Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

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

Bayesian extreme value models: Asymptotic behavior, hierarchical convergence, and predictive robustness

Omid M. Ardakani

Associate Professor of Economics, Parker College of Business, Georgia Southern University, Savannah, GA

ARTICLE INFO

Article history:
Received 12 November 2023
Revised 12 April 2024
Accepted 13 April 2024
Available online xxx

JEL classification:
C11
C46
G32
Q54

Keywords:
Bayesian analysis
Extreme value theory
Hierarchical Bayesian modeling
Predictive distribution
Prior sensitivity
Tail risk

ABSTRACT

The efficacy of Bayesian extreme value models is examined, with a focus on their ability to analyze tail behavior and their predictive accuracy. The hierarchical structure in Bayesian extreme value models helps model heterogeneity by borrowing strength across groups. The theoretical results indicate that the Kullback-Leibler divergence between posteriors drawn from various priors is bounded, reinforcing the stability of Bayesian extreme value models against prior selections. As the sample size grows, the influence of priors diminishes. In addition, the results establish that the hyperposterior distribution converges to a specific distribution as group size increases. This robustness is restricted to prior assumptions and integrates into its hierarchical structure. Another finding confirms the convergence of predictive distributions, especially for big data analyses. The advantage of hierarchical modeling over non-hierarchical counterparts is underscored as the variance component of the predictive loss function diminishes in hierarchical Bayesian extreme value models.

1. Introduction

Extreme value (EV) theory studies the behavior of tails and extreme events that traditional modeling cannot fully capture. Conventional methodologies sometimes fail to address econometric issues these extreme values present (Silverberg and Verspagen, 2007). The synthesis of Bayesian methods and extreme value theory has emerged as a solution, offering adaptability and capturing uncertainty. This paper studies Bayesian EV models, mainly concentrating on their asymptotic behaviors, the role of prior selections, and the significance of hierarchical structures.

Extreme event modeling has been extensively studied. Coles et al. (2001); Embrechts et al. (2013), and Novak (2011) offer foundations and applications of EV theory in various fields. Recently, the literature has incorporated information theory in capturing information and uncertainty in extreme events. Pons et al. (2020); Asadi et al. (2010), and Ardakani (2023) employ information-theoretic measures to bridge the gap between information theory and EV theory. Incorporating Bayesian methods into EV theory addresses some challenges in traditional approaches. As outlined by Gelman et al. (2013), the Bayesian paradigm offers a framework for incorporating prior information and beliefs, enabling more informed inference, especially in situations with limited data. Such situations are common in EV theory, which often focuses on extreme events. Further extending the benefits of Bayesian analysis, Ghosh et al. (2006) emphasizes the adaptability of Bayesian methods in adjust-

E-mail address: oardakani@georgiasouthern.edu

https://doi.org/10.1016/j.ecosta.2024.04.002
2452-3062/

ing to complex models and handling dependencies. These features are particularly beneficial when modeling extreme events with underlying processes.

The selection of prior distributions is influenced by a combination of expertise and formal rules. For instance, Jeffreys (1946) proposes an invariant form for the prior probability, which applies to various estimation scenarios for robustness. Furthermore, the systematic evaluation of prior selection methodologies by Kass and Wasserman (1996) underscores the importance of the prior choice in determining the posterior characteristics. In the theory of EV, prior beliefs play a crucial role due to the scarcity of data and the need to rely on them to "borrow strength." Connecting Bayesian methods with EV theory, Coles and Tawn (1991) present a framework for modeling extreme multivariate events, utilizing Bayesian principles to capture dependencies between variables. Their methodology highlights the synergy between Bayesian analysis and extreme event modeling. Similarly, in the context of high-frequency financial data, Mao and Zhang (2018) adopted a Bayesian perspective for a stochastic tail index model, resulting in enhanced tail behavior characterization. Meanwhile, Gelman and Hill (2006) provide insights into the hierarchical models, emphasizing their importance in accounting for the different layers of variability and complexity in the data. This feature is valuable in modeling extreme events.

The literature has made significant progress in applying Bayesian methods to extreme events and examining prior selection and hierarchical structures, but some areas still need to be studied. Specifically, the sensitivity of posterior distributions to the choice of priors in the context of EV theory remains an open question. Similarly, the asymptotic behavior of hyperposterior distributions, particularly within Bayesian EV models, hasn't been exhaustively studied. Additionally, while the advantages of hierarchical models in Bayesian analysis are acknowledged, there is a lack of comprehensive studies on their predictive performance. Addressing these gaps can illustrate how Bayesian methodologies and extreme value theory mutually reinforce one another. This study addresses these areas and elaborates on Bayesian inference for extreme value modeling.

The contributions of this study are as follows. First, it is highlighted that different choices of priors can influence the posteriors. The Kullback-Leibler (KL) divergence measures how posterior distributions derived from different priors vary. A result establishes an upper bound on the KL divergence between posteriors from two distinct priors. Using the Bernstein-von Mises theorem, it is shown that sample size influences this sensitivity. The implication of this finding is for large datasets where the influence of prior selection on inferences becomes less important. Second, the convergence of hierarchical modeling in the context of extreme values is illustrated. Specifically, the stability of hierarchical Bayesian EV models increases as the number of groups increases. A result shows the asymptotic behavior of the hyperposterior distributions in Bayesian EV models. The rate of convergence for the hyperposterior is provided after establishing its consistency under certain regularity conditions. Third, the predictive performance of hierarchical Bayesian EV models is studied. Predictive loss can be decomposed into bias, variance, and noise. The shrinkage is the advantage of hierarchical Bayesian models, especially when dealing with extreme events.

Simulation studies, examples, and empirical analysis illustrate the theoretical findings. The empirical study employs a hierarchical Bayesian model to examine stress testing of financial institutions. This application focuses on extreme negative returns and uses the Peaks Over Threshold method to study the tail risk characteristics. It then incorporates external monetary policy shocks and illustrates their impact on internal vulnerabilities (negative returns) and external shocks (monetary policy changes). This approach provides a view of potential risks, capturing intrinsic and extrinsic factors.

The paper is organized as follows. Section 2 studies the sensitivity of Bayesian posterior distributions to the selection of prior distribution. An upper bound on the KL divergence between posteriors resulting from two different priors is formulated in this section. The effect of priors on the posteriors is illustrated using an example employing the Gumbel distribution. Section 3 elaborates on hierarchical modeling within Bayesian EV analysis. This section examines the asymptotic behavior of the hyperposterior distribution, including its consistency and convergence rate. In addition, the predictive performance of hierarchical Bayesian EV models compared to non-hierarchical counterparts is examined in this section. Section 4 presents an empirical illustration emphasizing the stress testing of financial institutions and the modeling of monetary policy shocks. Section 5 provides concluding remarks and future research directions.

2. Bayesian extreme value models

2.1. Preliminaries

EV theory focuses on the tails of probability distributions. These tails represent extreme events. Coles and Powell (1996) demonstrate the difficulties in combining Bayesian inference with extreme value models. However, the Bayesian approach's potential in extreme value modeling cannot be overlooked. Given the lack of extreme data, the inclusion of prior knowledge through Bayesian methods can significantly enhance inferential quality (Gelman et al., 2013). Smith (1987) and Smith (1990) contribute towards the development of extreme value models and demonstrate the connection between the generalized extreme value (GEV) and the generalized Pareto (GP) distributions. The GEV and GP belong to the broader family of extreme value distributions. These distributions are used in the statistical analysis of extremes, sharing connections regarding applications and formulations (Coles et al., 2001; Embrechts et al., 2013).

1. **The GEV distribution** is pivotal for modeling block maxima. A "block" typically refers to a period over which maxima are observed and collected, such as daily, monthly, or yearly maxima in a time series. This definition is necessary for apply-

ing the GEV distribution, as it is designed to model the maximum (or minimum) value observed over predefined blocks of time or other units. Under specific conditions related to the distribution of independent and identically distributed random variables, the block maxima can be approximated using the GEV distribution (Leadbetter et al., 2012).The cumulative distribution function (CDF) of the GEV, contingent on the shape parameter ξ, is given by

$$F(z) = \begin{cases} \exp\left(-(1+\xi z)^{-1/\xi}\right) & \text{for } \xi \neq 0 \\ \exp\left(-\exp(-z)\right) & \text{for } \xi = 0, \end{cases} \tag{1}$$

where the standardized variable is given by $z = (x-\mu)/\sigma$, where μ is the location parameter and $\sigma > 0$ is the scale parameter.

2. **The GP distribution** is central to modeling exceedances over high thresholds. Due to its statistical properties tailored to this purpose, the GP is often the preferred distribution for such analyses (Pickands, 1975). Its CDF is given by

$$F(z) = \begin{cases} 1 - \left(1+\xi z\right)^{-1/\xi} & \text{for } \xi \neq 0 \\ 1 - \exp(-z) & \text{for } \xi = 0. \end{cases} \tag{2}$$

The Pickands-Balkema-de Haan theorem (Balkema and Haan, 1974; Pickands, 1975) states that if GEV applies to block maxima for a large n, then excesses over a high threshold are approximately GP distributed. Conversely, if GP applies to exceedances over a high threshold, then the maxima of block maxima will be GEV distributed. Asadi et al. (2010) develops information optimal models for the joint distribution that give information characterizations of extreme value distributions. Ardakani (2023) recently shows that the entropy of the block maxima converges to the entropy of the GEV distribution as the block size increases. Bayesian methods integrated into EV theory are useful for predictive accuracy, especially in tail risk scenarios by quantifying uncertainty, though conjugate priors for GEV are absent (Coles and Powell, 1996).

Bayesian approaches provide a flexible framework for estimating the parameters of GEV and GP models by incorporating prior information. The likelihood for a dataset $y_1, y_2, \ldots, y_n$ in the GEV context is given by

$$L(\mu, \sigma, \xi | y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} f(y_i | \mu, \sigma, \xi), \tag{3}$$

where f is the GEV probability density function (PDF). Similarly, for the GP, the likelihood for exceedances $y_1, y_2, \ldots, y_n$ over a threshold u is defined by

$$L(\sigma, \xi | y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} f(y_i - u | \sigma, \xi), \tag{4}$$

where f is the GP PDF.

The GEV distribution, encompassing the Gumbel, Fréchet, and Weibull families, satisfies the likelihood function's Lipschitz continuity due to its derivation from the block maxima approach, which is central to extreme value theory. Similarly, the GP distribution's role as a model for excesses over a threshold provides a natural fit for the continuity of the priors, as it is derived from the Pickands-Balkema-de Haan theorem, which assures its applicability to a wide range of tail behaviors. Both GEV and GP distributions, parameterized by location, scale, and shape parameters, offer flexible modeling of extreme events, accommodating a wide range of tail behaviors essential for capturing the distinct characteristics of each group within the hierarchical model. Also, the derivatives of the likelihood function with respect to the parameters are bounded within the parameter space, satisfying the regularity conditions necessary for the convergence of the hyperposterior distribution.

Choosing suitable priors is important for the Bayesian approach. This becomes more significant for extreme value models as it can significantly influence posterior distributions and predictions. Robert (2007) discusses prior selection and suggests certain priors may interact with the likelihood, potentially leading to unexpected results. Meanwhile, Gelman et al. (2013) discuss models' sensitivity with respect to their priors. Even though they do not specifically consider extreme value models, the general principles are broadly applicable. Additionally, Davison et al. (2012) discuss sensitivity to prior choice in extreme value modeling, concluding the need for robust model assessment using diagnostic tools to ensure that the results are independent of the prior choice. Scarrott and MacDonald (2012) study Bayesian hierarchical models for extreme events and discuss prior selection in this context.

2.2. Prior sensitivity and KL divergence

How does the selection of prior influence the inferential outcomes? As emphasized by Jaynes (2003), Bayesian models incorporate prior beliefs, and the choice of these priors can potentially affect the posterior distribution and, consequently, our conclusions about the data. Kass and Wasserman (1996) highlight the importance of reference Bayesian tests about prior selection. Gelman et al. (2013) underscore the necessity of examining the sensitivity of Bayesian models to priors. The sensitivity of posterior distributions reflects how a Bayesian model's inferences might differ when different prior beliefs or information is incorporated. The KL divergence of Kullback and Leibler (1951) is one tool that quantifies this difference between two probability distributions. The applications of this measure in statistics, econometrics, and finance can be seen in Soofi and Retzer (2002); Ardakani and Saenz (2022), and Ardakani (2022). The KL divergence has been used as loss and risk functions in various estimation problems (Ardakani et al., 2018). In the Bayesian context, it can measure the difference between two posteriors derived from distinct priors. A smaller KL divergence indicates that the prior choice has less influence on the posterior, suggesting the data is highly informative. The KL divergence between two distributions p and q is defined as

$$\mathcal{K}(p;q) = \int p(x)\log\left(\frac{p(x)}{q(x)}\right)\mathrm{d}x. \tag{5}$$

$\mathcal{K}(p;q) = 0$ for $p = q$. As they deviate, the KL divergence increases.

As highlighted by Bernardo (1979), the KL divergence has been pivotal in Bayesian model comparison and selection. The divergence provides a metric for model adequacy, measuring how one statistical model approximates another. Furthermore, Kullback (1959) introduced the concept, emphasizing its value in quantifying the inefficiency of using a distribution q to approximate the true distribution p. The relevance of KL divergence in Bayesian statistics is further elaborated by Cover and Thomas (1991), who discuss its applications in information theory and its usefulness in characterizing the difference between two distributions.

Proposition 1 (Prior sensitivity and KL divergence bound)**.** *Let $p(\theta|y,\pi)$ be the posterior distribution from a Bayesian EV model under a specified prior, π. Assume that:*

1. *The likelihood $p(y|\theta)$ is Lipschitz continuous in θ, i.e., there exists a Lipschitz constant C such that*
$$|p(y|\theta_1) - p(y|\theta_2)| \le C|\theta_1 - \theta_2|,$$
for any θ_1, θ_2 in the domain of θ, almost surely with respect to y.
2. *Priors π_A and π_B are absolutely continuous with respect to a σ-finite measure (e.g., the Lebesgue measure), ensuring the existence and smoothness of their density functions.*

Then, for any two priors, π_A and π_B, the Kullback-Leibler divergence between their posteriors is bounded almost surely by a function $f : Y \times \Pi \times \Pi \to \mathbb{R}^+$, crucial for assessing the robustness and sensitivity of the Bayesian EV model against variations in the priors.

$$\mathcal{K}\big(p(\theta|y,\pi_A);\, p(\theta|y,\pi_B)\big) \le f(y,\pi_A,\pi_B) \tag{6}$$

where,

$$f(y,\pi_A,\pi_B) = \max_{\theta}\left|\log\left(\frac{p(\theta|y,\pi_A)}{p(\theta|y,\pi_B)}\right)\right|. \tag{7}$$

Proof. Using (5),

$$\mathcal{K}\big(p(\theta|y,\pi_A);\, p(\theta|y,\pi_B)\big) = \int p(\theta|y,\pi_A)\log\left(\frac{p(\theta|y,\pi_A)}{p(\theta|y,\pi_B)}\right)\mathrm{d}\theta,$$

where the integral is taken with respect to the parameter θ and not with respect to the data y. By Bayes' theorem,

$$p(\theta|y,\pi) = \frac{p(y|\theta)\pi(\theta)}{p(y|\pi)},$$

where $p(y|\pi)$ is the marginal likelihood. Given the two assumptions, the ratio of the two posteriors, $p(\theta|y,\pi_A)/p(\theta|y,\pi_B)$, is bounded due to the boundedness of the likelihood and the absolute continuity of the priors. Given (7), f provides an upper bound for the KL divergence between the two posteriors almost surely. □

Function $f(y,\pi_A,\pi_B)$ provides an upper bound for the KL divergence between the two posteriors. This bound should be interpreted cautiously, especially when the maximum is attained at a θ with low posterior probabilities under both π_A and π_B. The upper bound serves as a diagnostic tool to evaluate the model's sensitivity to changes in the prior assumptions. In cases where the maximum is at a θ with low posterior probability, it is key to analyze the nature of such θ values and their neighborhoods to comprehend how such points affect the overall discrepancy between the posteriors derived under different priors.

In Proposition 1, we establish two assumptions crucial for the application of Bayesian extreme value models. The first assumption pertains to the likelihood function's Lipschitz continuity in parameter θ, which ensures that small changes in θ lead to proportionally small changes in the likelihood function. This property is critical for the stability and convergence of Bayesian inference procedures. The GEV distribution, characterized by its max-stability property, satisfies this assumption. Max-stability implies that the distribution of block maxima remains within the GEV family, providing a controlled and predictable behavior of the likelihood function as θ varies, which is essential for the Lipschitz condition.

The second assumption concerns the continuity of the priors. This assumption is fundamental to ensuring that the posterior distribution is well-defined and that Bayesian updating can proceed smoothly. In the context of extreme value theory, priors are typically chosen to reflect the domain knowledge about the tails of the distribution, such as the rate of decay of tail probabilities. The GP distribution used for threshold exceedance modeling aligns with this assumption. The GP distribution is derived under the Peak Over Threshold method, which models excesses over a high threshold. The continuity of the GP distribution ensures that the priors placed on its parameters can be continuous, adhering to the second assumption.

Furthermore, the GEV distribution's max-stability and the GP distribution's suitability for threshold exceedance modeling are not merely theoretical properties. Empirically, these characteristics have been observed in a wide range of extreme event data, from environmental phenomena like rainfall and flood levels to financial market risks such as stock market crashes. This empirical evidence further justifies our reliance on these distributions and the underlying assumptions in extreme value analysis. To provide a concrete example, consider the application of the GEV distribution to model annual maximum rainfall. The max-stability of the GEV ensures that even as we aggregate data over longer periods, the distribution of the annual maximum remains within the GEV family, satisfying the Lipschitz continuity assumption in a practical setting. Similarly, when using the GP distribution to model exceedances over a high water level threshold in flood risk analysis, the continuity of the prior over the shape and scale parameters of the GP distribution is crucial for accurately reflecting the tail behavior and ensuring robust Bayesian inference.

Example 1. The sensitivity of posterior distributions to prior choices is illustrated in this example within a Bayesian framework, focusing on EV modeling. The Gumbel distribution is selected for its prevalence in modeling the maxima of independent and identically distributed random variables, a common scenario in extreme value theory. Its suitability for modeling block maxima from various data types makes it a pertinent choice for demonstrating Bayesian EV modeling principles. The CDF of Gumbel is given by

$$F(y) = \exp\left(-\exp\left(-\frac{y-\mu}{\beta}\right)\right), \tag{8}$$

where μ represents the location parameter and $\beta > 0$ the scale parameter.

Synthetic data are simulated to resemble real-world extreme event analyses, where true parameter values often remain unknown. By assuming $\mu_{\text{true}} = 0$ and $\beta_{\text{true}} = 1$, we create a controlled environment to observe the influence of prior beliefs on posterior outcomes. These values are chosen to reflect standard settings in EV analyses, providing a familiar basis for comparison and interpretation within the field. To explore prior sensitivity, two priors are selected that represent different levels of informativeness:

- Prior A (π_A): A broad, less informative prior with $\mu \sim \mathcal{N}(0, 10)$ and $\beta \sim \mathcal{U}(0, 5)$. This prior reflects a state of relative uncertainty or weak prior beliefs about the parameters, mimicking situations where little prior information is available.
- Prior B (π_B): A more informative prior with $\mu \sim \mathcal{N}(0, 2)$ and $\beta \sim \mathcal{U}(0, 2)$. This choice represents stronger prior beliefs, potentially derived from previous studies or domain expertise, influencing the posterior more significantly.

The posterior distributions for μ and β were estimated using the Markov Chain Monte Carlo (MCMC) Metropolis algorithm, with a setup of 2,000 iterations and a thinning factor of 10, leading to an acceptance rate of approximately 0.554. To quantify the influence of the chosen priors on the resulting posterior distributions, we computed the KL divergence, yielding a value of 0.012. This computation, referred to as "discrete estimation," involves numerical approximation of the integral defining KL divergence, utilizing discrete samples from the MCMC procedure. Furthermore, we calculated the maximum value of the function f, which serves as an upper bound for the KL divergence, obtaining a value of 3.419. This metric helps assess the sensitivity of the Bayesian EV model to variations in prior beliefs. The results—KL divergence of 0.0116 and a maximum f value of 3.419—indicate the impact of prior selection on posterior inferences in extreme value modeling, underscoring the importance of careful prior specification in Bayesian analysis.

Lemma 1 (Data-driven bound and posterior consistency)**.** *Let Θ be a parameter space endowed with a σ-algebra and consider two priors π_A, π_B defined on Θ. Assume the following regularity conditions hold:*

1. *The likelihood function $p(y|\theta)$ is measurable, bounded, and continuous for all $\theta \in \Theta$.*
2. *The priors π_A and π_B are absolutely continuous with respect to a σ-finite measure μ on Θ, and both priors assign positive probability to any open neighborhood of the true parameter value $\theta_0 \in \Theta$.*
3. *The statistical model defined by $p(y|\theta)$ is identifiable; that is, if $\theta_1 \neq \theta_2$, then $p(y|\theta_1) \neq p(y|\theta_2)$ for some y.*

Then, for any true parameter value $\theta_0 \in \Theta$, as the sample size $n \to \infty$, the KL divergence between the posterior distributions corresponding to π_A and π_B converges to zero:

$$\lim_{n\to\infty} \mathcal{K}\Big(p(\theta|y, \pi_A); p(\theta|y, \pi_B)\Big) = 0. \tag{9}$$

Proof. Let $p_n(\theta|y, \pi)$ denote the posterior distribution based on a sample of size n. Under the posterior consistency, for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n > N$,

$$\mathbb{P}_{\theta_0}\Big(\mathcal{B}(\theta_0; \epsilon)\Big) > 1 - \epsilon,$$

where $\mathcal{B}(\theta_0; \epsilon)$ is a ball of radius ϵ centered at θ_0. Given this, for any $\delta > 0$, choose $\epsilon > 0$ small enough such that for all $n > N$,

$$\int_{\Theta\setminus\mathcal{B}(\theta_0;\epsilon)} p_n(\theta|y, \pi_A) d\theta < \delta$$

and

$$\int_{\Theta\setminus\mathcal{B}(\theta_0;\epsilon)} p_n(\theta|y, \pi_B) d\theta < \delta.$$

As $p_n(\theta|y, \pi_A)$ and $p_n(\theta|y, \pi_B)$ both concentrate around θ_0 as $n \to \infty$, the KL divergence between them tends to zero, i.e.,

$$\lim_{n\to\infty} \mathcal{K}\Big(p_n(\theta|y, \pi_A);\, p_n(\theta|y, \pi_B)\Big) = 0.$$

□

Lemma 1 emphasizes the role of posterior consistency and is general and not specific to EV models. When applying to EV models, the dataset's richness often varies and, in some instances, may be substantial. In such cases, the insights from this lemma become particularly relevant as they imply that different prior selections will eventually converge to similar posteriors, giving analysts more confidence in the model's robustness against the choice of priors. The prior choice can significantly affect the initial conclusions drawn from EV models, especially with smaller sample sizes. By integrating this knowledge, we can analyze the transient influence of prior specifications and ensure the robustness of the EV model's inferences in diverse data scenarios. For a clearer understanding of posterior consistency and its implications, refer to Ghosh et al. (2006) and Van der Vaart (2000).

3. Multivariate EV models and hierarchical Bayesian

This section integrates hierarchical structures, focusing on multivariate EV models. Multivariate EV models represent extreme events occurring simultaneously in different dimensions, providing insights into the dependencies amongst them (Coles and Tawn, 1991). By introducing hierarchical priors, hierarchical Bayesian models allow a representation of uncertainties and dependencies within different groups or datasets. The hierarchical models leverage a hierarchical prior that encompasses both group-specific parameters, θ_i, and a common hyperparameter, ϕ, shared across all groups. This shared hyperparameter, ϕ, allows the model to exploit information from all observations in each group to learn more about commonalities, thus improving the overall estimation.

3.1. Hierarchical model convergence

In a hierarchical Bayesian model applied to EV theory, the hierarchical prior models the dependencies and uncertainties among various groups. Here, the parameter θ_i is specific to each group, while the hyperparameter ϕ is common across all groups, providing a structured approach to borrow information across different groups or datasets. This mechanism is beneficial when data from multiple sources, regions, or periods exhibit diverse patterns, enabling the model to consider each dataset's specific characteristics while leveraging broader patterns from the collective analysis.

Proposition 2 (Hierarchical model convergence)**.** *Consider a hierarchical Bayesian model for extreme value analysis, comprising k groups, each with n_i independent and identically distributed observations based on a finite-dimensional parameter $\theta_i \in \mathbb{R}^d$. Let $\mathbf{y} = \{y_{i1}, \ldots, y_{in_i}\}_{i=1}^k$ denote the observed data, where y_{ij} represents the j-th observation in the i-th group. The model assumes the following:*

1. *Each group's observations are modeled by a distinct but related distribution, parameterized by θ_i, and are conditionally independent given θ_i.*
2. *The hierarchical structure is introduced through a hyperparameter ϕ, influencing all group-specific parameters θ_i.*
3. *The likelihood function $p(\mathbf{y}|\{\theta_i\}_{i=1}^k, \phi)$ is Lipschitz continuous and sufficiently smooth in θ_i and ϕ, with all required derivatives bounded.*
4. *The hyperprior $\pi(\phi)$ is proper, providing a regularizing effect and ensuring the well-posedness of the posterior distribution.*

Under these conditions, as $k \to \infty$ and $\min(n_i) \to \infty$, the hyperposterior distribution $p(\phi|\mathbf{y})$ converges in distribution to a Gaussian distribution centered around the true hyperparameter ϕ_{true}, with convergence rates dependent on both k and n_i.

Proof. We begin by expressing the joint posterior distribution of θ and ϕ, given the observed data $\mathbf{y}$, as

$$p(\theta, \phi|\mathbf{y}) \propto p(\mathbf{y}|\theta) \prod_{i=1}^{k} \pi_{\theta_i|\phi}(\theta_i)\pi_\phi(\phi).$$

This expression captures the hierarchical structure of the model, where $\pi_{\theta_i|\phi}$ represents the prior distribution of θ_i conditioned on the hyperparameter ϕ, and $\pi_\phi(\phi)$ denotes the hyperprior. To derive the hyperposterior distribution $p(\phi|\mathbf{y})$, we integrate out the θ parameters:

$$p(\phi|\mathbf{y}) = \int p(\theta, \phi|\mathbf{y})\, d\theta.$$

This integral reflects the marginalization over the group-specific parameters to focus on the distribution of the hyperparameter ϕ. Given the Bayesian Central Limit Theorem and assuming that the likelihood function and its derivatives are bounded and regular, the primary contribution to the joint distribution of ϕ and θ_i for large k is derived from the likelihood function $p(\mathbf{y}|\theta)$ through its influence on θ_i. Hence, we approximate the hyperposterior as

$$p(\phi|\mathbf{y}) \approx \prod_{i=1}^{k} p(y_i|\mu_{\theta_i}, \phi)\pi_\phi(\phi),$$

where each term $p(y_i|\mu_{\theta_i}, \phi)$ represents the likelihood of the data in the i-th group, given the hyperparameter ϕ. To further analyze the behavior of $p(\phi|\mathbf{y})$, we perform a Taylor expansion around the mode $\hat{\phi}$ of the log hyperposterior:

$$\log p(\phi|\mathbf{y}) \approx \log p(\hat{\phi}|\mathbf{y}) - \frac{1}{2}(\phi - \hat{\phi})^T \mathbf{H}(\phi - \hat{\phi}) + \mathcal{O}((\phi - \hat{\phi})^3),$$

where $\mathbf{H}$ is the Hessian matrix evaluated at $\hat{\phi}$, and $\mathcal{O}((\phi - \hat{\phi})^3)$ indicates higher-order terms. The dominance of the quadratic term for sufficiently large k, combined with the controlled behavior of higher-order terms, allows us to approximate the hyperposterior distribution as Gaussian in the vicinity of $\hat{\phi}$, with mean $\hat{\phi}$ and variance given by $-\mathbf{H}^{-1}$

$$p(\phi|\mathbf{y}) \approx \exp\left(\log p(\hat{\phi}|\mathbf{y}) - \frac{1}{2}(\phi - \hat{\phi})^T \mathbf{H}(\phi - \hat{\phi})\right).$$

This approximation underscores the Gaussian convergence of the hyperposterior distribution under the specified regularity conditions. □

It is important to note that the assumptions outlined in the proposition are consistent with the properties of GEV and GP distributions commonly used in extreme value analysis. Specifically, the GEV and GP distributions' flexibility, continuity, and bounded derivative properties ensure their compatibility with the hierarchical model's requirements.

This proposition introduces extreme value modeling within hierarchical Bayesian models. This framework builds upon Hoff (2009) to address the convergence properties of hyperposteriors within hierarchical setups. The consideration of the relative growth rates of the number of groups (k) and the number of observations per group (n_i) in this analysis echoes the results in Gelman and Hill (2006) and extends these discussions by explicitly focusing on extreme value distributions. Furthermore, the explicit handling of group-specific parameters alongside a common hyperparameter aligns with recent methodological developments discussed in Lavine et al. (2021). By providing a theoretical framework for the Gaussian convergence of hyperposteriors in hierarchical extreme value models, this proposition offers practical insights for practitioners dealing with multi-level extreme event data. Proposition 1 and Proposition 2 discuss the influence of priors. Proposition 1 evaluates the influence of different priors on a single model's posterior through the KL divergence, while Proposition 2 looks at the influence of hyperpriors on the convergence of the hyperposterior as more data (from more groups) becomes available.

Example 2. This example explores extreme weather patterns, specifically focusing on modeling the maximum daily rainfall in a collection of countries, denoted as $\mathcal{C}$. A hierarchical Bayesian EV model is employed using the Gumbel distribution suitable for block maxima analysis such as annual maximum rainfall events. For each country $i \in \mathcal{C}$, the maximum daily rainfall y_i is modeled as

$$y_i|\theta_i \sim GEV(\mu_i, \beta_i, \xi = 0),$$

where μ_i and β_i represent the location and scale parameters, respectively, and are encapsulated within $\theta_i = (\mu_i, \beta_i)$. The influence of broader climatic conditions on these parameters is modeled through a hyperparameter ϕ, leading to

$$\theta_i|\phi \sim f(\theta_i; \phi),$$

where f denotes the distribution function reflecting ϕ's influence.

The distributional aspects of this model is detailed as follows:

1. **Model prior distribution ($\pi_{\theta_i|\phi}$):** Specifies prior beliefs about the rainfall model parameters θ_i, conditioned on global climatic conditions ϕ.
2. **Hyperprior distribution (π_{ϕ}):** Defines prior assumptions about the hyperparameter ϕ, representing overarching climatic influences.
3. **Likelihood function ($p(\mathbf{y}|\theta)$):** Describes how the observed data $\mathbf{y}$ is generated from the model, given parameters θ_i.
4. **Joint posterior distribution ($p(\theta, \phi|\mathbf{y})$):** Combines the prior and likelihood information to update beliefs about the model parameters and hyperparameters based on observed data.

The hierarchical structure of the model is captured by the joint posterior distribution, which aggregates information across all countries:

$$p(\theta, \phi|\mathbf{y}) \propto p(\mathbf{y}|\theta) \prod_{i=1}^{k} \pi_{\theta_i|\phi}(\theta_i) \pi_{\phi}(\phi).$$

Data are simulated from the Gumbel distribution parameters ($\mu = 0$, $\beta = 1$) for 100 countries. The simulation assesses how ϕ influences the posterior estimates of θ_i as k increases. The resulting convergence of the posterior estimate for the central tendency of extreme rainfall across these countries is approximately 0.574 inches, shown in Figure 1.

As emphasized in Kass and Wasserman (1996) and Gelman and Hill (2006), as the sample size increases, the influence of the prior tends to diminish, leading to the posterior distribution being predominantly driven by the data. This behavior has been formalized under the Bernstein-von Mises theorem for standard Bayesian models. However, the convergence properties and robustness in the context of Bayesian EV models, especially in hierarchical settings, are less explored. The following theorem sheds light on the robustness of hierarchical Bayesian EV models in the presence of large datasets.

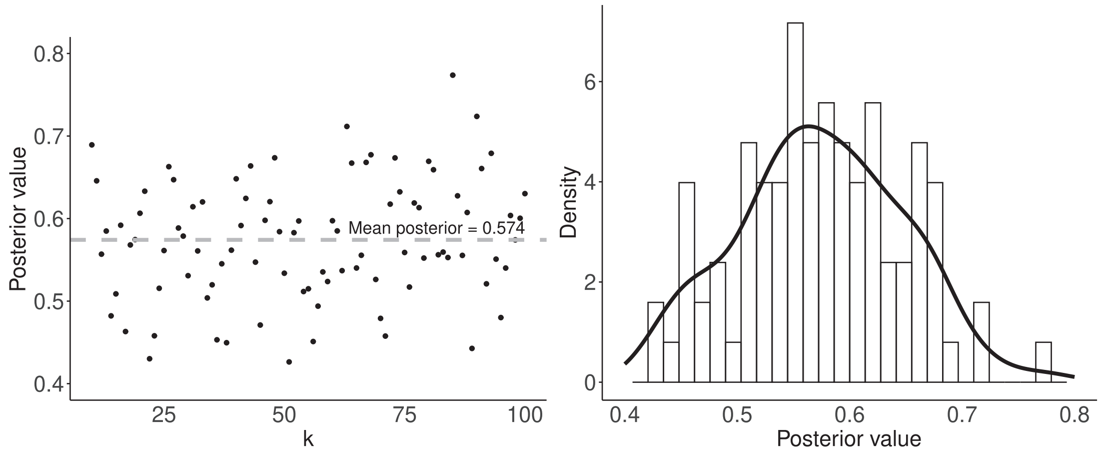

Fig. 1. The posterior convergence across k in a simulated Bayesian extreme value model.

Theorem 1 (Robustness of hierarchical Bayesian EV models under increasing group sizes)**.** *Consider a hierarchical Bayesian EV model with k groups. Let θ_i represent the group-specific parameters for the i-th group, influenced by the hyperparameter ϕ, and let $\mathbf{y}_i$ denote the observed extreme values for group i. Define $p(\theta_i|\mathbf{y}_i, \phi)$ as the posterior distribution of θ_i conditional on $\mathbf{y}_i$ and ϕ. We assume that all regularity conditions ensuring the existence and convergence of posterior distributions are satisfied and that the likelihoods, priors, and their derivatives are appropriately bounded and regular. The hyperposterior for ϕ is given by*

$$p(\phi|\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k) \propto \prod_{i=1}^{k} p(\mathbf{y}_i|\theta_i, \phi)\pi_{\theta_i|\phi}(\theta_i)\pi_\phi(\phi), \tag{10}$$

where $\pi_{\theta_i|\phi}(\theta_i)$ and $\pi_\phi(\phi)$ are the prior distributions for θ_i and ϕ, respectively.

If, as the number of groups k goes to infinity, the hyperposterior $p(\phi|\mathbf{y})$ converges in probability to a unique limiting distribution $\mathcal{D}(\phi)$, then, for any two distinct priors π_A and π_B, the Kullback-Leibler divergence between the resulting posteriors becomes negligible. Formally,

$$\lim_{k\to\infty} \mathcal{K}\Big(p(\theta|\mathbf{y}, \pi_A, \phi); p(\theta|\mathbf{y}, \pi_B, \phi)\Big) = 0 \quad \text{almost surely.} \tag{11}$$

Proof.

1. *Negligible influence of prior:* Assume that there exist regularity conditions, ensuring that the likelihood, the prior, and their derivatives are bounded and regular. Under these conditions, by the Bernstein-von Mises theorem, we have

$$p(\theta|\mathbf{y}, \pi_n) \xrightarrow{P} p(\theta|\mathbf{y}) \text{ as } n \to \infty,$$

   This indicates that the influence of the prior π on the posterior becomes negligible as the sample size n grows large.
2. *Convergence of hyperposterior:* In the hierarchical Bayesian model, where the hyperposterior is represented as in (10), assume sufficient regularity and boundedness of likelihood and priors. Under these assumptions, we establish

$$p(\phi|\mathbf{y}) \xrightarrow{P} \mathcal{D}(\phi) \text{ as } k \to \infty.$$

   This asserts the dominance of the hyperprior and the diminishing influence of individual group-specific priors on the hyperposterior as the number of groups k grows.
3. *KL divergence between different priors:* Under the established convergence conditions, we deduce that the KL divergence between posteriors corresponding to different priors converges to zero almost surely as the number of groups increases:

$$\lim_{k\to\infty} \mathcal{K}\Big(p(\theta|\mathbf{y}, \pi_A, \phi); p(\theta|\mathbf{y}, \pi_B, \phi)\Big) = 0 \text{ almost surely.}$$

4. *Stability and robustness of the model:* The convergence of the hyperposterior to $\mathcal{D}(\phi)$ in probability and the almost sure convergence of the KL divergence between different priors affirm the robustness and stability of the hierarchical Bayesian EV model, emphasizing the influence of the hyperprior over the selection of group-specific priors. □

Theorem 1 asserts that the hierarchical Bayesian EV model becomes robust to the choice of group-specific priors as the number of groups increases, emphasizing the influence of the hyperprior on determining the asymptotic behavior of the model.

Corollary 1 (Invariance of predictive distributions under large samples). *Assuming the hyperposterior distribution for the hyperparameters in Bayesian EV models converges to a unique distribution, the predictive distributions become asymptotically equivalent as the sample size increases.*

This result extends from the robustness principle detailed in Theorem 1. It posits that when we have two different sets of priors in Bayesian EV models, the predictions they lead to will eventually resemble each other closely as more data are added. The more data we have, the less influence our initial assumptions or priors have on the model's predictions. This characteristic is especially beneficial, providing model designers with flexibility: they can start with various initial beliefs, yet with enough data, they can be confident that these models will yield almost identical predictions. This not only underlines the robust nature of Bayesian EV models but also prioritizes empirical, data-driven insights over initial assumptions or beliefs, emphasizing the impact of data in determining predictive outcomes.

3.2. Consistency and convergence

In the context of hierarchical Bayesian models applied to extreme value theory, this section discusses the asymptotic behavior of the hyperposterior distribution. Leveraging the findings of Ghosal et al. (2000), this section establishes the hyperposterior's consistency and convergence rate under specific regularity conditions. The insights from Ghosh and Ramamoorthi (2003) on Bayesian nonparametric inference are particularly pertinent to the discussion on generalized hierarchical models. Shen and Wasserman (2001) further provides insights into convergence rates in specialized settings. For a broader perspective on Bayesian consistency, Walker and Hjort (2001) offers valuable guidance on achieving consistency under more generalized conditions.

Lemma 2 (Consistency of hyperposterior at true hyperparameter ϕ^*). *Assume ϕ^* is the true hyperparameter value from which the data are generated. Under the hierarchical Bayesian framework, let k denote the number of groups, and let the following regularity conditions hold:*

1. *The conditional likelihood $p(\theta|\phi)$ for the parameters given the hyperparameter is bounded and exhibits continuity in ϕ, ensuring stable model behavior as ϕ varies.*
2. *The hyperprior $\pi(\phi)$ is set to be continuous and assigns positive probability to neighborhoods around ϕ^*, ensuring that ϕ^* is a support point of the hyperprior.*

Then, the hyperposterior distribution $p(\phi|\mathbf{y})$ is consistent at the true hyperparameter ϕ^*, satisfying

$$\lim_{k\to\infty} \mathbb{P}\Big(|\phi - \phi^*| > \epsilon \,|\, \mathbf{y}\Big) = 0, \quad \forall \epsilon > 0, \tag{12}$$

indicating that as the number of groups increases, the hyperposterior distribution increasingly concentrates around the true hyperparameter value ϕ^*.

Proof. Consider the set $\Delta_\epsilon = \{\phi : |\phi - \phi^*| > \epsilon\}$ where the deviation from the true hyperparameter exceeds ϵ. For $\epsilon > 0$, define the measure $M_k(\epsilon)$ as

$$M_k(\epsilon) = \int_{\Delta_\epsilon} \exp\Big(k\Big|\log p(\mathbf{y}|\phi) - \log p(\mathbf{y}|\phi^*)\Big|\Big)\pi(\phi)\, d\phi,$$

which captures the prior mass allocated to Δ_ϵ. To establish consistency, we show that

1. $M_k(\epsilon)$ remains finite for all k, ensured by the boundedness of the conditional likelihood and the continuity of the hyperprior.
2. $\lim_{k\to\infty} \frac{1}{k}\log M_k(\epsilon) = -\infty$indicating exponential decay of the hyperposterior mass outside Δ_ϵ.

Applying Markov's inequality yields

$$\mathbb{P}\Big(\phi \in \Delta_\epsilon \,|\, \mathbf{y}\Big) \le \frac{M_k(\epsilon)}{\exp(k\Big|\log p(\mathbf{y}|\phi^*)\Big|)},$$

which approaches zero as $k \to \infty$, affirming the hyperposterior's concentration around ϕ^*. This is grounded in the principles established by Ghosal et al. (2000), demonstrates the hyperposterior's consistency at the true hyperparameter ϕ^*, underpinning the data generation process and thereby directly linking the choice of ϕ^* to the assumptions on data generation. □

Furthermore, using results from the Bayesian nonparametric literature, particularly Ghosal et al. (2000), the rate of convergence for the posterior can be determined by (1) demonstrating the Hellinger consistency of the posterior, (2) using the Laplace approximation to approximate the posterior around the maximum a posteriori (MAP) estimate, and (3) connecting the asymptotic behavior of the MAP with the MLE to infer the rate for the posterior. Given the specific regularity conditions, Σ can be related to the Fisher information or its Bayesian counterpart.

Lemma 3 (Rate of convergence). *Let ϕ^* be the true value of the hyperparameter and $\hat{\phi}$ be its posterior mean. Under certain regularity conditions on the hyperprior, likelihood, and assuming ϕ^* is an interior point of its parameter space, we have*

$$\sqrt{n}(\hat{\phi} - \phi^*) \xrightarrow{d} \mathcal{N}(0, \Sigma), \tag{13}$$

where Σ is a positive definite matrix. The rate of convergence of the hyperposterior to the true hyperparameter value is $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

Example 3. Consider multiple regions indexed by $i = 1, \ldots, K$. We collect data on the top incomes y_{ij} in each region, where $j = 1, \ldots, n_i$. We model these top incomes using the Pareto distribution to capture the heavy-tail nature of income distribution:

$$y_{ij} \sim \text{Pareto}(x_m, \alpha_i),$$

where x_m is the minimum income threshold to be considered in the top incomes category, and α_i is the shape parameter for region i, indicating the tail heaviness and inequality within that region's income distribution. To incorporate the variability and potential socio-economic factors across regions, we assume the shape parameters α_i are influenced by an overarching socio-economic factor, modeled as $\alpha_i \sim \text{Gamma}(\phi, \tau)$, where ϕ represents the mean effect of the socio-economic factor across all regions. Given a conjugate gamma prior for ϕ with parameters a and b and incorporating data from an increasing number of regions K, we aim to explore the behavior of the hyperposterior distribution of ϕ. The boundedness of the likelihood ratios, due to the properties of the Pareto distribution, and the continuity of the prior on ϕ ensure the regularity conditions are met. For any given region i, the posterior for α_i, given the observed top incomes and the hyperparameter ϕ, is

$$p(\alpha_i|\{y_{ij}\}, \phi) \propto \prod_{j=1}^{n_i} p(y_{ij}|\alpha_i) p(\alpha_i|\phi).$$

By integrating over all α_i's, we derive the marginal likelihood for ϕ:

$$p(\{y_{ij}\}_j|\phi) = \int \left(\prod_{j=1}^{n_i} p(y_{ij}|\alpha_i)\right) p(\alpha_i|\phi)\, d\alpha_i.$$

This process leads to the derivation of the hyperposterior for ϕ:

$$p(\phi|\{y_{ij}\}_{i,j}) \propto p(\phi) \prod_{i=1}^{K} p(\{y_{ij}\}_j|\phi).$$

Employing this hierarchical model, the hyperposterior distribution of ϕ converges to the true ϕ^* as the number of regions, K, increases, with a convergence rate of $\mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$.

3.3. Predictive performance

Hierarchical models are known for their adaptability in scenarios characterized by significant heterogeneity across groups or limited sample sizes within individual groups. These models leverage the shared structure among groups to refine predictive accuracy, as documented in the Bayesian literature. See Gelman et al. (2013) for examples. This ability to "borrow strength" from across groups renders hierarchical models valuable in extreme value analysis applications. While the hierarchical Bayesian approach directly addresses the challenges of heterogeneity and data sparsity, the broader statistical landscape also includes innovative methodologies that capture information properties within models. For instance, Ardakani et al. (2021) study the information properties of mixtures, further elaborating on how various statistical models utilize information. Such insights enrich the overarching framework within which hierarchical Bayesian models operate, especially in their capacity to process information for predictive purposes. Building on model information properties, the predictive performance of hierarchical Bayesian EV models is further scrutinized as follows.

Proposition 3 (Predictive performance of hierarchical Bayesian EV model)**.** *Let $p^H(y^*|y, \pi)$ be the predictive distribution from a hierarchical Bayesian EV model, $p^{NH}(y^*|y, \pi)$ be the predictive distribution from a non-hierarchical Bayesian EV model, and $L(y^*, \hat{y}^*)$ be a loss function, where y^* is the true future observation and $\hat{y}^*$ is the predicted future observation. Given heterogeneity across groups and small sample sizes within groups, we have*

$$\mathbb{E}\left[L\left(y^*, \hat{y}_H^*\right)\right] \leq \mathbb{E}\left[L\left(y^*, \hat{y}_{NH}^*\right)\right], \tag{14}$$

where $\hat{y}_H^*$ and $\hat{y}_{NH}^*$ represent the predicted future observations obtained from the hierarchical and non-hierarchical models, respectively.

Proof. In a hierarchical model, the posterior distribution of a group-specific parameter is informed by leveraging information from other groups, especially when data for the specific group is sparse (Gelman et al., 2013). This phenomenon, often referred to as "borrowing strength," can lead to more robust posterior distributions, especially with sparse group-specific data.

$$p^H(y^*|y, \pi) = \int\int p(y^*|\theta_i, \phi) p(\theta_i|\phi, y, \pi) p(\phi|y, \pi)\, \mathrm{d}\theta_i\, \mathrm{d}\phi.$$

Table 1
Predictive performance of hierarchical and non-hierarchical Bayesian EV.

	Median Posterior Predictive	Mean Posterior Predictive	Mean Squared Error
Hierarchical Bayesian EV	3.6489	3.6494	1.6595
Non-hierarchical Bayesian EV	3.6445	3.6480	1.6632

$$p^{NH}(y^*|y,\pi) = \int p(y^*|\theta_i)p(\theta_i|y,\pi)\,\mathrm{d}\theta_i.$$

Due to the borrowing of strength, the hierarchical model's predictive distribution is generally more robust with sparse data. This robustness, in turn, translates to a lower expected loss,

$$\mathbb{E}\big[L\big(y^*,\hat{y}^*_H\big)\big] \le \mathbb{E}\big[L\big(y^*,\hat{y}^*_{NH}\big)\big].$$

□

Remark 1. The equality

$$\mathbb{E}\big[L\big(y^*,\hat{y}^*_H\big)\big] = \mathbb{E}\big[L\big(y^*,\hat{y}^*_{NH}\big)\big]$$

holds under the following scenarios:

1. When each group has sufficient data such that the hierarchical borrowing of strength becomes unnecessary or redundant.
2. Even with sparse data, if there is minimal variability between groups, meaning all groups are essentially homogenous.
3. If the hierarchical model is incorrectly specified, including the specification of the hyperprior.
4. If the loss function does not penalize the types of mistakes that the non-hierarchical model is prone to.

Each scenario presents conditions under which the hierarchical structure does not offer an advantage over the non-hierarchical model in terms of expected loss.

Proposition 3 discusses the predictive advantages of hierarchical Bayesian EV models over their non-hierarchical counterparts. Hierarchical Bayesian models inherently share strength across groups. In situations where small individual group sample sizes can lead to uncertainty about those groups' parameters. However, hierarchical models can borrow information from the entire dataset to make more informed predictions for any specific group. This pooling of information can lead to a reduced predictive error compared to non-hierarchical models, especially when substantial variation exists between groups (Gelman and Hill, 2006).

On the other hand, non-hierarchical models make predictions for each group based purely on that group's data. For groups with limited data, these predictions can be uncertain or exhibit extreme values due to outlier observations. Hierarchical models alleviate this issue through their hyperparameters. These hyperparameters, which govern the distribution from which group-level parameters are drawn, are estimated based on data from all groups. This structure introduces a shrinkage effect in the estimates towards a grand mean, which can be highly beneficial when individual group sample sizes are small. However, a few caveats are worth noting. First, the benefits of hierarchical models are most pronounced when the data has an inherent hierarchical structure. This is observed when groups are not entirely independent and share common underlying influences. Second, hierarchical models tend to be more computationally demanding than non-hierarchical models. In extreme value analysis, the benefits of borrowing strength are even more pronounced given the rarity (and thus often sparse data) of extreme events. With the hierarchical model's ability to better inform individual group posteriors through shared hyperparameters, especially with sparse data, its predictive distribution offers better predictive performance (regarding a specific loss function) than a non-hierarchical model.

Example 4. To compare the predictive performance of a hierarchical Bayesian EV model with a non-hierarchical Bayesian EV model under conditions of heterogeneity and small sample sizes within groups, consider the following data generating process. First, data from the GEV distribution is simulated and used in extreme value analysis. Second, data are generated from three distinct groups characterized by different location parameters. The GEV distribution is parameterized in Eq. (1). The parameters used in this example are as follows.

- Number of observations per group: $n = 50$
- Location parameters: $\mu = \{1, 3, 5\}$
- Scale parameter: $\sigma = 1$

Using the generated data, two separate Bayesian models are fitted: (1) hierarchical Bayesian EV model (hierarchical model with group-specific intercepts) and (2) non-hierarchical Bayesian EV model (non-hierarchical model with group as a fixed effect). Posterior predictive samples are drawn for each group, and the point-wise median are computed for predictions. The Mean Squared Error (MSE) is then calculated for each model using these predictions.

Table 1 illustrates the predictive performance of both hierarchical and non-hierarchical models. The hierarchical model has a smaller MSE than the non-hierarchical model. Even though the difference appears modest, it is consistent with

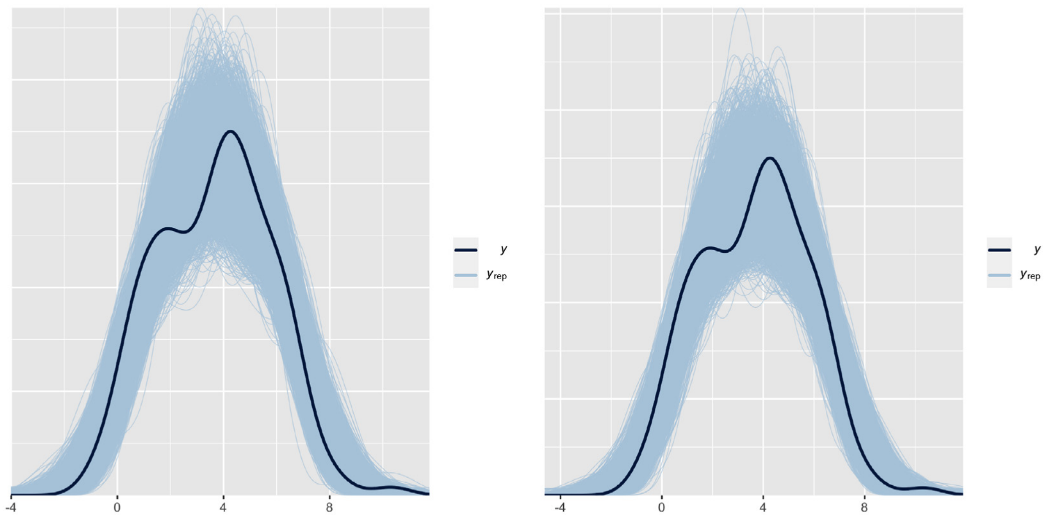

Fig. 2. Posterior predictive of hierarchical (left) and non-hierarchical (right) Bayesian EV.

Proposition 3 about the superior predictive capability of hierarchical models. The hierarchical model provides a more informed posterior distribution by pooling information from all groups.

Figure 2 presents the posterior predictive check plots of hierarchical and non-hierarchical models. The plot compares the observed data y with simulated data y_{rep} and visualizes if the model's simulations are consistent with the observed data.

Lemma 4 (Loss function decomposition)**.** *Given a predictive distribution $p(y^*|y,\pi)$ and a true distribution $q(y^*)$, the expected predictive loss can be decomposed as*

$$\mathbb{E}_q\left[L\left(y^*, p(y^*|y,\pi)\right)\right] = \left(\mathbb{E}_p[y^*] - \mathbb{E}_q[y^*]\right)^2 + \mathbb{E}_p\left[(y^* - \mathbb{E}_p[y^*])^2\right] + \mathbb{E}_q\left[(y^* - \mathbb{E}_q[y^*])^2\right], \tag{15}$$

where the first term denotes the squared bias, the second represents variance, and the third captures noise (James et al., 2013). The hierarchical Bayesian EV model notably reduces the variance component compared to a non-hierarchical model.

Proof. Define the squared error loss as

$$L(y^*, \hat{y}^*) = (y^* - \hat{y}^*)^2,$$

where $\hat{y}^* = \mathbb{E}_p[y^*|y,\pi]$. Expanding it and taking its expected value under $q(y^*)$, we get

$$\mathbb{E}_q\left[L\left(y^*, \hat{y}^*\right)\right] = \mathbb{E}_q\left[y^{*2}\right] - 2\mathbb{E}_q[y^*]\mathbb{E}_p[y^*|y,\pi] + \mathbb{E}_p[y^*|y,\pi]^2.$$

Using the properties of expectation, we can express the three terms of Lemma 4 in terms of the above expanded expected squared error loss.

The bias term is given by

$$\left(\mathbb{E}_p[y^*] - \mathbb{E}_q[y^*]\right)^2 = \left(\mathbb{E}_p[y^*|y,\pi] - \mathbb{E}_q[y^*]\right)^2.$$

The variance term for p is

$$\mathbb{E}_p\left[(y^* - \mathbb{E}_p[y^*])^2\right] = \mathbb{E}_p\left[y^{*2}\right] - \mathbb{E}_p[y^*|y,\pi]^2.$$

And the noise term for q is

$$\mathbb{E}_q\left[(y^* - \mathbb{E}_q[y^*])^2\right] = \mathbb{E}_q\left[y^{*2}\right] - \mathbb{E}_q[y^*]^2.$$

Combining these terms, we get the result in Lemma 4. The regularization effect stems from sharing information across groups. This pooling of information results in group-specific estimates being shrunk towards the overall mean, reducing variance, especially in extreme events. Detailed derivations are provided in Gelman and Hill (2006). □

Remark 2 (Shrinkage in hierarchical models)**.** In hierarchical modeling, *shrinkage* describes the regularization effect wherein group-specific estimates gravitate towards a shared central value. This is particularly notable when data for individual groups is sparse. As a result, the variance of these estimates is reduced, leading to more consistent and robust predictions across the model. This behavior underscores one of the major advantages of hierarchical structures, as they leverage the “borrowing of strength” from the entire dataset to make informed estimations even in the face of limited group-specific data. This variance reduction aligns with the decomposition highlighted in Lemma 4 and examines the benefits of adopting hierarchical over non-hierarchical approaches in extreme events.

Table 2
Bias, variance, and noise estimates for extreme rainfall.

	Bias	Variance	Noise
Region A	0.47	8.05	8.00
Region B	0.06	9.43	9.44

Example 5. Consider a scenario where we analyze extreme rainfall events across various regions. Each region is treated as a group in a hierarchical model. Extreme value theory informs us that rainfall exceeding certain thresholds can often be modeled with GEV distributions (Katz et al., 2002). For simplicity, assume two regions: Region A, with only a decade of rainfall data, and Region B, with five decades of data. In traditional non-hierarchical models, predicting extreme rainfall events for Region A might be challenging due to its limited data. However, by employing a hierarchical Bayesian EV model, information from Region B supplements the analysis for Region A, effectively reducing the variance in its predictions. The following data are simulated given both regions' extreme rainfall is modeled with a GEV distribution defined in Equation (1):

- Region A: $\mu = 5, \sigma = 1.5, \xi = 0.2$.
- Region B: $\mu = 5.2, \sigma = 1.6, \xi = 0.25$.

Given a sample size of $n = 100$ for Region A and $5n$ for Region B, we can see how close the hierarchical model's predictions are to the true extreme rainfall values for Region A. Bias quantifies the deviation of the predictions, influenced by the shared information from the true values. Variance is expected to diminish for Region A due to its hierarchical structure. Noise represents the natural variability in rainfall events that aren't captured by the model, including sudden anomalies caused by unforeseen climatic changes.

Bias, variance, and noise estimates for these regions are summarized in Table 2. The following points are noteworthy. For Region A, the bias is comparatively larger at 0.47, indicating that the predictions for this region deviate more from the true values than those for Region B, where the bias is just 0.06. This elevated bias in Region A might be attributed to the limited data available. However, despite this challenge, the hierarchical nature of the model assists in reducing the variance in predictions for Region A to 8.05, which is slightly lower than that for Region B (9.43). This illustrates the power of the hierarchical Bayesian EV model in leveraging information across regions to produce more stable estimates, especially when the data for a particular group (in this case, Region A) is sparse. Noise for Regions A and B are 8.00 and 9.44, implying that there are unaccounted factors in the rainfall events for both regions.

3.4. Influence of hyperpriors

The choice of prior affects posterior inferences, especially for extreme events or when the data is relatively uninformative about certain parameters. The impact of prior specification has been discussed in the literature. Notably, Kass and Wasserman (1996) study the robustness and sensitivity of Bayesian inferences to different prior choices. Gelman (2006) also elaborates on the implications of prior choice. Furthermore, hyperpriors allow for a hierarchical specification of priors. Gelman et al. (2013) elaborate on this hierarchical structure, examining prior and hyperprior choices in complex models. The section explains the influence of hyperprior informativeness on predictive loss in Bayesian EV analysis.

Lemma 5 (Influence of hyperprior)**.** *Given a Bayesian EV model with differentiable likelihood and hyperprior functions, assume the following:*

1. *The true parameter value θ^* lies within the mode of the likelihood function.*
2. *The hyperprior has mean θ^*.*
3. *The data is sufficiently informative such that the likelihood dominates the prior for large sample sizes.*

Under these conditions, if the hyperprior makes the prior more informative (e.g., decreases its variance) and pulls it towards θ^*, then an increase in the informativeness of the hyperprior will not increase the predictive loss. That is, if $L(\theta; \phi)$ denotes the predictive loss for hyperprior parameter ϕ, then

$$\frac{\partial}{\partial \phi} L(\theta; \phi) \leq 0. \tag{16}$$

Proof. Given the likelihood and prior:

$$p(\theta|y, \phi) \propto \mathcal{L}(y|\theta)\pi(\theta|\phi).$$

From condition 3, for large sample sizes, the likelihood will dominate, and hence

$$p(\theta|y, \phi) \approx \mathcal{L}(y|\theta).$$

When increasing the informativeness of the hyperprior (by increasing ϕ), the variance of the prior $\pi(\theta|\phi)$ decreases, centering around θ^*. We define the predictive loss function as

$$L(\theta; \phi) = \mathbb{E}_{p(\theta|y,\phi)}[\theta] - \theta^*.$$

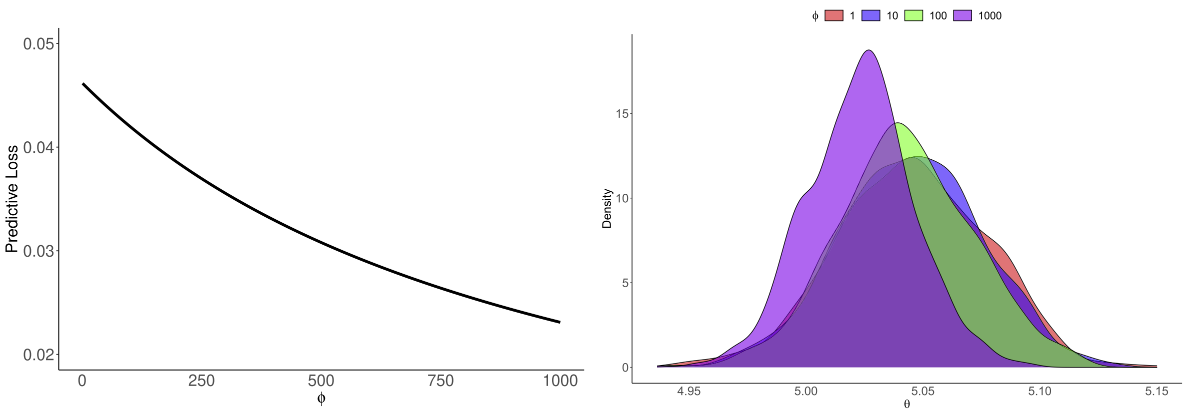

Fig. 3. Predictive loss vs. hyperprior (left) and posterior distributions across ϕ (right).

Differentiating with respect to ϕ,

$$\frac{\partial}{\partial \phi} L(\theta;\phi) = \frac{\partial}{\partial \phi} \mathbb{E}_{p(\theta|y,\phi)}[\theta].$$

Given conditions 1 and 2, and since the hyperprior makes the prior more informative and centered around θ^*, the expectation under the posterior, $\mathbb{E}_{p(\theta|y,\phi)}[\theta]$, moves closer to θ^* as ϕ increases. This implies that the change in expected predictive loss with respect to ϕ is non-positive:

$$\frac{\partial}{\partial \phi} L(\theta;\phi) \leq 0.$$

□

For GEV and GP distributions to be effectively utilized within a hierarchical Bayesian framework, certain regularity conditions must be satisfied. Both the GEV and GP distributions exhibit Lipschitz continuity in their likelihood functions with respect to their parameters. This property ensures the stability of the model as the parameters vary, particularly within the hierarchical structure where group-specific parameters θ_i are influenced by a common hyperparameter ϕ. The likelihood functions for GEV and GP distributions possess derivatives that are bounded within the parameter space. This boundedness is essential for the tractability of the model and ensures that the hyperposterior distribution's convergence can be effectively analyzed. The hierarchical model framework assumes a proper hyperprior, which regularizes the model and ensures the well-posedness of the posterior distribution. For GEV and GP distributions, the choice of hyperprior must reflect the underlying extreme value processes. These conditions underpin the hierarchical Bayesian approach to extreme value analysis, ensuring the model remains robust.

Example 6. Consider a Bayesian EV model with true parameter value $\theta^* = 5$. A hyperprior is chosen centered at θ^*, and the predictive loss as the hyperprior becomes more informative. Let the likelihood be defined as

$$\mathcal{L}(y|\theta) \sim \mathcal{N}(\theta, 1).$$

The hyperprior is normally distributed as

$$\pi(\theta|\phi) \sim \mathcal{N}(5, \frac{1}{\sqrt{\phi}}).$$

The hyperprior variance decreases as ϕ increases, making it more informative. The data are simulated from the likelihood and observe how the predictive loss changes with increasing ϕ.

The left panel in Figure 3 shows that the predictive loss declines as the hyperprior becomes more informative. The right panel plots the posterior distributions of the parameter θ under various hyperprior settings, as denoted by different values of ϕ. Each color represents a unique value of ϕ, and the densities depict how the beliefs about θ change with the informativeness of the hyperprior. As ϕ increases, making the hyperprior more informative, we observe a narrowing and shift in the posterior distribution, highlighting the influence of the hyperprior on our post-data beliefs about θ. The convergence of these distributions towards the true parameter value underscores the point in Lemma 5 about the non-increasing predictive loss with a more informative hyperprior.

4. Empirical illustration

This section examines the application of hierarchical Bayesian extreme value models to study the resilience of major financial entities to adverse market conditions and policy shifts. It focuses on stress-testing the returns of major banks, such

Table 3
Descriptive Statistics of Absolute Negative Returns.

Bank	n	Max	Mean	SD	25%	50%	75%	95%
JPM	2,061	20.7%	1.5%	1.8%	4.0%	9.0%	1.9%	4.9%
BAC	2,055	29.0%	1.8%	2.4%	0.5%	1.1%	2.2%	5.5%
C	2,097	39.0%	1.9%	2.6%	0.5%	1.1%	2.3%	5.6%

n corresponds to the total number of days with negative returns. Percentiles are expressed as percentages.

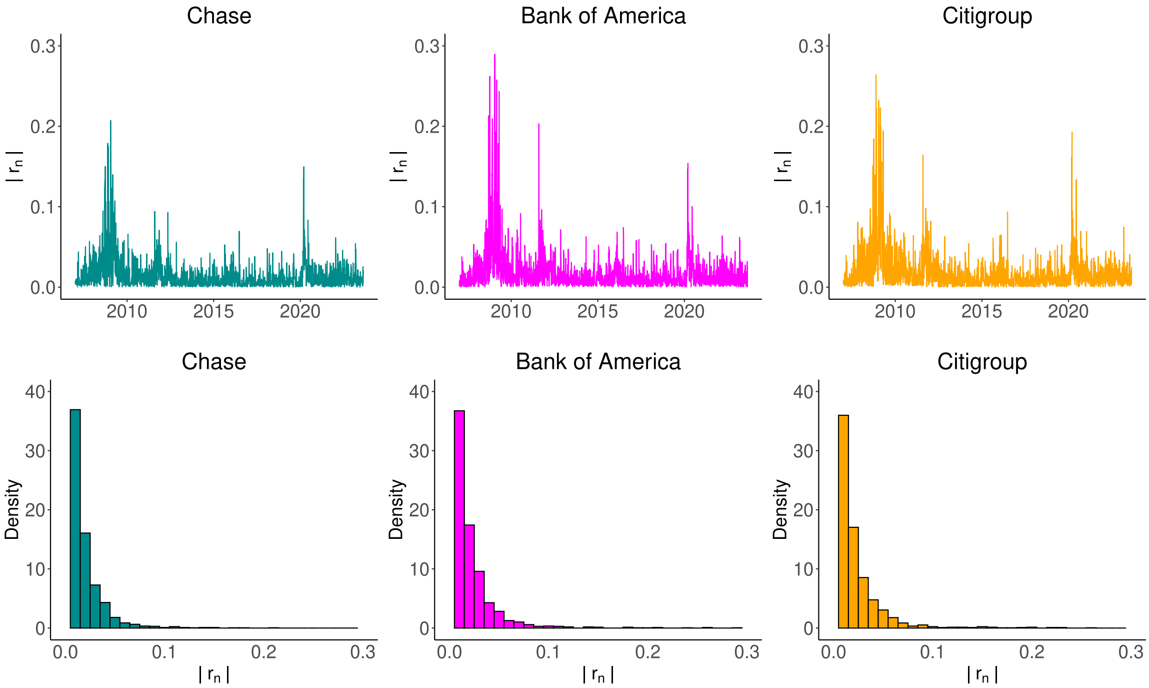

Fig. 4. Time series and histogram plots of absolute negative returns.

as JPMorgan Chase, Bank of America, and Citigroup, to evaluate their potential vulnerabilities to extreme market downturns. Moreover, it extends the analysis to simulate the impact of monetary policy shocks on these institutions, studying susceptibility to external policy changes.

4.1. Stress testing of financial institutions

This application performs a stress testing analysis on the returns of three banks: JPMorgan Chase (JPM), Bank of America (BAC), and Citigroup (C). These banks represent some of the largest financial institutions in the U.S. Stress testing, which assesses the potential vulnerability of financial institutions to extreme market shocks, has become a key component in risk management and regulatory frameworks (BLS, 2009). The hierarchical Bayesian approach considers the similarities and differences between banks while estimating their return distributions. Hierarchical models are beneficial when information is shared across groups (Gelman and Hill, 2006).

Banks' daily returns are obtained from Bloomberg. The data spans from January 3, 2007, to August 08, 2023. This empirical exercise focuses explicitly on negative returns to assess the banks' risk profiles. Table 3 provides descriptive statistics for the absolute negative returns. The number of observations, denoted as n, represents the total number of days with negative returns. Citigroup shows the highest maximum decline at 39%, with Bank of America and JPMorgan Chase following at 29% and 20.7%, respectively. The mean decline ranges from 1.5% for JPMorgan Chase to 1.9% for Citigroup, indicating the average magnitude of negative returns. For JPMorgan Chase, the 25th percentile at 4.0% and the 75th percentile at 1.9% indicate that the lower quartile of negative returns is more severe compared to the upper quartile. This pattern is reversed for Bank of America and Citigroup, where the 25th percentile values are lower (0.5%), and the 75th percentile values are higher (2.2% for BAC and 2.3% for C), suggesting a heavier tail on the upper side of their negative return distributions.

Figure 4 presents the time series and histogram plots of absolute negative returns. Bank of America has a slightly broader spread with a peak near 0.018 and extremities stretching to 0.290 than Chase, with extreme values reaching 0.207. Meanwhile, Citigroup exhibits the most significant tail risk, with the maximum negative return peaking at 0.390, suggesting

Table 4
Estimated GPD Parameters for absolute negative returns.

	ξ	σ	μ	β
Chase	0.218	0.011	-0.003	0.022
Bank of America	0.236	0.016	-0.015	0.032
Citigroup	0.403	0.008	0.005	0.029

ξ and β are scale and shape parameters of generalized Pareto.

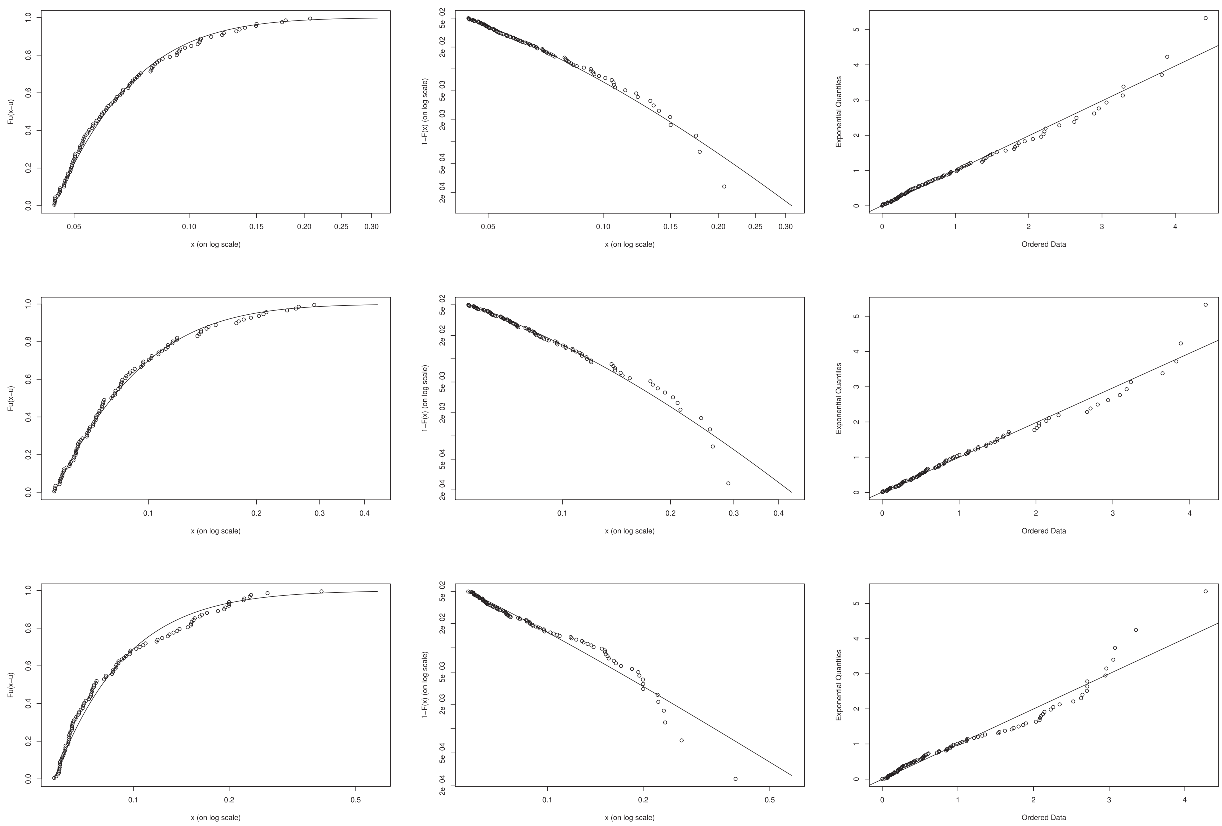

Fig. 5. Tail risk analysis of absolute negative returns using the Peaks Over Threshold method. Rows represent Chase, Bank of America, and Citigroup. The first column shows excess distribution. The second column presents tail of underlying distribution. The third column presents the QQ plots of Residuals.

sporadic severe downturns in its history. In extreme value analysis, distinguishing between distributions, particularly in the tails, is pivotal. Ardakani et al. (2020) introduce the MR plot as a big data tool for this purpose, offering an approach to determine differences between distributions. This tool can be instrumental in extreme value theory, aiding in selecting appropriate models for tail behavior analysis.

Table 4 provides the parameter estimates of the GP distribution. These estimates are derived from the Peak Over Threshold method, utilizing returns exceeding the 95th percentile. This method is used to focus on the most extreme negative returns. This approach allows for analyzing the tail behavior required for stress testing and understanding the potential for extreme losses. Utilizing the 95th percentile provides a balance between including sufficiently extreme events and maintaining a reasonable sample size. The scale (ξ) and shape (β) parameters provide insight into the severity and tail behavior of the return distributions. Notably, Citigroup has the highest shape parameter. It suggests the fattest tail among the three banks, indicating a higher probability of extreme negative returns, highlighting its vulnerability to severe market downturns. Conversely, while Bank of America and Chase have similar shape parameters, Bank of America's scale parameter is slightly higher, suggesting that its extreme negative returns are more severe than those of Chase.

Figure 5 presents the absolute negative returns of the three banks using the Peak Over Threshold method. Each row represents a bank, while the columns exhibit different aspects of the return distributions. The first column, excess distribution, provides how frequently returns exceed a particular threshold, indicating the intensity of extreme negative returns for each bank. The second column, the tail of the underlying distribution, gives a visual representation of the tail behavior of returns, offering insights into the potential for extreme losses. The QQ plot of residuals in the third column is a diagnostic tool to

Table 5
Posterior samples for hierarchical Bayesian model.

Parameter	Min	25%	Median	Mean	75%	Max
Intercept: All groups	-0.0084	0.0159	0.0174	0.0178	0.0191	0.0396
Intercept: Chase	-0.0256	-0.0044	-0.0026	-0.0031	-0.0013	0.0233
Intercept: Bank of America	-0.0220	-0.0007	0.0008	0.0004	0.0021	0.0271
Intercept: Citigroup	-0.0208	-0.0004	0.0012	0.0008	0.0026	0.0268
Residual SD	0.0225	0.0229	0.0230	0.0230	0.0231	0.0235

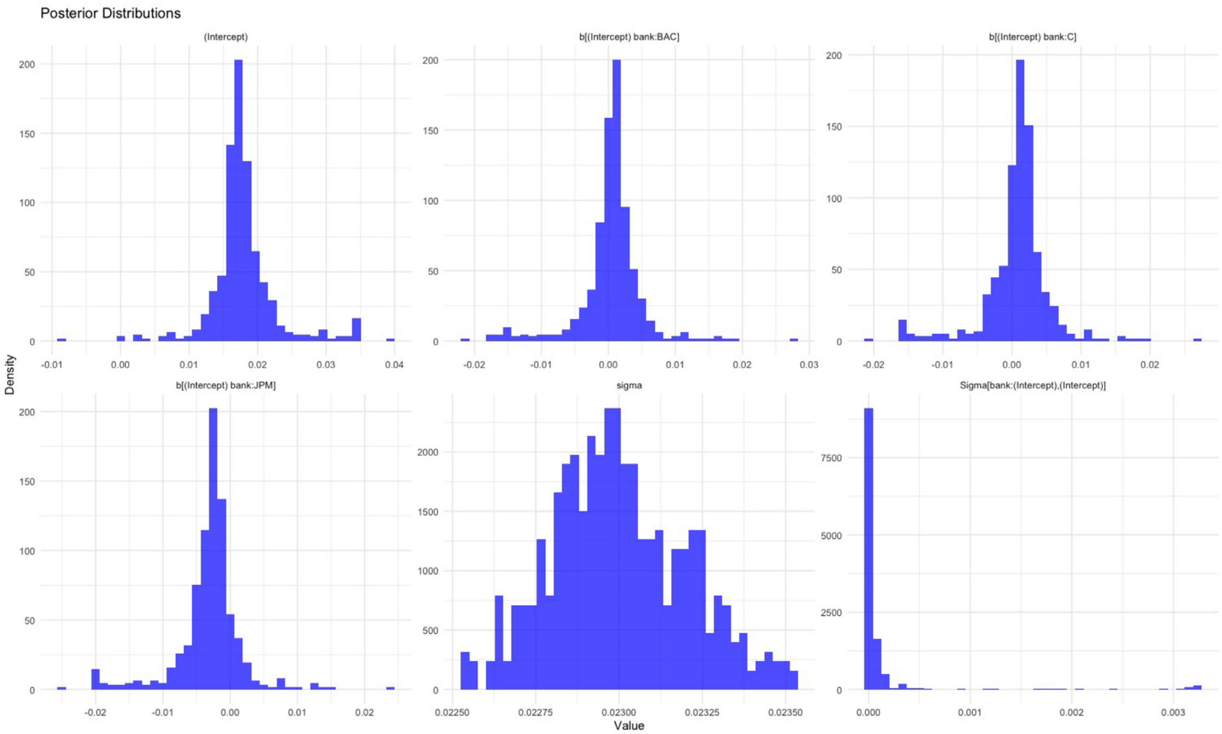

Fig. 6. The posterior distributions of parameters in the hierarchical Bayesian model.

check the fit of the generalized Pareto distribution to the observed excesses. Deviations from the diagonal line in this plot indicate departures from the assumed model. These plots offer a perspective into the tail risk characteristics of the banks' returns.

The hierarchical Bayesian EV model accounts for the shared structure between banks while permitting each bank to have its unique return profile. It also provides a framework for uncertainty quantification and parameter estimation. Table 5 represents the posterior samples from the hierarchical Bayesian model. This model considers differences across three banks, which is its primary advantage since it enables the sharing of strength across groups, enabling a more informed estimation even for groups with less data. The intercept across the three banks with no specific bank in consideration has a median of 0.0174, suggesting that the base return across all banks is positive. The specific bank effects are close to zero, indicating only minor deviations from the overall base return for each bank. However, their min-max range does show some variability, indicating the uncertainty in these estimations. Diagnostics ensure that the MCMC sampler has converged.

The hierarchical Bayesian EV model's ability to leverage shared information across banks while accommodating their individual risk profiles provides an understanding of financial risk. This approach enhances the robustness of the findings and allows for an informed estimation of risk, even for banks with less extreme data. Figure 6 presents the posterior distributions of each parameter from the hierarchical Bayesian model for bank returns. Each histogram shows the frequency of sampled values for its respective parameter. We can infer the central tendencies and uncertainties associated with each parameter from the distribution shapes. The posterior distributions are concentrated, indicating high certainty in parameter estimates. Figure 7 displays the mean predictions and associated 95% confidence intervals for the returns of each bank, as estimated by the hierarchical Bayesian model. The blue points represent the mean predictions, while the red error bars indicate the 95% confidence intervals. The width of the confidence intervals provides insight into the uncertainty associated with each prediction. Narrower intervals indicate more precise predictions, while wider intervals suggest greater uncertainty.

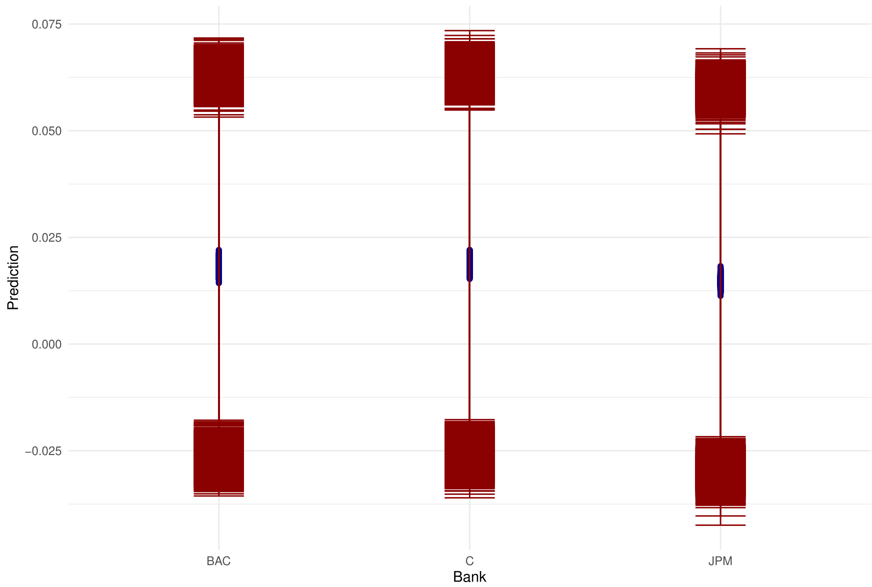

Fig. 7. Mean predictions and 95% confidence intervals for bank returns from the hierarchical Bayesian model.

The position of the mean point within the interval can also shed light on potential skews or asymmetries in the posterior predictive distribution.

Overall, findings provide several noteworthy observations.

1. While all three banks have their returns drop between 1.5% to 1.9% on average, Citigroup and Bank of America exhibit higher volatility in their negative returns. This variance can be indicative of the potential risks these institutions carry, especially in adverse market conditions.
2. Citigroup has faced the steepest declines, with the worst-case scenario seeing its returns drop by a staggering 39%, revealing its vulnerability.
3. The advantage of the hierarchical Bayesian modeling approach are twofold. First, it accounts for the unique profiles of each bank and, secondly, it leverages shared information across these banks to produce more informed estimations.

In this empirical example, we observe tail behaviors that closely mirror the theoretical foundations laid out in previous sections. Specifically, the distribution of extreme negative returns exemplifies the exceedance patterns predicted by the GP distribution. This observation is particularly pertinent in the likelihood function's Lipschitz continuity in parameters from Proposition 1, where the GP distribution's utility in modeling excesses over a predetermined threshold is underscored. The data's adherence to this exceedance pattern validates the continuity of the priors and reinforces the GP distribution's appropriateness in capturing the tail risk inherent in financial markets. Furthermore, the application of the block maxima approach involves aggregating returns data to isolate the most extreme events within specified periods. This approach is seamlessly modeled through the GEV distribution, designed to encapsulate block maxima behavior.

4.2. Modelling monetary policy shocks

The banking sector is influenced by monetary policy shocks—unexpected changes in the monetary policy tools. While these policies directly impact financial institutions, the extent to which these shocks influence individual banks can differ based on their operational dynamics, asset holdings, and market exposure (Bernanke and Gertler, 1995; Romer and Romer, 2004). For simplicity, the shock is modeled as a random percentage change between -5% and 5%. While this is a broad generalization, in a real-world scenario, such shocks would be based on factors such as central bank announcements, geopolitical events, etc. The purpose of simulating a monetary policy shock in the model is to analyze its impact on the returns of banks that are already vulnerable due to negative returns. By integrating this shock with the previous models, we can capture the financial risks associated with intrinsic (negative returns) and extrinsic (monetary policy) factors.

After incorporating the shock, the hierarchical Bayesian model is refitted to account for fixed effects (the shock) and random effects (bank-specific variations). The posterior distributions post-shock show the bank returns' distribution under such stressed scenarios. Figure 8 presents the posterior distributions after incorporating a monetary policy shock on the returns of the three major banks. Each histogram shows the distribution of estimated parameters after accounting for the shock. Parameters have more spread-out posterior distributions, suggesting greater variability in their post-shock values. The shape and spread of each distribution can be related to the sensitivity of each bank's returns to monetary policy changes

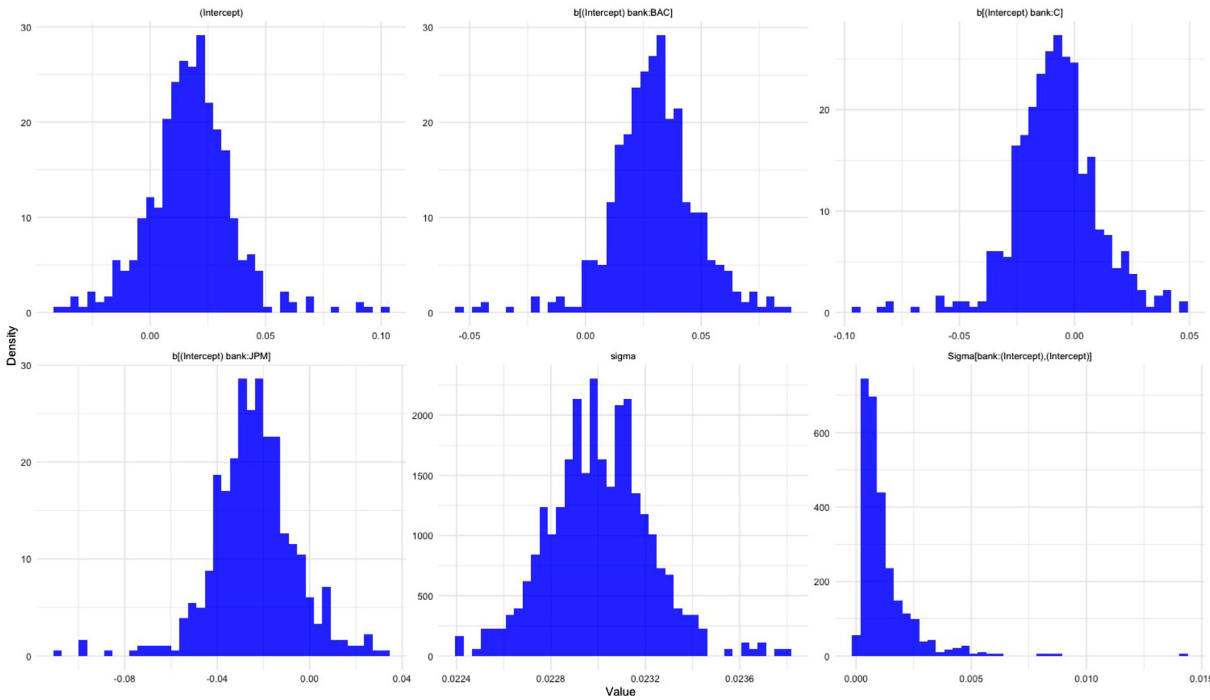

Fig. 8. Posterior distributions of model parameters post-monetary policy shock.

and their overall financial stability. Banks with more concentrated posterior distributions show less sensitivity to monetary policy shocks and are better hedged against such events, while banks with less concentrated distributions are more exposed to such shocks.

5. Concluding remarks

Bayesian extreme value models offer a framework to study tail behavior, evaluate risk, and make predictions in the presence of extreme events. By introducing Bayesian approaches, we can account for uncertainties in the parameter estimates. Furthermore, the hierarchical structure enables modeling heterogeneity across groups and borrowing strength from shared structures, making it suitable for datasets with shared global influences. This paper establishes the asymptotic behavior of these Bayesian extreme value models, examining the sensitivity of the models to the choice of priors, the influence of hyperpriors on model convergence, and the predictive robustness of hierarchical models under increasing group sizes.

The theoretical results illustrate that the Kullback-Leibler divergence between posteriors derived from different priors has an upper bound, underlining the stability of the Bayesian EV model with respect to prior choices. This idea is augmented by the Bernstein-von Mises theorem, suggesting that as the sample size grows, the influence of prior beliefs becomes almost negligible. A result pertains to the hierarchical Bayesian EV model when considering multiple groups. As the number of groups increases, the hyperposterior tends to a particular distribution. The hierarchical structure thus introduces another layer of robustness to the Bayesian EV models, especially when multiple groups are considered. The findings also suggest that the robustness of the Bayesian extreme value model is not merely constrained to prior beliefs but also extends to its hierarchical structure. As the number of groups increases, the effect of different priors on the posterior diminishes.

In addition, new findings focus on the predictive power of hierarchical Bayesian EV models. The predictive distributions become almost indistinguishable as the sample size grows. This implies that these models ensure reliability for h-step-ahead forecasting or big data analysis. The loss function can be decomposed into bias, variance, and noise. A result establishes the diminished variance component in a hierarchical Bayesian EV model. This underscores its superiority over its non-hierarchical counterpart, especially when high data heterogeneity exists. Lastly, a perspective on the influence of hyperpriors is presented. Under certain conditions, increasing the informativeness of the hyperprior doesn't increase the predictive loss. This insight can guide practitioners when choosing hyperpriors.

Future research can focus on variable selection techniques specific to Bayesian EV models. This can make the model more interpretable and improve its prediction performance. Also, future research can study advanced techniques for approximating the posterior distribution, such as variational inference or alternative MCMC methods, to improve computational efficiency. While Bayesian EV models are promising, their performance depends on the quality and relevance of the data. The accuracy and robustness of models rely on data preprocessing. Combining Bayesian EV models with other modeling approaches, such as machine learning techniques, could yield models with better prediction capabilities.

Declaration of competing interest

I, Omid M. Ardakani, as the sole author of this manuscript, hereby declare that I have no relevant or material financial interests related to the research described in this paper. This includes any financial relationships, direct or indirect, or other situations that could potentially be perceived as influencing the objectivity, integrity, and validity of this research.

CRediT authorship contribution statement

Omid M. Ardakani: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Acknowledgements

I extend my gratitude to the editor, the associate editor, and the two reviewers whose insightful comments and suggestions significantly enhanced the quality of this paper. Their detailed feedback was invaluable in refining my arguments and improving the overall clarity and rigor of the manuscript. Any errors that remain are solely my responsibility.

References

Ardakani, O., Asadi, M., Ebrahimi, N., Soofi, E., 2021. Variants of mixtures: Information properties and applications. Journal of the Iranian Statistical Society 20 (1), 27–59. doi:10.52547/jirss.20.1.27.
Ardakani, O.M., 2022. Option pricing with maximum entropy densities: The inclusion of higher-order moments. Journal of Futures Markets 1–16. doi:10.1002/fut.22361.
Ardakani, O.M., 2023. Capturing information in extreme events. Economics Letters 231, 111301.
Ardakani, O.M., Asadi, M., Ebrahimi, N., Soofi, E.S., 2020. MR plot: A big data tool for distinguishing distributions. Statistical Analysis and Data Mining: The ASA Data Science Journal 13 (4), 405–418. doi:10.1002/sam.11464.
Ardakani, O.M., Ebrahimi, N., Soofi, E.S., 2018. Ranking forecasts by stochastic error distance, information and reliability measures. International Statistical Review 86 (3), 442–468.
Ardakani, O.M., Saenz, M., 2022. On the comparison of inequality measures: evidence from the world values survey. Applied Economics Letters 1–10.
Asadi, M., Ashrafi, S., Ebrahimi, N., Soofi, E.S., 2010. Models based on partial information about survival and hazard gradient. Probability in the Engineering and Informational Sciences 24 (4), 561–584.
Balkema, A.A., Haan, L.d., 1974. Residual life time at great age. The Annals of Probability 2 (5), 792–804.
Bernanke, B., Gertler, M., 1995. Inside the black box: The credit channel of monetary policy transmission. Journal of Economic perspectives 9 (4), 27–48.
Bernardo, J.M., 1979. Reference posterior distributions for Bayesian inference. Journal of the Royal Statistical Society. Series B (Methodological) 41 (2), 113–147.
BLS, 2009. Basel committee on banking supervision: Principles for sound stress testing practices and supervision. Technical Report. Bank for International Settlements.
Coles, S., Bawa, J., Trenner, L., Dorazio, P., 2001. An introduction to statistical modeling of extreme values. Springer, London.
Coles, S.G., Powell, E.A., 1996. A review and new developments. International Statistical Review 64 (1), 119–136.
Coles, S.G., Tawn, J.A., 1991. Modelling extreme multivariate events. Journal of the Royal Statistical Society Series B: Statistical Methodology 52 (2), 377–392.
Cover, T.M., Thomas, J.A., 1991. Information theory and statistics: Elements of information theory. Wiley.
Davison, A.C., Padoan, S.A., Ribatet, M., 2012. Statistical modeling of spatial extremes. Statistical Science 27 (2), 161–186.
Embrechts, P., Kluppelberg, C., Mikosch, T., 2013. Modelling extremal events: for insurance and finance.
Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Analysis 1 (3), 515–534.
Gelman, A., Carlin, J.B., Stern, H.S., DusonA, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian data analysis.
Gelman, A., Hill, J., 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge university press.
Ghosal, S., Ghosh, J.K., van der, V.A.W., 2000. Convergence rates of posterior distributions. Annals of Statistics 28 (2), 500–531.
Ghosh, J.K., Delampady, M., Samanta, T., 2006. An introduction to Bayesian analysis: theory and methods. Springer, New York.
Ghosh, J.K., Ramamoorthi, R.V., 2003. Bayesian Nonparametrics. Springer Series in Statistics.
Hoff, P.D., 2009. A first course in Bayesian statistical methods. Springer, New York.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning. Springer, New York.
Jaynes, E.T., 2003. Probability theory: The logic of science. Cambridge university press.
Jeffreys, H., 1946. An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 186 (1007), 453–461.
Kass, R.E., Wasserman, L., 1996. The selection of prior distributions by formal rules. Journal of the American Statistical Association 91 (435), 1343–1370.
Katz, R.W., Parlange, M.B., Naveau, P., 2002. Statistics of extremes in hydrology. Advances in Water Resources 25, 1287–1304.
Kullback, S., 1959. Information theory and statistics. John Riley and Sons, New York.
Kullback, S., Leibler, R., 1951. On information and sufficiency. Annals of Mathematical Statistics 22, 7986.
Lavine, R.A., Herring, A.H., Dunson, D.B., 2021. Roy, A., Lavine, I., Herring, A. H., & Dunson, D. B. (2021). Perturbed factor analysis: Accounting for group differences in exposure profiles. The Annals of Applied Statistics 15 (3), 1386.
Leadbetter, M.R., Lindgren, G., Rootzén, H., 2012. Extremes and related properties of random sequences and processes. Springer Science & Business Media.
Mao, G., Zhang, Z., 2018. Stochastic tail index model for high frequency financial data with Bayesian analysis. Journal of Econometrics 205 (2), 470–487.
Novak, S.Y., 2011. Extreme value methods with applications to finance. CRC Press.
Pickands, J., 1975. Statistical inference using extreme order statistics. Annals of Statistics 119–131.
Pons, F.M.E., Messori, G., Alvarez-Castro, M.C., Faranda, D., 2020. Sampling hyperspheres via extreme value theory: implications for measuring attractor dimensions. Journal of Statistical Physics 179 (5-6), 1698–1717.
Robert, C.P., 2007. The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer, New York.
Romer, C.D., Romer, D.H., 2004. A new measure of monetary shocks: Derivation and implications. American Economic Review 94 (4), 1055–1084.
Scarrott, C., MacDonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. REVSTAT - Statistical Journal 10 (1), 22–60.

Shen, X., Wasserman, L., 2001. Rates of convergence of posterior distributions. The Annals of Statistics 29 (3), 687–714.
Silverberg, G., Verspagen, B., 2007. The size distribution of innovations revisited: an application of extreme value statistics to citation and value measures of patent significance. Journal of Econometrics 139 (2), 318–339.
Smith, R.L., 1987. Estimating tails of probability distributions. The Annals of Statistics 15 (3), 1174–1207.
Smith, R.L., 1990. Extreme value theory. Handbook of Applicable Mathematics 7 (437-471), 18.
Soofi, E.S., Retzer, J.J., 2002. Information indices: unification and applications. Journal of Econometrics 107, 1740.
Van der Vaart, A.W., 2000. Asymptotic statistics. Cambridge University Press.
Walker, S., Hjort, N.L., 2001. On Bayesian consistency. Journal of the Royal Statistical Society Series B: Statistical Methodology 63 (4), 811–821.