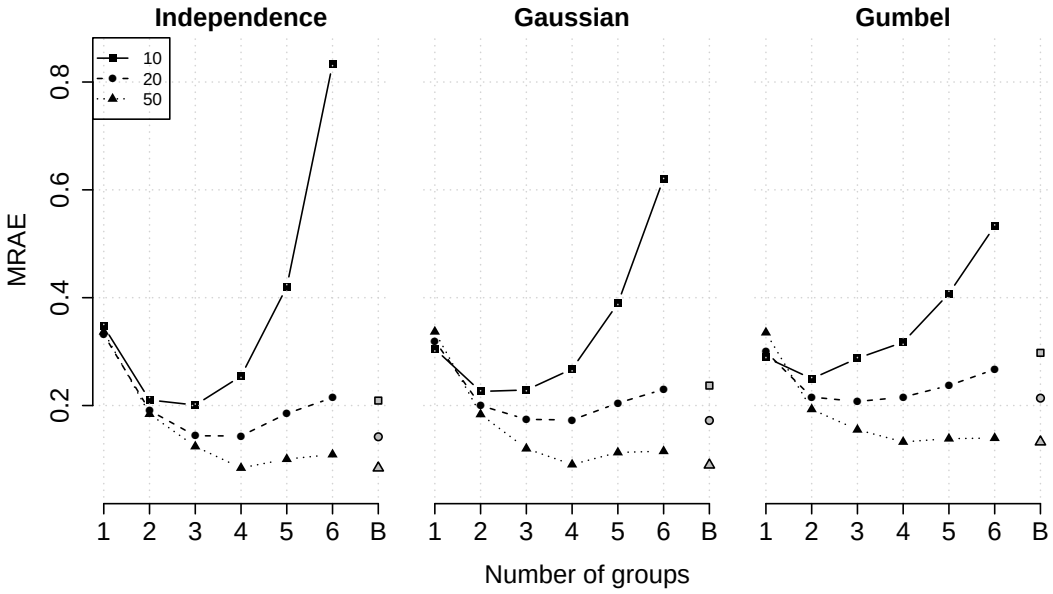

Fig 1: Median MRAE when estimating $Q_{it}^{0.99}$ using the model with $G \in \{1, \ldots, 6\}$ and the BIC selected model (B) over 100 replications: independence ($\mathbf{C}^{Ind}$), Gaussian ($\mathbf{C}_{0.5}^{Gauss}$) and Gumbel ($\mathbf{C}_{2}^{Gum}$) dependence.

4. Financial risk management. Extreme value theory offers suitable instruments for stress testing of extreme losses (Committee on the Global Financial System, 2001). The GEV distribution in (2) is commonly employed in the financial literature to model maximum losses over a fixed horizon, both in unconditional static (Bali, 2003; McNeil, Frey and Embrechts, 2015) and conditional dynamic (Zhao, Zhang and Chen, 2018) settings. In a stress testing setup, a GEV regression model ties maximum losses to a time-varying market risk factor to study how changes to the latter affect large losses. An example for such a risk factor is the semi-variance, the average of the squared deviations of values that are less than the mean. It is a measure for downside risk, unlike the variance which provides a measure of volatility. The semi-variance is a widely used measure of downside risk in finance (Barndorff-Nielsen, Kinnebrouk and Shephard, 2010), which is strongly related to macroeconomic uncertainty (Jurado, Ludvigson and Ng, 2015; Segal, Shaliastovich and Yaron, 2015).

We consider an institutional investor diversifying across 48 U.S. industry portfolios and interested in a stress test analysis of the portfolios' annual maximum daily loss using the portfolios' annual realized semi-variance as a risk factor. Table 3 shows the 48 U.S. industry portfolios for which we collect daily returns from 1950 to 2018[1]. Let $Z_{i,t,j}$ be the daily return of the ith industry portfolio on the jth day of the tth year, with $i = 1, \ldots, 48$ and $t = 1, \ldots, 69$. We define the annual maximum loss $Y_{i,t} = \max_{j=1,\ldots,s}(-Z_{i,t,j})$ and the annual realized semi-variance $RS_{i,t} = \sum_{j=1}^{s} Z_{i,t,j}^2 \mathbb{I}\{Z_{i,t,j} < 0\}$, where $\mathbb{I}\{\cdot\}$ denotes the indicator function. Accurate stress testing requires a model that properly characterizes the variation in the extreme quantiles of such industry portfolios to changes in the annual realized semi-variance. We model the panel of maximum losses with the following grouped panel GEV

[1]Details on the composition of the portfolios and data are available from the Kenneth French Data Library at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

model

$$\begin{aligned}\log\left(\mu_{i,t}\left(\tau_i\right)\right) &= \kappa_{0,(\tau_i)} + \kappa_{1,(\tau_i)}\log\left(RS_{i,t}\right)\\ \log\left(\sigma_{i,t}\left(\tau_i\right)\right) &= \gamma_{0,(\tau_i)}\\ \log\left(\xi_{i,t}\left(\tau_i\right)\right) &= \delta_{0,(\tau_i)} + \delta_{1,(\tau_i)}\log\left(RS_{i,t}\right)\end{aligned}$$

with $\tau_i \in \{1,\ldots,G\}$. The realized semi-variance in the location parameter accounts for the heteroskedasticity characterizing financial data. As periods of high volatility should lead to larger extremes, we expect the κ_1 parameters to be positive. Similarly, we model the shape parameter $\xi_{i,t}$ as a function of the realized semi-variance to account for the changing nature of tail risk. Previous analyses have shown that the shape parameter changes over time and it is positively associated to market volatility and economic uncertainty (Zhao, Zhang and Chen, 2018; Massacci, 2017), therefore we expect the δ_1 parameters to be positive. We consider the logarithm of $RS_{i,t}$ as it is convenient from a modeling perspective (Bee, Dupuis and Trapin, 2019).

We first estimate the model using all the available observations simultaneously, i.e., setting $G = 1$. Table 4 shows that the QML estimate for $\kappa_{1,(g)}$ is positive and strongly statistically significant. The parameter $\delta_{1,(g)}$ is positive but not statistically significant. This would lead us to conclude that the variation in the realized semi-variance is not informative of tail risk in the stock market, as measured by the shape parameters $\xi_{i,t}$. However, a more careful analysis shows that the omitted heterogeneity is leading to misguided inference. We explore possible group structures and estimate the panel GEV model with the EM algorithm using $G \in \{2,3,4,5,6\}$. The BIC-based optimal number of groups is $G^* = 4$. Table 3 shows the optimal group assignments $\widehat{\boldsymbol{\tau}}^*$ and Table 4 shows the estimated parameters. The results suggest that there is strong group heterogeneity in the panel, particularly on the constant terms of the regression, i.e., the parameters κ_0, γ_0, δ_0. Estimates for $\kappa_{1,(g)}$ are still positive and strongly statistically significant for each $g \in \{1,\ldots,G\}$. The size of the δ_1 parameters is now larger and statistically significant at the 5% level for two of the four groups. The benefit of the grouped panel GEV model from a stress testing perspective is clear when comparing the 90th $(Q_{it}^{0.90})$ and 95th $(Q_{it}^{0.95})$ quantiles computed by (5) for $G = 1$ and for the optimal group structure $\widehat{\boldsymbol{\tau}}^*$. Table 5 reports summary statistics for the quantile exceedances on each industry portfolio, i.e., $V_i^p = \frac{1}{T}\sum_{t=1}^{T}\mathbb{I}\{Y_{it} > Q_{it}^p\}$, with $p \in \{0.90, 0.95\}$. While the median number of exceedances across portfolios is similar when using one group or the optimal assignments $\widehat{\boldsymbol{\tau}}^*$, the spread is considerably larger in the case of the former, both for the 90th and 95th quantiles, highlighting the greater accuracy of the latent group panel GEV model.

Standard practice in finance would group industry portfolios based on the Standard Industrial Classification (SIC) (Oh and Patton, 2020), and for comparison we carry out our previous calculations based on this group assignment $\boldsymbol{\tau}^{SIC}$. Table 3 shows the group assignments and Table 4 shows the estimated parameters. Estimates uncover the heterogeneity in the δ_1 parameters associated to the realized semi-variance, thus improving upon the $G = 1$ estimates, but the group assignments are quite different from those identified with the EM algorithm and the corresponding maximized log-likelihood value is much smaller. As for quantile exceedances on each industry portfolio, Table 5 shows that the medians are similar, but the spread is greater when using $\boldsymbol{\tau}^{SIC}$ rather than $\widehat{\boldsymbol{\tau}}^*$, particularly on the 90th quantile. This suggests that the SIC grouping only partially explains the heterogeneity in the panel GEV model and that a data-driven procedure to group the portfolios is beneficial.

TABLE 3

Industry portfolios and corresponding group assignments obtained by the EM algorithm ($\widehat{\boldsymbol{\tau}}^*$) and defined by the SIC ($\boldsymbol{\tau}^{SIC}$).

Name	*Description*	$\widehat{\boldsymbol{\tau}}^*$	$\boldsymbol{\tau}^{SIC}$	*Name*	*Description*	$\widehat{\boldsymbol{\tau}}^*$	$\boldsymbol{\tau}^{SIC}$
Agric	Agriculture	4	1	Ships	Shipbuilding, Railroad Equipment	4	2
Food	Food Products	2	1	Guns	Defense	1	2
Soda	Candy & Soda	4	1	Gold	Precious Metals	4	5
Beer	Beer & Liquor	1	1	Mines	Non-Metallic and Industrial Metal Mining	4	5
Smoke	Tobacco Products	2	1	Coal	Coal	4	2
Toys	Recreation	1	1	Oil	Petroleum and Natural Gas	4	2
Fun	Entertainment	1	5	Util	Utilities	2	2
Books	Printing and Publishing	3	1	Telcm	Communication	3	3
Hshld	Consumer Goods	3	1	PerSv	Personal Services	1	1
Clths	Apparel	3	1	BusSv	Business Services	3	3
Hlth	Healthcare	4	4	Comps	Computers	4	3
MedEq	Medical Equipment	3	4	Chips	Electronic Equipment	4	3
Drugs	Pharmaceutical Products	4	4	LabEq	Measuring and Control Equipment	4	3
Chems	Chemicals	1	2	Paper	Business Supplies	1	2
Rubbr	Rubber and Plastic Products	1	2	Boxes	Shipping Containers	1	2
Txtls	Textiles	3	1	Trans	Transportation	3	5
BldMt	Construction Materials	3	5	Whlsl	Wholesale	2	1
Cnstr	Construction	1	5	Rtail	Retail	3	1
Steel	Steel Works	4	5	Meals	Restaraunts, Hotels, Motels	3	5
FabPr	Fabricated Products	4	2	Banks	Banking	2	5
Mach	Machinery	3	2	Insur	Insurance	3	5
ElcEq	Electrical Equipment	1	2	RlEst	Real Estate	1	5
Autos	Automobiles and Trucks	1	2	Fin	Trading	3	5
Aero	Aircraft	1	2	Other	-	1	5

TABLE 4

Parameter estimates and log-likelihood values (LLH) of the models for the industry portfolios: single group $G = 1$, optimal group structure $\widehat{\boldsymbol{\tau}}^*$ from the EM algorithm, and SIC-based $\boldsymbol{\tau}^{SIC}$. Standard errors in parentheses.

	$G=1$	$\widehat{\boldsymbol{\tau}}^*$				$\boldsymbol{\tau}^{SIC}$				
		$g=1$	$g=2$	$g=3$	$g=4$	$g=1$	$g=2$	$g=3$	$g=4$	$g=5$
$\kappa_{0,(g)}$	-0.88	-1.23	-1.12	-0.98	-0.89	-0.91	-0.89	-0.74	-0.82	-0.91
	(0.07)	(0.05)	(0.16)	(0.04)	(0.05)	(0.09)	(0.07)	(0.39)	(0.25)	(0.06)
$\kappa_{1,(g)}$	0.43	0.49	0.49	0.45	0.43	0.43	0.43	0.40	0.42	0.43
	(0.02)	(0.01)	(0.04)	(0.01)	(0.01)	(0.02)	(0.02)	(0.09)	(0.06)	(0.01)
$\gamma_{0,(g)}$	-0.37	-0.47	-0.73	-0.57	-0.21	-0.50	-0.46	-0.30	-0.32	-0.42
	(0.08)	(0.07)	(0.19)	(0.09)	(0.08)	(0.08)	(0.11)	(0.19)	(0.15)	(0.08)
$\delta_{0,(g)}$	-3.63	-6.03	-5.83	-4.89	-5.61	-5.11	-3.54	-3.62	-6.53	-5.09
	(1.36)	(2.25)	(3.73)	(2.72)	(2.18)	(1.36)	(1.11)	(4.46)	(5.69)	(1.81)
$\delta_{1,(g)}$	0.33	0.81	0.84	0.68	0.65	0.71	0.41	0.30	0.87	0.60
	(0.22)	(0.42)	(0.52)	(0.55)	(0.33)	(0.27)	(0.23)	(0.78)	(0.91)	(0.29)
LLH	-4225.01	-4065.24				-4199.68				

5. Effect of climate change on extreme temperatures. There is a large literature on global mean temperature change and an increasing literature focusing on trends of temperature extremes, see respectively e.g. Hansen et al. (2010) and Papalexiou et al. (2018), and references therein. The global mean temperature has recently been increasing at an increasing rate, but local temperature extremes have not always changed at the same rate, or even in the same direction. Temperature extremes have well-documented detrimental health and social impacts (IPCC, 2008). Nighttime warming reduces crop yields (García et al., 2015; Sadok

TABLE 5

Summary statistics for the average number of exceedances of the 90th and 95th quantile for 48 U.S. industry portfolios: minimum (Min), 25th quantile (Q_1), median (Med), 75th quantile (Q_3), and maximum (Max). The expected number of exceedances is 0.10 and 0.05 for the 90th and 95th quantile, respectively.

	90th quantile					95th quantile				
	Min	Q_1	Med	Q_3	Max	Min	Q_1	Med	Q_3	Max
$G=1$	0.057	0.098	0.116	0.134	0.261	0.014	0.041	0.058	0.087	0.161
$\widehat{\boldsymbol{\tau}}^*$	0.072	0.101	0.116	0.130	0.203	0.018	0.043	0.058	0.072	0.125
$\boldsymbol{\tau}^{SIC}$	0.072	0.089	0.116	0.130	0.247	0.018	0.043	0.058	0.072	0.161

and Krishna Jagadish, 2020) and can have large economic consequences for crop-producing regions. Assessing the global mean to local extreme temperature relationship makes the effects of climate change more relatable to economically vulnerable constituents and can guide public policy at the local level. Extreme value theory and the latent group panel GEV model developed in Section 2.3 allow us to seek this relationship for the crop-producing region in the U.S. Midwest in Figure 4.

The GEV distribution in (2) is a widely used model for annual daily minimum temperatures (Zwiers and Kharin, 1998; Wang et al., 2016). Including the global land temperature anomaly as a covariate in the GEV location and scale parameters allows us to infer the sought-after global mean/local extreme relation at a given location, but short series yield large standard errors and a panel model could provide more precise estimates. We consider the negative annual daily minimum temperature $Y_{i,t}$ (in °C), $i = 1, \ldots, N$, $t = 1, \ldots, T$, at $N = 127$ stations in seven crop-producing states in the U.S. Midwest (Ohio, Indiana, Illinois, Iowa, Missouri, Nebraska, and Kansas) during $T = 99$ years; see Figure 4 for the locations of the weather stations. Data are available for the 1912 to 2010 period in the `USHCNTemp` dataset of the R package `SpatialExtremes` (Ribatet, 2019). Elevation and latitude should explain some of the variation in extremes across the U.S. Midwest, so we let the parameters of the panel GEV model to vary as a function of $\mathbf{X}_{i,t} = (\text{elev}_i, \text{lat}_i, \text{anom}_t)$ with

$$
\begin{aligned}
\mu_{i,t}(\tau_i) &= \kappa_{0,(\tau_i)} + \kappa_{1,(\tau_i)}\text{elev}_i + \kappa_{2,(\tau_i)}\text{lat}_i + \kappa_{3,(\tau_i)}\text{anom}_t \\
\log(\sigma_{i,t}(\tau_i)) &= \gamma_{0,(\tau_i)} + \gamma_{1,(\tau_i)}\text{elev}_i + \gamma_{2,(\tau_i)}\text{lat}_i + \gamma_{3,(\tau_i)}\text{anom}_t \\
\xi_{i,t}(\tau_i) &= \delta_{0,(\tau_i)}
\end{aligned} \tag{10}
$$

for $\tau_i \in \{1, \ldots, G\}$, where elev_i and lat_i denote, respectively, the elevation (in 10^3 feet) and (normalized) latitude at station i, and anom_t denotes the annual global land anomaly (in °C) in year t. Elevation and latitude are provided in the `USHCNTemp` dataset. Annual global land anomalies are available at `https://www.ncdc.noaa.gov/monitoring-references/faq/anomalies.php#anomalies`. We expect $\kappa_{0,(g)}$ to capture some of the group-specific idiosyncrasies, and $\kappa_{1,(g)}$ and $\kappa_{2,(g)}$ to be positive. The tail index $\delta_{0,(g)}$ should be negative as annual daily minimum temperature has a lower bound. The effect of the global anomaly on the location and scale parameters of the GEV, as measured by $\kappa_{3,(g)}$ and $\gamma_{3,(g)}$, respectively, is of interest.

Figure 2 shows BIC values of optimal panel fits for different group sizes. Table 6 shows the estimated parameters for panel model (10) with $G = 1$ group as well as the estimated parameters with $G = G^* = 4$, the optimal number of groups based on the BIC criterion. Figure 3 shows the (local) estimates of $\kappa_{3,(g)}$ when a GEV model with covariates as in (10) is fitted to data at each of the 127 locations, i.e., $G = 127$ groups, as well as the estimates based on the GEV panel model with $G^* = 4$. The panel estimates with $G^* = 4$ have better precision by pooling information across all stations within a group, showing 29, 39, 37 and 28% mean reduction in estimated standard errors compared to local estimates for groups $g = 1$ to $g = 4$,

respectively. The panel fit allows us to infer that a 1°C increase in the annual global land anomaly results in mean increases in the annual daily minimum temperature between 1.7 and 2.8°C in the area under study. Figure 4 shows the panel groupings. Largest mean increases in annual daily minimum temperature are predominantly in the western-most states (Nebraska, Kansas, Iowa and Missouri). The panel fit with $G = 1$ in Table 6 yields $\hat{\kappa}_{3,(g)} = -2.2$°C for the entire area under study, but the qq-plots at each station (not shown) for this $G = 1$ model are quite poor and inference is questionable. Station-wise qq-plots (not shown) based on the $G^* = 4$ panel model are very good, so the panel model allows us to confidently infer that a 1°C increase in the annual global anomaly results in mean increases to annual daily minimum temperature in some U.S. Midwest regions that are up to almost three-fold that amount. Continued exacerbated local effects of likely global increases would be particularly problematic for these important crop-producing U.S. Midwest states where temperature-driven crop yield variability is already well documented (Kukal and Irmak, 2018; Petersen, 2019). Changes to the annual global land anomaly do not have a significant effect on the variability of annual daily minimum temperatures as $\gamma_{3,(g)}$ estimates in Table 6 are small when compared to their associated standard errors.

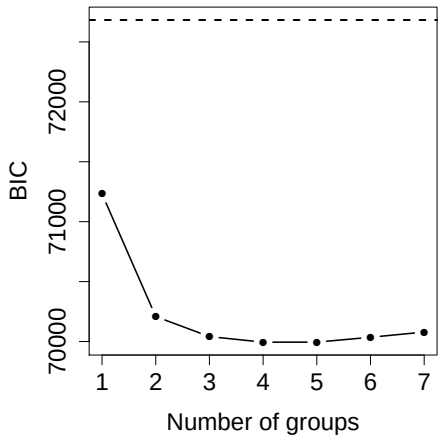

Fig 2: BIC values for the optimal panel fits as a function of different group sizes (solid line). The horizontal dashed line corresponds to the BIC of the local model, i.e. when $G = N = 127$.

6. Flood risk assessment. Floods are major natural hazards that threaten human lives and cause huge damages to the environment and the economy. Effective flood protection is therefore crucial and this requires an accurate assessment of the risk related to high river flows.

The GEV distribution H in (2) is a widely used model for yearly maxima of river discharges thanks to its mathematical justification and good properties in applications (e.g. Katz, Parlange and Naveau, 2002). At a gauging station with many consecutive years of observations, the parameters μ, σ and ξ of this distribution family can be fitted locally as in (3), that is, using only data from this particular station.

In many hydrological applications, analysis of flood risk is required at stations with little or no data, and in such cases estimates of the return level $RL^S(\mu, \sigma, \xi)$ for long return periods S based on the the local fit may exhibit huge variances. Regionalization is a common

TABLE 6

Parameter estimates of the models for negative annual daily minimum temperatures: single group $G = 1$ and optimal group structure $\widehat{\boldsymbol{\tau}}^*$ from the EM algorithm.

Parameter	$G = 1$	$\widehat{\boldsymbol{\tau}}^*$			
		$g = 1$	$g = 2$	$g = 3$	$g = 4$
$\kappa_{0,(g)}$	-23.7	-24.2	-22.9	-22.1	-22.6
	(0.5)	(0.8)	(0.5)	(0.5)	(0.6)
$\kappa_{1,(g)}$	3.7	3.3	2.9	4.0	5.1
	(0.5)	(0.5)	(0.5)	(0.8)	(0.5)
$\kappa_{2,(g)}$	17.7	18.3	18.4	17.8	10.7
	(0.4)	(0.8)	(0.5)	(0.5)	(0.8)
$\kappa_{3,(g)}$	-2.2	-2.0	-2.8	-2.2	-1.7
	(0.7)	(0.8)	(0.7)	(0.8)	(0.8)
$\gamma_{0,(g)}$	1.59	1.87	1.55	1.53	1.64
	(0.06)	(0.12)	(0.08)	(0.07)	(0.08)
$\gamma_{1,(g)}$	-0.09	0.13	-0.22	0.03	-0.01
	(0.06)	(0.08)	(0.09)	(0.15)	(0.08)
$\gamma_{2,(g)}$	-0.30	-0.9	-0.4	-0.4	-0.4
	(0.09)	(0.2)	(0.1)	(0.1)	(0.1)
$\gamma_{3,(g)}$	0.04	0.10	0.10	0.03	0.02
	(0.07)	(0.09)	(0.08)	(0.09)	(0.1)
$\delta_{0,(g)}$	-0.24	-0.25	-0.27	-0.23	-0.25
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)

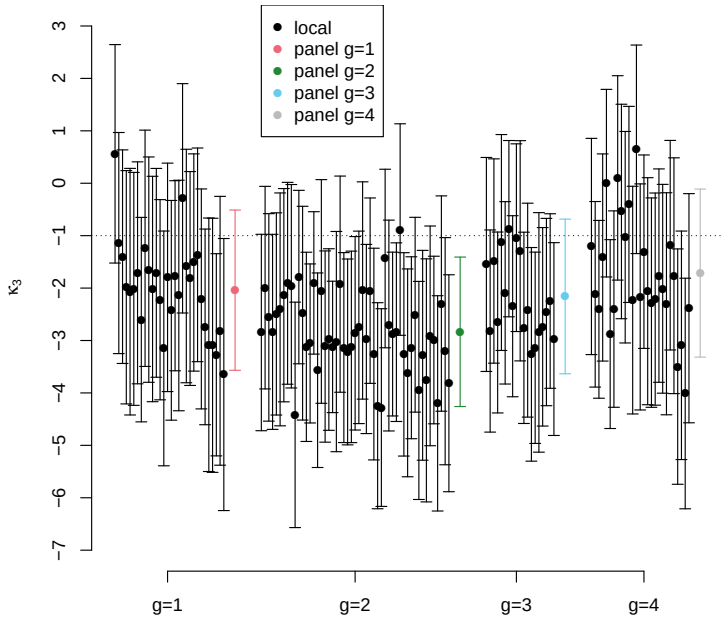

Fig 3: Effect of global anomaly on negative annual daily minimum temperature based on GEV model fitted to data at each of the 127 locations (black, plotted by group) and panel estimates (color). Grouping and colors shown in Figure 4. Horizontal dotted line indicates a 1:1 mean annual global anomaly to mean annual minimum temperature increase.

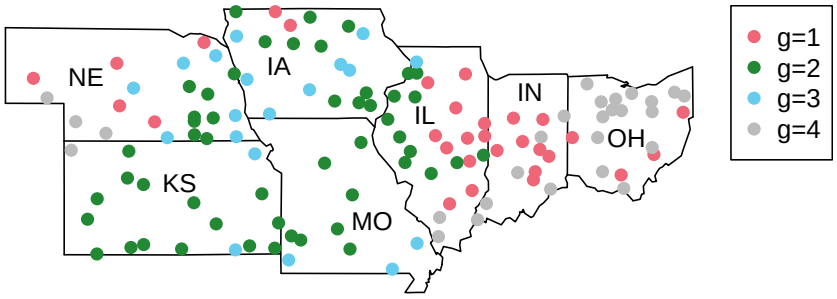

Fig 4: Seven U.S. Midwest states and 127 weather stations. The $G^* = 4$ different colors of the stations correspond to the optimal grouping for the panel GEV model fitted to the negative annual daily minimum temperature.

alternative way of estimating such high quantiles. It relies on identifying groups of stations that are similar to each other and on sharing their information on extreme flows to obtain more accurate estimates. There exists a vast literature that contains many different methods to construct such groupings and different ways of sharing the information (e.g., Burn, 1990; Merz and Blöschl, 2005; Asadi, Engelke and Davison, 2018). For a fixed grouping, a widely used model for the ith station is

$$\begin{aligned} \log(\mu_{i,t}(\tau_i)) &= \boldsymbol{\kappa}_{(\tau_i)}^\top \log(\mathbf{X}_{i,t}), \\ \log(\sigma_{i,t}(\tau_i)) &= \boldsymbol{\gamma}_{(\tau_i)}^\top \log(\mathbf{X}_{i,t}), \\ \xi_{i,t}(\tau_i) &= \xi_{(\tau_i)}, \end{aligned} \tag{11}$$

with $\tau_i \in \{1, \dots, G\}$. We consider yearly maxima of river flow data $Y_{i,t}$, $i = 1, \dots, N$, $t = 1, \dots, T$, at $N = 31$ stations in the upper Danube catchment during $T = 50$ years; see Figure 5 for the river network and the locations of the gauging stations. This data set has been used in Asadi, Davison and Engelke (2015), Gnecco et al. (2021), Mhalla, Chavez-Demoulin and Dupuis (2020), Röttger, Engelke and Zwiernik (2021) and Engelke and Hitz (2020) for univariate and multivariate extreme value analyses. In addition to the river flow measurements, for each observation $Y_{i,t}$ we use a corresponding covariate vector $\mathbf{X}_{i,t}$ that contains the latitude of the station, and the size, the mean altitude and the mean slope of the corresponding sub-catchment. Since the data are yearly maxima, we can interpret this as a grouped panel GEV regression as in (8), and we see that the covariate vectors are time-invariant, that is, $\mathbf{X}_{i,t} \equiv \mathbf{X}_i$.

For risk assessment at each of the $N = 31$ stations there are several possibilities. One may locally fit at each station separately a GEV distribution. As discussed above, this becomes infeasible or highly sub-optimal if the data record is too short at some stations. To illustrate this, we choose six stations and randomly delete 80% of their $T = 50$ observations. A local fit at each station can be seen as a grouped panel with $G = N = 31$ groups, that is, every station is in its own group and no information is shared. To borrow information across the 31 stations, one can use a covariate model as in (11) for all stations simultaneously. This corresponds to a panel model with $G = 1$ group only. Since the effect of the covariates might however not

be the same for all stations, a regionalization approach with a good group assigment of the stations should provide a superior fit. Asadi, Davison and Engelke (2015) choose such a grouping with $G = 4$ groups in an *ad hoc* way and fit model (11) to the data.

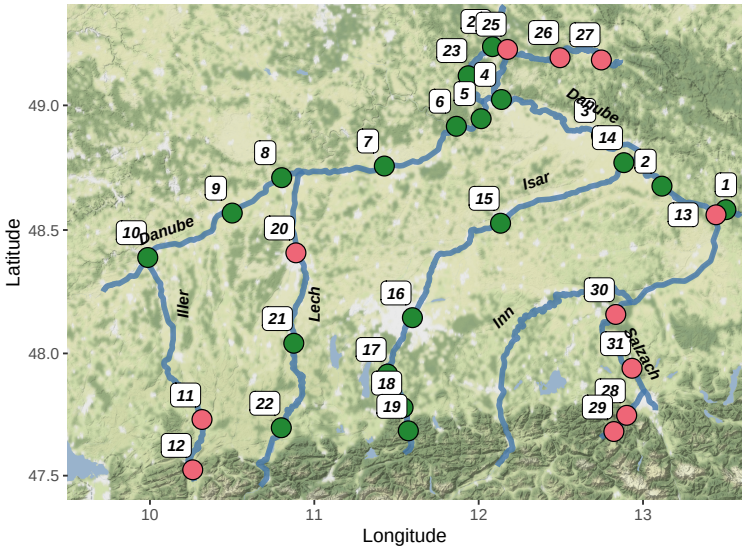

Fig 5: Upper Danube river basin and 31 gauging stations. The $G^* = 2$ different colors of the stations correspond to the optimal grouping for the GEV panel.

Our grouped panel approach allows us to find the optimal group assignments as described in Section 2.3. The number of groups for this data set is $G^* = 2$ and is chosen as the panel that minimizes the BIC; see Figure 7 for the BIC for different numbers of groups. The optimal number of groups chosen by our algorithm is smaller than in Asadi, Davison and Engelke (2015) and therefore allows for more efficient pooling of information. Note that even if we consider a fixed group size $G = 4$, the group assignments of our panel with four groups differ from the grouping in their approach. While they use an *ad hoc* grouping based on domain knowledge, we optimize our assignment in a purely data-driven way according to the second step in the EM algorithm in Section 2.3.

Figure 6 shows the QQ-plots of the different fitted models for four exemplary stations, where two of them have missing data. It can be seen that the local fit (figures on the right) performs well for stations where enough data are available, but naturally it is not able to capture well the tail if data at a station are scarce. For the other boundary case of a global fit with only one group, $G = 1$, the left-hand side of Figure 6 shows that such a model is not flexible enough to model the extremes at all locations well. Our optimized panel with data-driven group assignments is shown in the center column of Figure 6. We see that it combines the advantages of the local and the global fits. It is flexible enough to model the tail at all stations sufficiently well. Moreover, the fact of pooling the information from all stations in a group helps to obtain a good fit even for stations with a lot of missing data. The BIC values in Figure 7 provide further evidence that our optimized grouped panel GEV regression performs better than the local and global models, and the model of Asadi, Davison and Engelke (2015).

Figure 5 shows the final group assignments of our optimal panel model in two different colors. We recognize a clear spatial pattern in the sense that stations on the same river tend

to fall into the same group. This is astonishing since our method does not enforce this in any way. This shows the big advantage of our methodology, namely, that it does not require domain knowledge to produce sensible panel groupings.

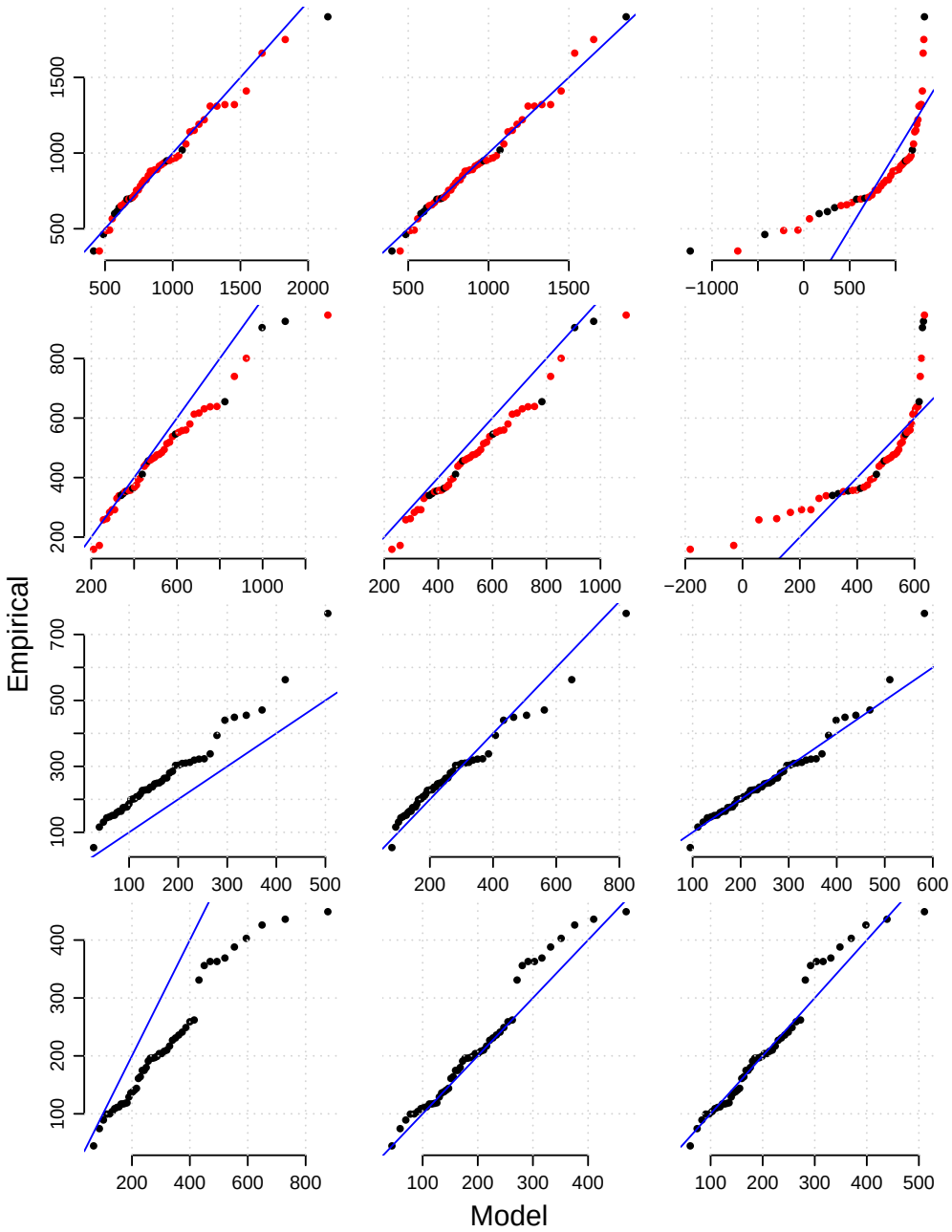

Fig 6: The rows correspond to QQ-plots of four stations: fit with $G = 1$ group (left), fit with $G^* = 2$ groups (center) and local fit (right). Red points are missing data that are not used for fitting in any of the models.

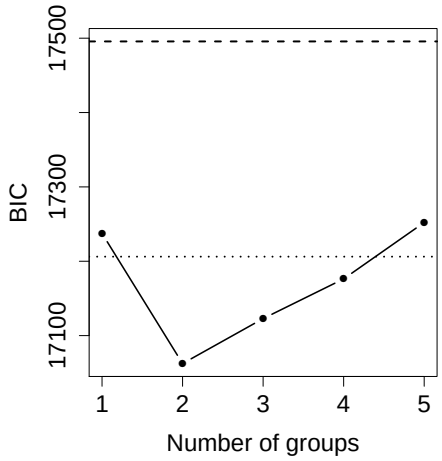

Fig 7: BIC values for the optimal panel fit as a function of different number of groups (solid line). The horizontal lines correspond to the BIC of the local model (dashed line) and the model in Asadi, Davison and Engelke (2015) (dotted line), respectively.

7. Discussion. Sections 4 to 6 present three applications demonstrating the practical usefulness of our model in very diverse settings, but the benefits of properly grouping individuals to estimate extreme events extend to other situations. We mention a few of the many applications: applying our grouped panel GEV regression model to i) annual maximum sub-hourly precipitation would reduce the estimation error of design storms at more recent gauging sites where only short series of precipitation data are available (Alila, 1999; Buishand, 1991); ii) annual maximum three-day snow fall depth data would reduce the estimation error of the extreme return levels required in avalanche hazard mapping (Bocchiola et al., 2008); iii) annual fastest-mile wind speed data would reduce the estimation error of at-site minimum design wind loads (Cheng, 1998).

We present our approach and applications based on the GEV regression model where extremes are defined as the maximum (or minimum) observation from blocks of equal size. A different modeling framework for extremes is the peaks-over-threshold method that uses the generalized Pareto (GP) distribution (Pickands, 1975). In the Appendix we outline how such a grouped panel GP regression model can be implemented.

The parameters in our grouped panel GEV or GP model are linear functions of the covariates. If more flexibility is needed, it is straightforward to generalize our approach to other forms of regression functions, such as additive models (Chavez-Demoulin and Davison, 2005), neural networks (Cannon, 2010; Velthoen et al., 2021) or random forests (Gnecco, Terefe and Engelke, 2022). Variable selection through lasso-type techniques is another possible extension (de Carvalho et al., 2021).

Future research may also extend our methodological framework to modeling of panels of multivariate extremes where the dependence structure of the maxima depends on a vector of covariates as in de Carvalho and Davison (2014). Latent groups with different dependence regimes could then be defined and our EM algorithm could be extended to estimate the regression parameters and group allocations in the dependence structure following the lines of Oh and Patton (2020).

REFERENCES

ALILA, Y. (1999). A hierarchical approach for the regionalization of precipitation annual maxima in Canada. *Journal of Geophysical Research* **104** 31,645–31,655.

ASADI, P., DAVISON, A. C. and ENGELKE, S. (2015). Extremes on river networks. *The Annals of Applied Statistics* **9** 2023–2050.

ASADI, P., ENGELKE, S. and DAVISON, A. C. (2018). Optimal regionalization of extreme value distributions for flood estimation. *Journal of Hydrology* **556** 182–193.

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

BARNDORFF-NIELSEN, O. E., KINNEBROUK, S. and SHEPHARD, N. (2010). *Measuring downside risk: realised semivariance*, In *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle* edited by T. Bollerslev, J. Russell and M. Watson) ed. 117-136. Oxford University Press.

BEE, M., DUPUIS, D. J. and TRAPIN, L. (2019). Realized Peaks over Threshold: a Time-Varying Extreme Value Approach with High-Frequency-Based Measures. *Journal of Financial Econometrics* **17** 254–283.

BOCCHIOLA, D., BIANCHI JANETTI, E., GORNI, E., MARTY, C. and SOVILLA, B. (2008). Regional evaluation of three-day snow depth for avalanche hazard mapping in Switzerland. *Natural Hazards and Earth System Sciences* **8** 685–705.

BÜCHER, A. and SEGERS, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes* **20** 839–872.

BÜCHER, A. and SEGERS, J. (2018). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli* **24** 1427–1462. MR3706798

BUISHAND, T. A. (1991). Extreme rainfall estimation by combining data from several sites. *Hydrological Sciences Journal* **36** 345–365.

BURN, D. H. (1990). An appraisal of the "region of influence" approach to flood frequency analysis. *Hydrological Sciences Journal* **35** 149-165.

CAMERON, C. A. and TRIVEDI, P. K. (2015). Count panel data. In *The Oxford Handbook of Panel Data* (B. H. Baltagi, ed.) 8 Oxford University Press, Oxford.

CANNON, A. J. (2010). A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology. *Hydrological Processes* **24** 673-685.

CARREAU, J., NAVEAU, P. and NEPPEL, L. (2017). Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. *Water Resources Research* **53** 4407–4426. https://doi.org/10.1002/2017WR020758

CHANDLER, R. E. and BATE, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika* **94** 167–183.

CHAVEZ-DEMOULIN, V. and DAVISON, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C* **54** 207–222.

CHENG, E. D. H. (1998). Macroscopic extreme wind regionalization. *Journal of Wind Engineering and Industrial Aerodynamics* **77-78** 13–21.

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

DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statistical Science* **27** 161–186.

DE CARVALHO, M. and DAVISON, A. C. (2014). Spectral density ratio models for multivariate extremes. *Journal of the American Statistical Association* **109** 764–776.

DE CARVALHO, M., PEREIRA, S., PEREIRA, P. and DE ZEA BERMUDEZ, P. (2021). An Extreme Value Bayesian Lasso for the Conditional Left and Right Tails. *Journal of Agricultural, Biological and Environmental Statistics* 1–18.

DOMBRY, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* **21** 420–436. MR3322325

DOMBRY, C. and FERREIRA, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli* **25** 1690–1723. MR3961227

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events: for Insurance and Finance*. Springer Science & Business Media.

ENGELKE, S. and HITZ, A. S. (2020). Graphical models for extremes (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **82** 871–932.

GARCÍA, G., DRECCER, M., MIRALLES, D. and SERRAGO, R. (2015). High night temperatures during grain number determination reduce wheat and barley grain yield: a field study. *Global Change Biology* **21** 4153–4164. https://doi.org/10.1111/gcb.13009

GNECCO, N., TEREFE, E. M. and ENGELKE, S. (2022). Extremal Random Forests. https://doi.org/10.48550/ARXIV.2201.12865

GNECCO, N., MEINSHAUSEN, N., PETERS, J. and ENGELKE, S. (2021). Causal discovery in heavy-tailed models. *The Annals of Statistics* **49** 1755–1778.
GREENE, W. (2009). Discrete choice modeling. In *Palgrave Handbook of Econometrics* 473–556. Springer.
GU, J. and VOLGUSHEV, S. (2019). Panel data quantile regression with grouped fixed effects. *Journal of Econometrics* **213** 68–91.
HAMBUCKERS, J. and KNEIB, T. (2021). Smooth Transition Regression Models for Non-Stationary Extremes. *Journal of Financial Econometrics* **nbab005**.
HANSEN, J., RUEDY, R., SATO, M. and LO, K. (2010). Global surface temperature change. *Reviews of Geophysics* **48**. RG4004.
HSIAO, C. (2007). Panel data analysis - advantages and challenges. *Test* **16** 1–22.
HSIAO, C. (2014). *Analysis of Panel Data* **54**. Cambridge University Press.
IPCC (2008). Climate change 2007. Synthesis report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change Core Writing Team, eds. R. K. Pachauri and A. Reisinger.
JURADO, K., LUDVIGSON, S. C. and NG, S. (2015). Measuring uncertainty. *American Economic Review* **105** 1177–1216.
KATZ, R. W. (2013). Statistical methods for nonstationary extremes. In *Extremes in a Changing Climate* 15–37. Springer.
KATZ, R. W., PARLANGE, M. B. and NAVEAU, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources* **25** 1287–1304.
KUKAL, M. S. and IRMAK, S. (2018). Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production. *Scientific Reports* **8**. https://doi.org/10.1038/s41598-018-21848-2
LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer.
MASSACCI, D. (2017). Tail risk dynamics in stock returns: links to the macroeconomy and global markets connectedness. *Management Science* **63** 3072–3089.
MCLACHLAN, G. and KRISHNAN, T. (2008). *The EM algorithm and extensions*, 2nd ed. *Wiley series in probability and statistics*. Wiley, Hoboken, NJ.
MCNEIL, A. J., FREY, R. and EMBRECHTS, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools-Revised Edition*. Princeton University Press.
MERZ, R. and BLÖSCHL, G. (2005). Flood frequency regionalisation—spatial proximity vs. catchment attributes. *Journal of Hydrology* **302** 283 - 306.
MHALLA, L., CHAVEZ-DEMOULIN, V. and DUPUIS, D. J. (2020). Causal mechanism of extreme river discharges in the upper Danube basin network. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **69** 741-764.
MHALLA, L., HAMBUCKERS, J. and LAMBERT, M. (2020). Extremal connectedness and systemic risk of hedge funds. *Working Paper (Available at SSRN)*.
OH, D. H. and PATTON, A. J. (2020). Dynamic Factor Copula Models with Estimated Cluster Assignments. *Working Paper*.
COMMITTEE ON THE GLOBAL FINANCIAL SYSTEM (2001). Stress testing by large financial institutions: current practice and aggregation issues Technical Report, Bank for International Settlements.
OVEREEM, A., BUISHAND, A. and HOLLEMAN, I. (2008). Rainfall depth-duration-frequency curves and their uncertainties. *Journal of Hydrology* **348** 124–134.
PAKEL, C., SHEPHARD, N. and SHEPPARD, K. (2011). Nuisance parameters, composite likelihoods and a panel of GARCH models. *Statistica Sinica* **21** 307–329.
PAPALEXIOU, S. M., AGHAKOUCHAK, A., TRENBERTH, K. E. and FOUFOULA-GEORGIOU, E. (2018). Global, Regional, and Megacity Trends in the Highest Temperature of the Year: diagnostics and Evidence for Accelerating Trends. *Earth's Future* **6** 71–79.
PETERSEN, L. K. (2019). Impact of Climate Change on Twenty-First Century Crop Yields in the U.S. *Climate* **7**.
PICKANDS, J. III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131.
RAND, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* **66** 846–850.
REICH, B. J. and SHABY, B. A. (2019). A Spatial Markov Model for Climate Extremes. *Journal of Computational and Graphical Statistics* **28** 117–126. https://doi.org/10.1080/10618600.2018.1482764
RIBATET, M. (2019). SpatialExtremes: Modelling Spatial Extremes R package version 2.0-7.2.
ROHRBECK, C. and TAWN, J. A. (2021). Bayesian Spatial Clustering of Extremal Behavior for Hydrological Variables. *Journal of Computational and Graphical Statistics* **30** 91–105. https://doi.org/10.1080/10618600.2020.1777139

RÖTTGER, F., ENGELKE, S. and ZWIERNIK, P. (2021). Total positivity in multivariate extremes. Available from https://arxiv.org/abs/2112.14727.
SADOK, W. and KRISHNA JAGADISH, S. V. (2020). The Hidden Costs of Nighttime Warming on Yields. *Trends in Plant Science* **25** 644–651.
SEGAL, G., SHALIASTOVICH, I. and YARON, A. (2015). Good and bad uncertainty: macroeconomic and financial market implications. *Journal of Financial Economics* **117** 369–397.
SMITH, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72** 67-90.
SU, L., SHI, Z. and PHILLIPS, P. C. (2016). Identifying latent structures in panel data. *Econometrica* **84** 2215–2264.
VELTHOEN, J., DOMBRY, C., CAI, J.-J. and ENGELKE, S. (2021). Gradient boosting for extreme quantile regression.
VIGNOTTO, E., ENGELKE, S. and ZSCHEISCHLER, J. (2021). Clustering bivariate dependencies of compound precipitation and wind extremes over Great Britain and Ireland. *Weather and Climate Extremes* **32**. https://doi.org/10.1016/j.wace.2021.100318
WANG, W. and SU, L. (2021). Identifying latent group structures in nonlinear panels. *Journal of Econometrics* **220** 272–295.
WANG, J., HAN, Y., STEIN, M. L., KOTAMARTHI, V. R. and HUANG, W. K. (2016). Evaluation of dynamically downscaled extreme temperature using a spatially-aggregated generalized extreme value (GEV) model. *Climate Dynamics* **47** 2833–2848.
WHITE, H. (1994). *Estimation, Inference and Specification Analysis*. *Econometric Society Monographs No. 22*. Cambridge University Press.
WOOLDRIDGE, J. M. (1994). Estimation and inference for dependent processes. *Handbook of Econometrics* **4** 2639–2738.
ZHAO, Z., ZHANG, Z. and CHEN, R. (2018). Modeling maxima with autoregressive conditional Fréchet model. *Journal of Econometrics* **207** 325–351.
ZWIERS, F. W. and KHARIN, V. V. (1998). Changes in the Extremes of the Climate Simulated by CCC GCM2 under CO2 Doubling. *Journal of Climate* **11** 2200–2222.

Acknowledgments. The authors wish to thank the Associate Editor and two anonymous referees for helpful comments that improved the paper.

Funding. The first author was supported by the Natural Sciences and Engineering Research Council of Canada RGPIN-2016-04114 and the HEC Foundation. The second author was supported by the Swiss National Science Foundation (Grant 186858). The third author was supported by a Modigliani Research Grant of the UniCredit Foundation for the project "On the determinants of time-varying tail risk".

APPENDIX A: CONSISTENCY OF EM ALGORITHM

Let $(\mathbf{Y}_t, \mathbf{X}_t)$ be an $N \times (K+1)$ matrix of random variables. Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ be a P-dimensional vector of parameters and let $\boldsymbol{\tau} \in \mathcal{G}$ be a vector of N integers denoting the assignment of each individual to one of G groups.

We reformulate the iteration of the EM algorithm for the joint estimation of the model parameters and group assignments in Section 2.3 in terms of M-estimators. At the $(j+1)$th iteration, we have,

1.1 **Maximization**. $\widehat{\boldsymbol{\theta}}^{(j+1)} = \arg\min_{\boldsymbol{\theta}} S_T\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\tau}}^{(j)}\right)$

1.2 **Expectation**. $\widehat{\tau}_i^{(j+1)} = \arg\min_{g \in \{1,\dots,G\}} S_T\left(\widehat{\boldsymbol{\theta}}^{(j+1)}, \widehat{\boldsymbol{\tau}}_{i,g}^{(j)}\right)$

where $\widehat{\boldsymbol{\tau}}_{i,g}^{(j)}$ is equal to $\widehat{\boldsymbol{\tau}}^{(j)}$ except that the ith element is equal to g, and

$$S_T(\boldsymbol{\theta}, \boldsymbol{\tau}) = \frac{1}{T} \sum_{t=1}^{T} \rho\left(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\tau}\right)$$

is the objective function used in the M-estimation, with $\rho(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\tau})$ some function of $(\mathbf{Y}_t, \mathbf{X}_t)$, the parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{\tau} \in \mathcal{G}$. The EM estimates $(\widehat{\boldsymbol{\theta}}_T, \widehat{\boldsymbol{\tau}}_T)$ are obtained by iterating these two steps until convergence of the EM algorithm. In Section 2.3, we define the criterion function $\rho(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\tau})$ as

$$\rho(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\tau}) = -\sum_{i=1}^{N} \log h(Y_{it} \mid X_{it}, \boldsymbol{\theta}, \boldsymbol{\tau})$$

where $\log h$ is the log-likelihood of the GEV distribution and X_{it}, $\boldsymbol{\theta}$, and $\boldsymbol{\tau}$ enter the GEV parameters μ, σ, and ξ as in (9).

We provide conditions under which $\left(\widehat{\boldsymbol{\theta}}_T, \widehat{\boldsymbol{\tau}}_T\right)$ are consistent for their true counterparts $(\boldsymbol{\theta}_0, \boldsymbol{\tau}_0)$, where

$$\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}} S(\boldsymbol{\theta}, \boldsymbol{\tau}_0)$$

$$S(\boldsymbol{\theta}, \boldsymbol{\tau}) = \mathbb{E}\left[S_T(\boldsymbol{\theta}, \boldsymbol{\tau})\right] = \mathbb{E}\left[\rho(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\tau})\right].$$

Note that labels attached to the groups are arbitrary and there exists a set of correct group assignments, that we denote $\mathcal{G}_0$. We can thus consistently estimate $\boldsymbol{\tau}_0$ up to re-labeling of the groups.

Assumption 1: $\{(\mathbf{Y}_t, \mathbf{X}_t) : t = 1, 2, \dots\}$ is a stationary ergodic sequence.
Assumption 2: $\boldsymbol{\Theta}$ is compact.
Assumption 3: For each $\boldsymbol{\tau} \in \mathcal{G}$ and $(\mathbf{Y}_t, \mathbf{X}_t)$, $\rho(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\tau})$ is continuous in $\boldsymbol{\Theta}$. For each $\boldsymbol{\tau} \in \mathcal{G}$ and for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\rho(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\tau})$ is measurable.
Assumption 4: For each $\boldsymbol{\tau} \in \mathcal{G}$ and for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, it holds that:

(i) $\mathbb{E}\left[|\rho(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\tau})|\right] < \infty$;
(ii) $\mathbb{E}\left[\|\nabla_{\boldsymbol{\theta}} \rho(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\tau})\|_1\right] < \infty$;
(iii) $\mathbb{E}\left[\|\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \rho(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\tau})\|_1\right] < \infty$.

Assumption 5: For any $\boldsymbol{\tau} \in \mathcal{G}$, let $\eta_{\boldsymbol{\tau}}(\epsilon)$ be an ϵ-neighborhood of $\boldsymbol{\theta}^*(\boldsymbol{\tau})$. For all $\epsilon > 0$,

(i) $\inf_{\boldsymbol{\theta} \in \{\boldsymbol{\Theta} \setminus \eta_{\boldsymbol{\tau}}(\epsilon)\}} S(\boldsymbol{\theta}, \boldsymbol{\tau}) > S(\boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau}), \quad \forall \boldsymbol{\tau} \in \mathcal{G}.$

Moreover, for $\boldsymbol{\tau} = \boldsymbol{\tau}_0$ we have

(ii) $\inf_{(\boldsymbol{\theta}, \boldsymbol{\tau}) \in \boldsymbol{\Theta} \times \{\mathcal{G} \setminus \mathcal{G}_0\}} S(\boldsymbol{\theta}, \boldsymbol{\tau}) > S(\boldsymbol{\theta}_0, \boldsymbol{\tau}_0)$

THEOREM 1. *Under Assumptions 1–5, we have that* $\widehat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0$ *and* $\widehat{\boldsymbol{\tau}}_T \xrightarrow{p} \boldsymbol{\tau}_0$.

Proof. Define the feasible and unfeasible profile estimators

$$\widetilde{\boldsymbol{\theta}}_T(\boldsymbol{\tau}) = \arg\min_{\boldsymbol{\theta}} \frac{1}{T} \sum_{t=1}^{T} \rho(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\tau})$$

$$\boldsymbol{\theta}^*(\boldsymbol{\tau}) = \arg\min_{\boldsymbol{\theta}} S(\boldsymbol{\theta}, \boldsymbol{\tau}).$$

We first show that $\widetilde{\boldsymbol{\theta}}_T(\boldsymbol{\tau}) \xrightarrow{p} \boldsymbol{\theta}^*(\boldsymbol{\tau})$ for a given $\boldsymbol{\tau}$. Theorem 3.4 of White (1994) implies this result if we can show that: (i) $\boldsymbol{\Theta}$ is compact; (ii) $\rho(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\tau})$ is measurable and continuous on $\boldsymbol{\Theta}$; (iii) $\boldsymbol{\theta}^*(\boldsymbol{\tau})$ is the unique minimizer of $S(\boldsymbol{\theta}, \boldsymbol{\tau})$ on $\boldsymbol{\Theta}$; (iv) $S_T(\boldsymbol{\theta}, \boldsymbol{\tau})$ converges to

$S(\boldsymbol{\theta}, \boldsymbol{\tau})$ uniformly on $\boldsymbol{\Theta}$. Conditions (i)-(ii) are satisfied by Assumptions 2 and 3. Condition (iii) is satisfied by Assumption 5(i). To show that condition (iv) is satisfied we verify Theorem 4.1 of Wooldridge (1994): Our Assumption 1 satisfies their stationary and ergodic condition; our Assumptions 2 and 3 match their Assumptions (i)-(ii); our Assumption 4(i) matches their Assumption (iii).

As $\widetilde{\boldsymbol{\theta}}_T(\boldsymbol{\tau}) \xrightarrow{p} \boldsymbol{\theta}^*(\boldsymbol{\tau})$ for each $\boldsymbol{\tau} \in \mathcal{G}$, $\widehat{\boldsymbol{\tau}}_T \xrightarrow{p} \boldsymbol{\tau}_0$ implies $\widehat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0$, see also Theorem 4.3 of Wooldridge (1994). To prove consistency of $\widehat{\boldsymbol{\tau}}_T$ to $\boldsymbol{\tau}_0$ we follow Oh and Patton (2020) showing that (i) $\boldsymbol{\tau}_0$ uniquely minimizes $S(\boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau})$ and (ii) $\mathrm{S}_T(\widetilde{\boldsymbol{\theta}}_T(\boldsymbol{\tau}), \boldsymbol{\tau})$ converges pointwise to $S(\boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau})$ for each $\boldsymbol{\tau} \in \mathcal{G}$. Condition (i) is satisfied by Assumption 5(ii), so we focus on showing condition (ii). We have

$$S_T(\widetilde{\boldsymbol{\theta}}_T(\boldsymbol{\tau}), \boldsymbol{\tau}) - S(\boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau}) = S_T(\widetilde{\boldsymbol{\theta}}_T(\boldsymbol{\tau}), \boldsymbol{\tau}) - S_T(\boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau}) + S_T(\boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau}) - S(\boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau})$$

Under Assumptions 1 and 4(i), the weak law of large numbers guarantees that $S_T(\boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau}) - S(\boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau}) = o_p(1)$. Further,

$$S_T(\widetilde{\boldsymbol{\theta}}_T(\boldsymbol{\tau}), \boldsymbol{\tau}) - S_T(\boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau}) = \frac{1}{T} \sum_{t=1}^{T} \left[\rho\left(\mathbf{Y}_t, \mathbf{X}_t, \widetilde{\boldsymbol{\theta}}_T(\boldsymbol{\tau}), \boldsymbol{\tau} \right) - \rho\left(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau} \right) \right]$$

From a mean value expansion of $\rho\left(\mathbf{Y}_t, \mathbf{X}_t, \widetilde{\boldsymbol{\theta}}_T(\boldsymbol{\tau}), \boldsymbol{\tau} \right)$ around $\boldsymbol{\theta}^*(\boldsymbol{\tau})$ we obtain

$$\begin{aligned} S_T(\widetilde{\boldsymbol{\theta}}_T(\boldsymbol{\tau}), \boldsymbol{\tau}) - S_T(\boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau}) = & \left(\tfrac{1}{T} \textstyle\sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} \rho\left(\mathbf{Y}_t, \mathbf{X}_t, \boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau} \right) \right)' \left(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\tau}) - \boldsymbol{\theta}^*(\boldsymbol{\tau}) \right) \\ & + \tfrac{1}{2} \left(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\tau}) - \boldsymbol{\theta}^*(\boldsymbol{\tau}) \right)' \left(\tfrac{1}{T} \textstyle\sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \rho\left(\mathbf{Y}_t, \mathbf{X}_t, \dot{\boldsymbol{\theta}}(\boldsymbol{\tau}), \boldsymbol{\tau} \right) \right) \left(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\tau}) - \boldsymbol{\theta}^*(\boldsymbol{\tau}) \right) \end{aligned}$$

where $\dot{\boldsymbol{\theta}}(\boldsymbol{\tau}) = \lambda \widetilde{\boldsymbol{\theta}}(\boldsymbol{\tau}) + (1 - \lambda) \boldsymbol{\theta}^*(\boldsymbol{\tau})$ for some $\lambda \in [0, 1]$. Assumptions 1, 5(ii), 5(iii) guarantee that the gradient and hessian terms converge in probability to a finite limit by the weak law of large numbers. Since $\widetilde{\boldsymbol{\theta}}(\boldsymbol{\tau}) \xrightarrow{p} \boldsymbol{\theta}^*(\boldsymbol{\tau})$, we have $S_T(\widetilde{\boldsymbol{\theta}}_T(\boldsymbol{\tau}), \boldsymbol{\tau}) - S_T(\boldsymbol{\theta}^*(\boldsymbol{\tau}), \boldsymbol{\tau}) = o_p(1)$.

APPENDIX B: A GROUPED PANEL GP REGRESSION MODEL

We outline here an approach to extend the methodology of this paper to grouped panels of generalized Pareto (GP) distributions. Let $(Z_{i,t}, \mathbf{X}_{i,t})$ be the observations for the ith individual at time t, with $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Here, the $Z_{i,t}$ can be seen as daily observations as opposed to the $Y_{i,t}$ that represented block maxima over a certain block size (e.g., a year). For an intermediate probability level p_0, the GP distribution (Pickands, 1975) is used to model the exceedances $(Z_{i,t} \mid Z_{i,t} > Q_{i,t}^{p_0}, \mathbf{X}_{i,t})$ over its conditional intermediate quantile $Q_{i,t}^{p_0}$ at level p_0 given the covariate vector $\mathbf{X}_{i,t}$. More precisely, the distribution function of these exceedances is approximated by a GP distribution with covariate dependent parameters given by

$$W(z \mid \sigma_{i,t}, \xi_{i,t}) = 1 - \left(1 + \frac{\xi_{i,t}}{\sigma_{i,t}} z \right)_{+}^{-1/\xi_{i,t}}, \quad z > 0,$$

where $\sigma_{i,t} > 0$ and $\xi_{i,t} \in \mathbb{R}$ are the conditional scale and shape parameters, respectively. To model a grouped panel GP regression, similarly to (8), we may assume that these parameters satisfy

$$\begin{aligned} \sigma_{i,t}(\tau_i) &= e_\sigma\left(\boldsymbol{\gamma}_{(\tau_i)}^\top \mathbf{X}_{i,t} \right) \\ \xi_{i,t}(\tau_i) &= e_\xi\left(\boldsymbol{\delta}_{(\tau_i)}^\top \mathbf{X}_{i,t} \right) \end{aligned} \tag{A.1}$$

where $\tau_i \in \{1, \ldots, G\}$ denotes the group membership of the ith individual and $\boldsymbol{\theta}_{(g)} = \left(\boldsymbol{\gamma}_{(g)}, \boldsymbol{\delta}_{(g)}\right) \in \boldsymbol{\Theta} \subset \mathbb{R}^P$ is the vector of regression parameters for the gth group, $g \in \{1, \ldots, G\}$.

We can then employ an EM algorithm, similar to the one of the grouped panel GEV model in Section 2.3, that is based on the exceedances of the data set $(Z_{i,t}, \mathbf{X}_{i,t})$, $i = 1, \ldots, N$, $t = 1, \ldots, T$, over their respective p_0 quantiles. That is, we would replace the GEV log-likelihoods in the EM algorithm by the log-likelihood of the GP distribution given by

$$\log w(z \mid \sigma_{i,t}, \xi_{i,t}) = -\log \sigma_{i,t} - \left(1 + \frac{1}{\xi_{i,t}}\right) \log \left\{1 + \frac{\xi_{i,t}}{\sigma_{i,t}}(Z_{i,t} - Q_{i,t}^{p_0})\right\}.$$

A difficulty in this panel GP model is that it relies on a two-step procedure. First one needs a good estimation of the intermediate conditional quantiles $Q_{i,t}^{p_0}$ to define the exceedances that are provided to the EM algorithm. This can either be done by an additional panel quantile regression (Gu and Volgushev, 2019), or with any other quantile regression method that the modeler considers suitable.