Statistica Sinica Preprint No: SS-2026-0109	
Title	Inference for Extreme Value Index Through Individualized Fusion Learning
Manuscript ID	SS-2026-0109
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202026.0109
Complete List of Authors	Hongfang Sun, Yu Chen, Tao Xu and Zhengjun Zhang
Corresponding Authors	Zhengjun Zhang
E-mails	zjz@stat.wisc.edu
Notice: Accepted author version.	

INFERENCE FOR EXTREME VALUE INDEX THROUGH INDIVIDUALIZED FUSION LEARNING

Hongfang Sun[1], Yu Chen[2], Tao Xu[3], and Zhengjun Zhang[4,5,6]

[1] *School of Mathematical Sciences, Ministry of Education Key Laboratory of NSLSCS, Nanjing Normal University*

[2]*School of Public Affairs, University of Science and Technology of China*

[3]*College of Computer and Information Sciences, Fujian Agriculture and Forestry University*

[4] *School of Data Science and Artificial Intelligence, Beijing University of Chinese Medicine*

[5]*School of Economics and Management, University of Chinese Academy of Sciences*

[6]*Department of Statistics, University of Wisconsin*

Abstract: Accurate estimation of the extreme value index (EVI) is central to the analysis of tail risks. Although substantial theoretical progress has been made, most existing estimators rely solely on tail observations from the variable of interest, which are often too limited or unstable to yield reliable inferences. We propose a novel framework, Individualized Fusion Learning (iFusion), to enhance EVI estimation across multiple data sources. The core idea is to enhance estimation for a target variable by strategically leveraging the information from the tails of related sources, thereby improving efficiency while maintaining statistical validity. Under an independence assumption, we construct an iFusion estimator based on the Hill estimator and establish its consistency and asymptotic normality. We extend the method to accommodate tail-dependent sources via an adjusted fusion strategy. Simulation studies and an application to ozone concentration data demonstrate the significantly improved performance of the iFusion method, particularly in variance reduction and out-of-sample prediction, highlighting its immediate practical value for multi-source extreme value analysis.

Key words and phrases: Asymptotic normality, Extreme value index, Individualized fusion learning, Multi-source information borrowing, Variance reduction.

Corresponding authors: Yu Chen, E-mail:cyu@ustc.edu.cn. Zhengjun Zhang, E-mail: zjz@stat.wisc.edu; ORCID: 0000-0003-2615-1539

1. Introduction

The era of big data has ushered in unprecedented opportunities for statistical inference, propelled by rapid advancements in data acquisition, storage, and sharing technologies. One compelling consequence is the increasing prevalence of multiple related datasets collected from heterogeneous sources. This naturally leads to combining or integrating such datasets to improve estimation and inference, particularly when the target dataset alone may not suffice. Terms such as data fusion, integration, aggregation, and pooling have emerged in the statistics and machine learning communities to describe this growing area of research. Concurrently, various modeling paradigms – such as transfer learning, semi-supervised learning, and fusion learning – have been developed to meet the challenges posed by such settings. The appropriateness of these approaches often depends on the structure of the data and the inferential goals.

The concept of data fusion has found broad applications in recent years across multiple domains, including causal inference (Shi et al., 2023), treatment effect (Sun and Miao, 2022; Zhang and Zhu, 2025; Wu et al., 2025), regression modeling (Li et al., 2022; Tian and Feng, 2023; Jin et al., 2024), parameter estimation (Shen et al., 2020; Li and Luedtke, 2023), and factor modeling (Duan et al., 2024). Motivated by this progress, we explore in this paper how fusion learning—specifically, individualized fusion learning (iFusion)—can enhance the estimation of the extreme value index (EVI), a fundamental parameter in extreme value theory. By borrowing strength from auxiliary sources with related tail behavior, we aim to improve estimation efficiency while maintaining statistical rigor.

1.1 Motivation

Despite its wide-ranging success, data fusion remains underutilized in analyzing rare events. This is particularly noteworthy in extreme value statistics, where the adequate sample size is often limited due to the very nature of tail data—only extreme observations are retained, and these unique occurrences are infrequent. The inherent limitations of

extreme value statistics underscore the need for alternative methods, such as data fusion, to improve inference. To this end, we address two fundamental questions: (i) What inferential problem in extreme value theory stands to benefit most from the idea of data fusion? (ii) How can data fusion be practically and theoretically implemented in this context?

Regarding the first question, we focus on estimating the EVI, a cornerstone of extreme value theory. The EVI determines the rate at which the tail of a distribution decays and serves as a building block for extrapolation techniques such as the Weissman-type estimator and for risk assessment tools like extreme quantiles and expectiles. Accurate estimation of EVI is thus essential for reliable inference in fields such as finance, insurance, and environmental science—domains where multiple datasets with potentially related tail behavior are routinely encountered.

To address the second question, we adopt the framework of fusion learning. Classical fusion learning aggregates inference results across multiple sources to improve efficiency. A more flexible variant – individualized fusion learning (iFusion) – was recently introduced by Shen et al. (2020). Unlike traditional methods, iFusion does not assume identical parameters across sources, allowing for source-specific variability. This feature is especially valuable in extreme value analysis, where tail indices may differ slightly but exhibit functional similarities. For a broader review of iFusion, see Cai et al. (2020). In this paper, we leverage iFusion to improve EVI estimation for a specific target dataset by incorporating information from related but potentially heterogeneous sources.

1.2 Related work

The estimation of EVI has a long and rich history. Classical extreme value theory primarily focuses on independent and identically distributed (IID) samples. However, several works have developed EVI estimation methods to accommodate more complex dependence structures and data settings; see Hsing (1991), Drees (2000), Einmahl et al. (2016), de Haan and Zhou (2021), and the monograph by Kulik and Soulier (2020).

More recently, attention has turned to the estimation of EVI in multi-source or large-scale settings. Chen et al. (2022) applied a divide-and-conquer framework, where the average of subsample-based EVI estimators is used. Li et al. (2024) proposed a subsampling-based method to alleviate memory constraints. In semi-supervised setups, Ahmed and Einmahl (2019) and Ahmed et al. (2025) utilized unlabeled data to enhance EVI estimation based on Hill and pseudo-MLE estimators. Daouia et al. (2024) proposed a weighted pooled Hill estimator, improving over naive averaging by introducing data-driven weights.

Table 1: Summary of relevant literature on the extension of EVI estimation

Extension to multiple data sources	
Chen et al. (2022)	Divide-and-conquer algorithm; Hill estimator
Li et al. (2024)	Subsampling-based methodology; MLE
Ahmed and Einmahl (2019)	Semi-supervised framework; Hill estimator
Ahmed et al. (2025)	Semi-supervised framework; pseudo-MLE estimator
Daouia et al. (2024)	Weighted pooled Hill estimator
This paper	Individualized fusion learning; Hill estimator

1.3 Contributions and paper organization

The works mentioned above typically assume a common EVI across all data subsets or sources. In contrast, we consider a more general and practically relevant scenario: the target dataset has limited extreme observations, and the auxiliary sources may have slightly different EVIs. This setting presents several challenges specific to tail inference. EVI estimation relies on only an intermediate number of upper-order observations, so the effective information from each source is determined by its tail sample size rather than its full sample size. Source similarity must therefore be assessed relative to tail-estimation uncertainty, and auxiliary sources that are difficult to distinguish from the target may

still induce a non-negligible local shift in the fused estimator. Moreover, source-specific Hill estimators may exhibit heterogeneous second-order biases, while dependence across sources must be characterized through their joint tail behavior. These features are not directly addressed by existing distributed or semi-supervised approaches. Accordingly, we adapt the iFusion framework using tail-based confidence distributions, data-driven screening weights, and asymptotic analyses tailored to EVI estimation. The resulting estimator accommodates parameter heterogeneity while enhancing inference for the target. Both numerical simulations and real-data analyses further highlight its advantages in reducing variance and improving out-of-sample prediction.

To position our work within the broader literature, Table 1 provides a comparative summary of recent approaches for EVI estimation using multiple sources. The main contributions of this work are twofold: (i) We propose a novel iFusion-based methodology to estimate the EVI for a target dataset by borrowing strength from multiple auxiliary sources. The resulting estimator is shown to be consistent and asymptotically normal. Notably, its asymptotic variance characterizes the efficiency gain from auxiliary data; (ii) For settings involving tail dependence, we modify our fusion strategy to account for inter-source dependence in extreme observations and establish the associated theoretical properties.

The rest of the paper is organized as follows. Section 2 provides the preliminary concepts and results, including a brief review of extreme value theory in Section 2.1 and the confidence distribution framework for extreme value index in Section 2.2. Section 3 defines the multiple data source framework and presents the iFusion-based estimator. Section 4 provides theoretical guarantees, and Section 5 extends our method to tail-dependent settings. Section 6 reports simulation results, and Section 7 presents real data applications. Section 8 concludes the paper with a few remarks and a discussion of potential directions for further development of the iFusion methodology for extreme value inference. All proofs and additional materials are available in the Supplementary

Material.

2. Preliminaries

2.1 Extreme value theory

Consider a distribution F which belongs to the max-domain of attraction (MDA) of an extreme value distribution G_γ with index $\gamma \in \mathbb{R}$, denoted by $F \in \mathrm{MDA}(G_\gamma)$. Mathematically, there exist sequences of constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) := \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\}$$

for all $1 + \gamma x > 0$. For $\gamma = 0$, the right hand side is interpreted as $\exp(-e^{-x})$. The parameter γ is called the extreme value index (EVI).

In this paper, we restrict our attention to the heavy-tailed case, i.e., $\gamma > 0$. In the classical setting, the Hill estimator (Hill, 1975) is a common choice for EVI in the heavy-tailed case due to its simplicity and practicality. Suppose $Y_1, Y_2, \cdots, Y_m$ are IID observations drawn from F, and we order the observations as $Y_{1,m} \le Y_{2,m} \le \cdots \le Y_{m,m}$. For a suitable intermediate sequence $k_H = k_H(m)$, i.e., $\lim_{m\to\infty} k_H = \infty, \lim_{m\to\infty} k_H/m = 0$, the Hill estimator is defined as

$$\widehat{\gamma}_H := \frac{1}{k_H}\sum_{i=1}^{k_H} \log(Y_{m-i+1,m}) - \log(Y_{m-k_H,m}).$$

We refer to k_H as the effective sample size or tail sample size, which represents the number of upper-order statistics used in the estimation. To reveal the asymptotic properties of the Hill estimator, the regularly varying conditions are usually required. We present the following characterization of the regularly varying conditions: the first-order condition $\mathbf{C}_1(\gamma)$ and second-order condition $\mathbf{C}_2(\gamma, \rho, A)$.

$\mathbf{C}_1(\gamma)$: The survival function $\overline{F} = 1 - F$ is regularly varying at infinity with index $-1/\gamma$, i.e., for all $x > 0$,

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-1/\gamma}. \tag{2.1}$$

$\mathbf{C}_2(\gamma, \rho, A)$: The survival function $\overline{F}$ satisfies, for all $x > 0$,

$$\lim_{t\to\infty} \frac{1}{A(1/\overline{F}(t))}\left(\frac{\overline{F}(tx)}{\overline{F}(t)} - x^{-1/\gamma}\right) = x^{-1/\gamma}\frac{x^{\rho/\gamma}-1}{\gamma\rho},$$

where $\gamma > 0, \rho < 0$, and A is a positive or negative function with $\lim_{t\to\infty} A(t) = 0$.

For the sake of presentation, we restate the consistency and asymptotic normality results for the Hill estimator from de Haan and Ferreira (2006) as the following proposition.

Proposition 1. *(de Haan and Ferreira, 2006) Let $Y_1, Y_2, \ldots, Y_m$ be IID random variables with distribution F, and the sequence $k_H = k_H(m)$ satisfies $k_H \to \infty, k_H/m \to 0$ as $m \to \infty$. Then as $m \to \infty$, it follows that*

(i) under condition $\mathbf{C}_1(\gamma)$, $\widehat{\gamma}_H \xrightarrow{p} \gamma$.

(ii) under condition $\mathbf{C}_2(\gamma, \rho, A)$,

$$\sqrt{k_H}\,(\widehat{\gamma}_H - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right), \quad \text{with } \lambda := \lim_{m\to\infty} \sqrt{k_H} A(m/k_H). \tag{2.2}$$

The asymptotic normality in Proposition 1 requires the second-order condition to hold, which is stronger than the condition for consistency. In addition, Eq.(2.2) demonstrates asymptotic normality with a non-zero mean. Furthermore, de Haan and Ferreira (2006) provides a detailed discussion of the cases where the bias part $\lambda/(1-\rho)$ is replaced by zero. Meanwhile, the bias-corrected Hill estimator was proposed in Caeiro et al. (2005). Its potential incorporation into the iFusion framework, together with the additional issues involved in such an extension, is discussed in Remark 1.

2.2 Confidence distribution of extreme value index

Confidence distributions (CDs) are a core ingredient of the iFusion methodology, as source-wise inferences are summarized in the form of CDs and then strategically combined to enhance individualized inference (see Shen et al., 2020). We first briefly review the concept of a CD before constructing one for the EVI.

In classical statistical inference, point or interval estimates are commonly used for unknown parameters. A CD estimates the unknown parameter of interest in distributional form. Unlike the former two, a CD, which is a distribution estimate, provides meaningful information, including point estimation, confidence interval, and p-values (see Xie and Singh, 2013). To illustrate, consider a mean inference problem with random sample $Z_i \sim N(\mu, \sigma^2), i = 1, \ldots, n$, where σ^2 is known but μ is unknown. It is well known that $\bar{Z}$ is a standard point estimator and $[\bar{Z} + \Phi^{-1}(\alpha/2)\sigma/n^{1/2}, \bar{Z} + \Phi^{-1}(1-\alpha/2)\sigma/n^{1/2}]$ is an interval estimator. We can also use $N(\bar{Z}, \sigma^2/n)$ or more formally in its cumulative distribution form $H_\Phi(\mu) = \Phi(\sqrt{n}(\mu - \bar{Z})/\sigma)$ as a distribution function to estimate μ. As a distribution function defined on the entire parameter space, a CD contains a wealth of information for inference and reproduces the usual point and interval procedures as special summaries. It plays a role analogous to a Bayesian posterior, but is constructed to retain frequentist validity. For a comprehensive introduction to confidence distributions, see Cui and Xie (2023).

We focus on choosing a CD for the heavy-tailed regime $\gamma > 0$. The Hill estimator $\widehat{\gamma}_H$ is a point estimator of γ, and its asymptotic normality provides a natural basis for constructing a CD for γ. Although Eq.(2.2) in Proposition 1 allows for a non-zero mean, we adopt the bias-negligible regime $\lambda = 0$ (i.e., $\sqrt{k_H}A(m/k_H) \to 0$ as $m \to \infty$) as a working approximation to obtain a simple and tractable CD. Under this condition, the sample-dependent normal distribution $N(\widehat{\gamma}_H, \widehat{\gamma}_H^2/k_H)$ is an approximate CD for γ. We use this CD in the subsequent iFusion implementation; importantly, the assumption $\lambda = 0$ is not imposed when establishing the theoretical properties of the proposed estimator. A similar bias-negligible strategy can be found in Ahmed and Einmahl (2019).

Remark 1. The bias-negligible approximation $\lambda = 0$ is adopted only to construct a simple and tractable working CD; it is not an intrinsic restriction of the iFusion framework. In principle, each source-wise CD may instead be centered at a bias-reduced Hill estimator and equipped with a corresponding variance estimate, with the screening weights

recalibrated accordingly. The resulting source-wise CDs can then be combined using the same general screening-and-fusion principle. Such a modification may attenuate or remove the leading source-specific second-order bias terms under appropriate conditions. However, estimating the required second-order parameters may be unstable when tail samples are limited and introduces additional uncertainty into the variance estimation and screening procedure. A formal extension would therefore require re-establishing the screening properties and asymptotic distribution of the fused estimator. We use the classical Hill-based working CD in this paper and leave a systematic study of bias-corrected iFusion to future research.

3. Methodology

3.1 Framework and notations

For multiple data sources or individuals, consider N individual subjects with a dataset $\mathbb{S} = \{\mathcal{S}_1, \cdots, \mathcal{S}_N\}$, where $\mathcal{S}_j$ contains n_j observations collected from the j-th individual, $j = 1, \cdots, N$. The total sample size $n = \sum_{j=1}^{N} n_j$. For the j-th individual, the observations are assumed to be IID from an individual-specific distribution F_j:

$$\begin{aligned} X_1^{(1)}, X_2^{(1)}, \ldots, X_{n_1}^{(1)} &\sim \text{IID } F_1, \\ X_1^{(2)}, X_2^{(2)}, \ldots, X_{n_2}^{(2)} &\sim \text{IID } F_2, \\ \vdots \qquad\qquad\qquad &\qquad\quad \vdots \\ X_1^{(N)}, X_2^{(N)}, \ldots, X_{n_N}^{(N)} &\sim \text{IID } F_N. \end{aligned}$$

Moreover, the samples are independent across individuals; that is, $X_\ell^{(j)}$ are mutually independent across $1 \le j \le N$ and $1 \le \ell \le n_j$. This setup allows for heterogeneity across individuals, in the sense that F_j may differ with j, which is more commonly encountered in practice. We consider the continuous and right heavy-tailed distribution F_j with an unknown positive EVI, denoted by γ_j. Notably, we only choose one individual as the target object to estimate the corresponding EVI instead of all the indices. Without loss

of generality, we treat individual-1 as the target, and accordingly aim to make a valid and efficient inference about γ_1.

Estimation for the EVI is a well-studied problem, with several methods detailed in de Haan and Ferreira (2006). In the classical setting, inference on γ_1 is based solely on the target data source $\mathcal{S}_1 := \{X_1^{(1)}, X_2^{(1)}, \cdots, X_{n_1}^{(1)}\}$, and a variety of standard EVI estimators can be directly applied to $\mathcal{S}_1$. In modern data-rich applications, however, information is often available beyond the target individual, potentially including observations from other individuals that may share similar tail behavior. This setting motivates us to develop the estimation of γ_1 through a special fusion learning method called iFusion.

Let k_j denote the tail sample size[1] for individual-j and $k := \sum_{j=1}^{N} k_j$ be the total tail sample size, which is typically chosen as an intermediate sequence relative to the full sample size n. For theoretical analysis, iFusion methodology advocates classifying the N parameters $\{\gamma_1, \gamma_2, \ldots, \gamma_N\}$ according to the scaled distance to the target γ_1 at the $k^{-1/2}$ resolution, and partitions them into three disjoint sets:

- Clique: $\mathcal{C}_1 = \{\gamma_j : k^{1/2}|\gamma_j - \gamma_1| = o(1),\ j = 1, \cdots, N\}$;
- Boundary set: $\mathcal{B}_1 = \{\gamma_j : k^{1/2}|\gamma_j - \gamma_1| \to c, 0 < c < \infty,\ j = 1, \cdots, N\}$, where c is some constant;
- Disperse set: $\mathcal{D}_1 = \{\gamma_j : k^{1/2}|\gamma_j - \gamma_1| \to \infty,\ j = 1, \cdots, N\}$.

The expression of clique $\mathcal{C}_1$ is akin to the so-called "near tie" regime. Equivalently, it can be characterized through a shrinking neighborhood around the target parameter, i.e.,

$$\mathcal{C}_1 = \{\gamma_j :\ \gamma_j \in \mathbf{B}_r(\gamma_1),\ j = 1, \ldots, N\},$$

where $\mathbf{B}_r(\gamma_1)$ is a ball centered at γ_1 with radius $r = o\left(k^{-1/2}\right)$. In particular, $\mathcal{C}_1$ always contains γ_1, and for any $\gamma_j \in \mathcal{C}_1,\ j \neq 1$, it is indistinguishable from γ_1 by its $\sqrt{k}$-consistent estimator based on the current sample size. The two extreme cases correspond to $|\mathcal{C}_1| = 1$ with only the target source retained and $|\mathcal{C}_1| = N$ with all sources asymptotically tied to

[1] In extreme value inference, the *tail sample size* refers to the number of upper-tail observations used in estimation; see Section 2.1 for the Hill estimator and the role of k_H.

the target. The set $\mathcal{C}_1$ only requires the parameters to be sufficiently close to the target one, without imposing exact equality. In contrast, the inclusion of the individuals in $\mathcal{D}_1$ may incur non-negligible bias.

For notational simplicity, we omit the subscript and refer to the sets as $\mathcal{C}, \mathcal{B}$, and $\mathcal{D}$. We denote $\mathbb{I}_{\mathcal{C}}$ as the index set for the elements in $\mathcal{C}$, i.e., $\mathbb{I}_{\mathcal{C}} = \{j : \gamma_j \in \mathcal{C}\}$, and $\mathbb{I}_{\mathcal{B}}, \mathbb{I}_{\mathcal{D}}$ are similar. Further, let $N_{\mathcal{C}}, N_{\mathcal{B}}$ and $N_{\mathcal{D}}$ denote the cardinalities of the sets $\mathcal{C}$, $\mathcal{B}$ and $\mathcal{D}$, respectively.

3.2 iFusion-based estimation

In this subsection, we proceed to implement the iFusion approach by combining the CDs of EVI and derive a general form of the iFusion estimator.

For the j-th data source, let $\widehat{\gamma}_j$ denote the estimator of the EVI γ_j by applying the classical Hill method to the dataset $\mathcal{S}_j$. Under the CD approximation introduced in Section 2.2, we associate $\widehat{\gamma}_j$ with an asymptotic normal confidence density

$$h_j(\gamma; \mathcal{S}_j) = \frac{1}{\sqrt{2\pi}\widehat{\sigma}_j} \exp\left\{-\frac{(\gamma - \widehat{\gamma}_j)^2}{2\widehat{\sigma}_j^2}\right\}, \quad \widehat{\sigma}_j^2 = \widehat{\gamma}_j^2/k_j. \tag{3.3}$$

Given screening weights $w_{1j} \in [0, 1]$, iFusion combines the source-wise information by maximizing the weighted confidence density and obtains the estimator

$$\widehat{\gamma}_1^{(c)} = \arg\max_{\gamma} \prod_{j=1}^{N} h_j(\gamma; \mathcal{S}_j)^{w_{1j}} = \arg\max_{\gamma} \sum_{j=1}^{N} w_{1j} \log h_j(\gamma; \mathcal{S}_j). \tag{3.4}$$

Here w_{1j} is the screening weight for individual-j with respect to individual-1, with large w_{1j} indicating the higher degree of relevance of individual-j to individual-1. When the individual confidence density functions take the form of Eq.(3.3), simple algebra yields

$$\widehat{\gamma}_1^{(c)} = \Big(\sum_{j=1}^{N} w_{1j} k_j \widehat{\gamma}_j^{-2}\Big)^{-1} \sum_{j=1}^{N} w_{1j} k_j \widehat{\gamma}_j^{-1} := \sum_{j=1}^{N} \eta_j^{(c)} \widehat{\gamma}_j, \tag{3.5}$$

where $\eta_j^{(c)} = w_{1j} k_j \widehat{\gamma}_j^{-2} \Big/ \Big(\sum_{j=1}^{N} w_{1j} k_j \widehat{\gamma}_j^{-2}\Big)$. It is evident that $\eta_j^{(c)}$ satisfies $0 \le \eta_j^{(c)} \le 1$ and $\sum_{j=1}^{N} \eta_j^{(c)} = 1$. Hence $\widehat{\gamma}_1^{(c)}$ is a weighted linear combination of the source-wise EVI estimators, with weights $\eta_j^{(c)} \propto w_{1j} \widehat{\sigma}_j^{-2}$, showing that the effective contribution of each

source is governed by both screening relevance and statistical efficiency. In particular, a source that is substantially biased is typically less compatible with the target and thus receives a small screening weight w_{1j}, preventing it from being overly trusted even if its variance is small.

To achieve the maximum efficiency gain by iFusion, it is desirable that the screening weights w_{1j} satisfy the following condition, which also serves as an ideal benchmark.

C₁(w): For each $j = 1, \cdots, N$, there exist nonnegative sequences a_j and b_j satisfying $a_j = o(k^{-1/2})$, $b_j = o(k^{-1/2})$, such that

$$w_{1j} = \begin{cases} 1 - a_j, & \text{if } j \in \mathbb{I}_{\mathcal{C}}, \\ b_j, & \text{otherwise}, \end{cases} \tag{3.6}$$

where $\mathbb{I}_{\mathcal{C}}$ is the index set of clique $\mathcal{C}$.

The weights are advocated to balance the contributions of the information in the three sets $\mathcal{C}, \mathcal{B}, \mathcal{D}$. A unified and practically implementable choice of weights is provided in Section S1 of the Supplementary Material.

4. Theoretical Properties

To study the theoretical properties of our estimators, we begin by specifying a set of regularity conditions on the multi-source setting. Specifically, we first assume that the number of data sources N is fixed, and further collect the required sample-size conditions in the following regime.

C(n, k): As total sample size $n \to \infty$, the source-specific sample sizes satisfy $n_j/n \to \pi_j \in (0, 1)$, the tail sample sizes satisfy

$$k = k(n) \to \infty, \quad \frac{k}{n} \to 0, \quad \frac{k_j}{k} \to c_j \in (0, 1), \quad j = 1, \ldots, N.$$

Consequently, $k_j \to \infty$, $k_j/n_j \to 0$, $j = 1, \ldots, N$.

Theorem 1. *Suppose the marginal distribution F_j satisfies first-order condition $\mathbf{C}_1(\gamma_j)$ with $\gamma_j > 0$ and the screening weight w_{1j} satisfies condition $\mathbf{C}_1(w)$, for each $j = 1, \cdots, N$.*

Then, under condition $\mathbf{C}(n,k)$,

$$\widehat{\gamma}_1^{(c)} \xrightarrow{p} \gamma_1.$$

While the consistency result in Theorem 1 establishes the validity of our iFusion estimator, it does not reveal its advantages over the classical Hill estimator $\widehat{\gamma}_1$. To address this limitation, we need to analyze the rate of the estimator and investigate its asymptotic distribution, which is studied in the following two subsections.

4.1 Oracle estimation

To evaluate the performance of the iFusion estimator, we first introduce the oracle estimator as an ideal baseline and explore its asymptotic distribution in this subsection.

Suppose that the set $\mathbb{I}_{\mathcal{C}}$ is known a priori, that is, we have perfect knowledge of which data sources are most beneficial to include in the fusion. We regard this case as the oracle case, and the oracle estimator of γ_1 is defined by pretending the membership of $\mathbb{I}_{\mathcal{C}}$ is completely known, i.e., setting the screening weights to $w_{1j} = 1$ for $j \in \mathbb{I}_{\mathcal{C}}$ (and $w_{1j} = 0$ otherwise). Under this idealized setting, the oracle estimator can be expressed in a mathematical form:

$$\widehat{\gamma}_1^{(o)} = \arg\max_{\gamma} \prod_{j\in\mathbb{I}_{\mathcal{C}}} h_j(\gamma;\mathcal{S}_j) = \arg\max_{\gamma} \sum_{j\in\mathbb{I}_{\mathcal{C}}} \log h_j(\gamma;\mathcal{S}_j).$$

Under the assumption of normally distributed individual confidence densities, it is straightforward to verify that

$$\widehat{\gamma}_1^{(o)} = \Big(\sum_{j\in\mathbb{I}_{\mathcal{C}}} k_j\widehat{\gamma}_j^{-2}\Big)^{-1} \sum_{j\in\mathbb{I}_{\mathcal{C}}} k_j\widehat{\gamma}_j^{-1} \quad = \sum_{j\in\mathbb{I}_{\mathcal{C}}} \eta_j^{(o)}\widehat{\gamma}_j, \tag{4.7}$$

where the weight reduces to $\eta_j^{(o)} = \Big(\sum_{j\in\mathbb{I}_{\mathcal{C}}} k_j\widehat{\gamma}_j^{-2}\Big)^{-1} k_j\widehat{\gamma}_j^{-2}$.

Example 1. We illustrate the intuitive form of the estimator $\widehat{\gamma}_1^{(o)}$ with a toy example. Suppose $\mathcal{C} = \{\gamma_1,\gamma_2,\gamma_3\}$, and the corresponding index set $\mathbb{I}_{\mathcal{C}} = \{1,2,3\}$ is known, with the equal effective sample sizes, i.e., $k_1 = k_2 = k_3$. To construct an estimator for γ_1, a natural choice is to take the average of all estimators, i.e., $1/3 \cdot \sum_{j=1}^{3} \widehat{\gamma}_j$. However, the

oracle estimator for γ_1 can be written as

$$\widehat{\gamma}_1^{(o)} = \frac{\widehat{\gamma}_1^{-2}}{\sum_{j=1}^3 \widehat{\gamma}_j^{-2}} \cdot \widehat{\gamma}_1 + \frac{\widehat{\gamma}_2^{-2}}{\sum_{j=1}^3 \widehat{\gamma}_j^{-2}} \cdot \widehat{\gamma}_2 + \frac{\widehat{\gamma}_3^{-2}}{\sum_{j=1}^3 \widehat{\gamma}_j^{-2}} \cdot \widehat{\gamma}_3,$$

and the weight $\eta_j^{(o)} = \widehat{\gamma}_j^{-2} \Big/ \sum_{j=1}^3 \widehat{\gamma}_j^{-2}$. A further assumption that all the indices are exactly equal does not guarantee the weight equality due to the sample difference. Hence, $\widehat{\gamma}_1^{(o)}$ takes a different form than simple average, despite $\eta_j^{(o)} \xrightarrow{p} 1/3$.

Theorem 2. *Suppose that the set $\mathbb{I}_{\mathcal{C}}$ is known, the marginal distributions F_j satisfy second-order condition $\mathbf{C}_2(\gamma_j, \rho_j, A_j)$ with $\lim_{n_j \to \infty} \sqrt{k_j} A_j\big(n_j/k_j\big) = \lambda_j$ for $j \in \mathbb{I}_{\mathcal{C}}$. Then, under condition $\mathbf{C}(n,k)$,*

(i) $\sqrt{k}(\widehat{\gamma}_1^{(o)} - \gamma_1) \xrightarrow{d} N\big(\mu_1^{(o)}, \Delta_1^{(o)}\big)$, where

$$\mu_1^{(o)} = \big(\sum_{j \in \mathbb{I}_{\mathcal{C}}} c_j \gamma_j^{-2}\big)^{-1} \sum_{j \in \mathbb{I}_{\mathcal{C}}} c_j^{1/2} \gamma_j^{-2} \Big(\frac{\lambda_j}{1-\rho_j}\Big), \quad \Delta_1^{(o)} = \big(\sum_{j \in \mathbb{I}_{\mathcal{C}}} c_j \gamma_j^{-2}\big)^{-1}.$$

(ii) For any nonempty index set $\mathcal{F} \subseteq \mathbb{I}_{\mathcal{C}}$, define

$$\widehat{\gamma}_1^{\mathcal{F}} = \arg\max_{\gamma > 0} \log \prod_{j \in \mathcal{F}} h_j\left(\gamma; \mathcal{S}_j\right). \tag{4.8}$$

Then $\widehat{\gamma}_1^{(o)} = \widehat{\gamma}_1^{\mathbb{I}_c}$ attains the minimal first-order asymptotic variance among all such $\widehat{\gamma}_1^{\mathcal{F}}$. In the bias-negligible case $\lambda_j = 0,\ j \in \mathbb{I}_{\mathcal{C}}$, it also attains the minimal first-order asymptotic MSE.

There are a few things to note regarding these results. Results in Theorem 2(i) imply that the oracle estimator is asymptotically normal. As a special case where $\lambda_j = 0, j \in \mathbb{I}_{\mathcal{C}}$, the above asymptotic normality yields a zero mean. We use the global normalization $k^{1/2}$ primarily to keep a unified scaling across all methods and results. The oracle result can be equivalently written with the clique-level tail size $k_C := \sum_{j \in \mathbb{I}_{\mathcal{C}}} k_j$, and the difference is only a constant rescaling under our fixed-N framework. We can view $\widehat{\gamma}_1^{(o)}$ as a desirable consistent estimator of γ_1, in the sense that it achieves the minimal asymptotic variance among the class of linear combinations $\big\{\sum_{j \in \mathbb{I}_{\mathcal{C}}} \eta_j \widehat{\gamma}_j\big\}$, see Remarks 3 and 4 for details.

In addition, we can estimate the asymptotic variance of $\widehat{\gamma}_1^{(o)}$ by

$$\widehat{\Delta}_1^{(o)}/k = \Big(\sum_{j\in\mathbb{I}_{\mathcal{C}}} \frac{k_j}{k}\widehat{\gamma}_j^{-2}\Big)^{-1}\Big/k = \Big(\sum_{j\in\mathbb{I}_{\mathcal{C}}} k_j\widehat{\gamma}_j^{-2}\Big)^{-1}.$$

Result(ii) of Theorem 2 further shows that choosing the full clique index set, $\mathcal{F} = \mathbb{I}_{\mathcal{C}}$, yields the smallest asymptotic variance among all estimators given in the form of Eq.(4.8). Note that the individual estimator $\widehat{\gamma}_1$ itself is a special case of $\widehat{\gamma}_1^{\mathcal{F}}$ with $\mathcal{F} = \{1\}$. Thus, the oracle estimator borrows information from all sources whose EVIs are sufficiently close to the target EVI at the $k^{-1/2}$ scale.

Remark 2.(Comparison with the classical Hill estimator) Proposition 1 gives the asymptotic distribution of $\widehat{\gamma}_1$ can be expressed as $k_1^{1/2}\big(\widehat{\gamma}_1 - \gamma_1\big) \xrightarrow{d} N\Big(\frac{\lambda_1}{1-\rho_1}, \gamma_1^2\Big)$. For the sake of comparison, we present the following alternative form by replacing k_1 with k,

$$k^{1/2}\big(\widehat{\gamma}_1 - \gamma_1\big) \xrightarrow{d} N\Big(\frac{c_1^{-1/2}\lambda_1}{1-\rho_1}, c_1^{-1}\gamma_1^2\Big) =: N\big(\mu_H, \Delta_H\big).$$

For our oracle estimator, the bias and variance terms can be written as

$$\mu_1^{(o)} = \sum_{j\in\mathbb{I}_{\mathcal{C}}} \eta_j c_j^{-1/2}\Big(\frac{\lambda_j}{1-\rho_j}\Big),\ \Delta_1^{(o)} = \sum_{j\in\mathbb{I}_{\mathcal{C}}} \eta_j^2 c_j^{-1}\gamma_j^2,\ \text{with } \eta_j = \frac{c_j\gamma_j^{-2}}{\sum_{j\in\mathbb{I}_{\mathcal{C}}} c_j\gamma_j^{-2}}.$$

Considering the simplest cases where the parameters in set $\mathcal{C}$ share the same value, and each individual takes the same tail sample size, which leads to $\eta_j = 1/N_{\mathcal{C}}$, and the final result is that $\mu_1^{(o)} = \mu_H, \Delta_1^{(o)} = \Delta_H/N_{\mathcal{C}} \le \Delta_H$.

Remark 3.(Comparison with Chen et al. (2022)) Recall that Chen et al. (2022) considered the case where all samples are from the same heavy-tailed distribution F and thus share a common EVI γ. Their framework differs from ours in that the number of data sources N may be infinite, while each k_j can be a fixed integer. To align their setup with ours, suppose that for $j = 1, \cdots, N$, the source distributions are identical, i.e., $F_j =: F$ and that F satisfies the second-order condition $\mathbf{C}_2(\gamma, \rho, A)$. It then follows that $\gamma_1 = \gamma_2 = \cdots = \gamma_N =: \gamma$ and $\mathcal{C} = \{\gamma_j : j = 1, 2, \ldots, N\}$. The distributed Hill estimator (Chen et al., 2022) can be viewed as a simple average of subsample EVI estimators,

$\widehat{\gamma}_{\mathrm{DH}} = \frac{1}{N}\sum_{j=1}^{N}\widehat{\gamma}_j$. Then we find that the two estimators exhibit the same asymptotic properties (as $n \to \infty$):

$$k^{1/2}\big(\widehat{\gamma}_{(\cdot)} - \gamma\big) \xrightarrow{d} N\left(\frac{\sqrt{N}\lambda}{1-\rho}, \gamma^2\right).$$

Remark 4.(Comparison with Daouia et al. (2024)) It is worth noting that Theorem 2 is, to some extent, consistent with Theorem 1 in Daouia et al. (2024). The general framework of the observations considered in our work is similar to that of Daouia et al. (2024); however, their study focuses on the homogeneous case in which $\gamma_j, j = 1, \cdots, m$ are equal to a common γ, where m denotes the number of subsamples in their notation. Their main object of study is the class of weighted pooled Hill estimators that aggregate information across subsamples:

$$\widehat{\gamma}_n(\boldsymbol{\omega}) = \widehat{\gamma}_n(\omega_1, \ldots, \omega_m) = \sum_{j=1}^{m}\omega_j\widehat{\gamma}_j = \boldsymbol{\omega}^\top\widehat{\boldsymbol{\gamma}}_n, \text{ with } \boldsymbol{\omega}^\top\mathbf{1} = 1.$$

Under the assumption that the data from all individuals are independent, the theoretically variance-optimal weights reduce to

$$\boldsymbol{\omega}^{(\mathrm{Var})} := \Big(\omega_1, \omega_2, \cdots, \omega_m\Big)^\top, \text{ with } \omega_j = \Big(\sum_{\ell=1}^{m} k_\ell\gamma_\ell^{-2}\Big)^{-1} k_j\gamma_j^{-2}, \ j = 1, 2, \cdots, m. \quad (4.9)$$

Plugging in $\widehat{\gamma}_j$ for γ_j in Eq.(4.9) gives an estimator $\widehat{\boldsymbol{\omega}}$, which is identical to our weights $\eta_j^{(o)}$ in Eq.(4.7).

4.2 iFusion estimation

In this subsection, we consider a more general and practical case. Actually, the membership of $\mathbb{I}_{\mathcal{C}}$ is typically unknown in practice, and it is often challenging to distinguish indices in $\mathbb{I}_{\mathcal{C}}$ from those in $\mathbb{I}_{\mathcal{B}}$. For simplicity, we first impose the assumption that the sets $\mathcal{B}$ and $\mathcal{C}$ are distinguishable, and refer to the following condition as the separation condition.

C(d): The boundary set $\mathcal{B} = \varnothing$, or equivalently, $k^{1/2}d \to \infty$ with $d = \min_j\big\{|\gamma_1 - \gamma_j| : \gamma_j \notin \mathcal{C}\big\}$.

Theorem 3. *Suppose that w_{1j} satisfies condition $\mathbf{C_1}(w)$, and the separation condition $\mathbf{C}(d)$ also holds. The second-order condition $\mathbf{C}_2(\gamma_j, \rho_j, A_j)$ holds for distribution $F_j, j = 1, \ldots, N$. Then, under condition $\mathbf{C}(n, k)$,*

(i) $k^{1/2}(\widehat{\gamma}_1^{(c)} - \gamma_1) \xrightarrow{d} N\big(\mu_1^{(o)}, \Delta_1^{(o)}\big)$, where $\mu_1^{(o)}$ and $\Delta_1^{(o)}$ are defined in Theorem 2.

(ii) $\widehat{\gamma}_1^{(c)}$ inherits the first-order asymptotic variance optimality of the oracle estimator $\widehat{\gamma}_1^{(o)}$ among estimators $\widehat{\gamma}_1^{\mathcal{F}}$ with $\varnothing \neq \mathcal{F} \subseteq \mathbb{I}_{\mathcal{C}}$. In the bias-negligible case where $\lambda_j = 0,\ j \in \mathbb{I}_{\mathcal{C}}$, it also inherits the corresponding first-order asymptotic MSE optimality.

Theorem 3 demonstrates that the iFusion estimator $\widehat{\gamma}_1^{(c)}$ with appropriate weights performs as well as the oracle approach asymptotically, even without knowing the membership of $\mathbb{I}_{\mathcal{C}}$, only requiring that the parameters in $\mathcal{C}$ can be separated from those in $\mathcal{B}$. However, the parameters in $\mathcal{B}$ are not easy to separate from those in $\mathcal{C}$ by using data alone, and including sources in $\mathcal{B}$ may reduce the standard deviation but also introduce a deterministic shift of the same order. It is worth emphasizing that, although the separation may not hold in the application, Theorem 3 is still significant in that it serves as a bridge between the asymptotic distributions of oracle estimator $\widehat{\gamma}_1^{(o)}$ and iFusion estimator $\widehat{\gamma}_1^{(c)}$.

We now turn to the case that $\mathcal{B} \neq \varnothing$, i.e., the separation condition does not hold. In this setting, the condition $\mathbf{C_1}(w)$ as Eq.(3.6) is typically difficult to satisfy. We therefore adopt a less restrictive requirement and impose the following condition.

$\mathbf{C}_2(w)$: For each $j = 1, \cdots, N$, the screening weights satisfy

$$w_{1j} = \begin{cases} 1 - a_j, & \text{if } j \notin \mathbb{I}_{\mathcal{D}}; \\ b_j, & \text{otherwise} \end{cases} \tag{4.10}$$

for nonnegative sequences a_j and b_j that satisfy $a_j = o(k^{-1/2}), b_j = o(k^{-1/2})$.

Eq.(4.10) implies that no strict distinction is drawn between sets $\mathcal{B}$ and $\mathcal{C}$. In fact, for both conditions $\mathbf{C}_1(w)$ and $\mathbf{C}_2(w)$, we provide the specific construction of screening weights in Section S1 of Supplementary Material.

Remark 5. Although Theorem 1 is stated under condition $\mathbf{C}_1(w)$, its consistency conclusion remains valid when $\mathbf{C}_1(w)$ is replaced by $\mathbf{C}_2(w)$. Condition $\mathbf{C}_2(w)$ only relaxes the distinction between the clique and boundary sources, while the disperse sources are still sufficiently downweighted. Therefore, $\widehat{\gamma}_1^{(c)} \xrightarrow{p} \gamma_1$ also holds under condition $\mathbf{C}_2(w)$.

Theorem 4. *Assume that the screening weights w_{1j} satisfy condition $\mathbf{C}_2(w)$ and that each marginal distribution F_j satisfies the second-order condition $\mathbf{C}_2(\gamma_j, \rho_j, A_j)$, $j = 1, \ldots, N$. Suppose further that, for each $j \in \mathbb{I}_{\mathcal{B}}$, $B_j := \lim_{k\to\infty} k^{1/2}(\gamma_j - \gamma_1)$ exists and is finite. Then, under condition $\mathbf{C}(n, k)$,*

$$k^{1/2}\big(\widehat{\gamma}_1^{(c)} - \gamma_1\big) - B^{(c)} \xrightarrow{d} N\Big(\mu_1, \Delta_1\Big),$$

where $B^{(c)} = \big(\sum_{j\notin\mathbb{I}_{\mathcal{D}}} c_j\gamma_j^{-2}\big)^{-1}\sum_{j\in\mathbb{I}_{\mathcal{B}}} c_j\gamma_j^{-2}B_j$, *and*

$$\mu_1 = \big(\sum_{j\notin\mathbb{I}_{\mathcal{D}}} c_j\gamma_j^{-2}\big)^{-1}\sum_{j\notin\mathbb{I}_{\mathcal{D}}} c_j^{1/2}\gamma_j^{-2}\Big(\frac{\lambda_j}{1-\rho_j}\Big), \quad \Delta_1 = \big(\sum_{j\notin\mathbb{I}_{\mathcal{D}}} c_j\gamma_j^{-2}\big)^{-1}.$$

Theorem 4 explores the asymptotic behavior of the estimator $\widehat{\gamma}_1^{(c)}$ and reveals an intriguing result. Firstly, the weights are taken in the form of Eq.(4.10) due to the fact that the separation $\mathbf{C}(d)$ does not hold. Secondly, there exists a boundary-induced local shift $B^{(c)}$ that involves unknown parameter values in the boundary set $\mathcal{B}$.

5. Extension to the Dependent Case

The theoretical results in Section 4 are developed within the framework of independence across individuals, as described in Section 3.1. In this section, the assumptions on the data setting are relaxed to accommodate more complex scenarios, allowing for dependence structures among individuals. The rest of the settings remain the same. Notably, a straightforward divide-and-conquer implementation is no longer directly applicable in this scenario, since the estimation of any pairwise dependence structure requires the simultaneous availability of the pairwise samples. In other words, we need to collect

specific observations from all the data sources instead of the estimators. This requirement runs counter to the principle of distributed inference and can be viewed as the cost of accounting for dependence structures.

Once interdependence among individuals is considered, a key step is to characterize the dependence structure between the observations. In the context of extreme value from multiple data sources, the emphasis is on the tail dependence properties of the underlying variables. In multivariate extreme value theory, several analytical tools have been developed to describe the tail dependence structure of a random vector, including the exponent measure and stable tail dependence function (see de Haan and Ferreira, 2006). Here, we make use of the R-function, a transformation of the stable tail dependence function, to capture this structure.

For a pair of individuals (i, j), $i, j = 1, \cdots, N$, define the transformed variables $U = \overline{F}_i(X^{(i)})$ and $V = \overline{F}_j(X^{(j)})$, we describe their pairwise tail dependence through the bivariate survival copula

$$\overline{C}_{ij}(u, v) = \mathbb{P}(U \leq u, V \leq v) = \mathbb{P}(\overline{F}_i(X^{(i)}) \leq u, \overline{F}_j(X^{(j)}) \leq v), \quad u, v \in [0, 1].$$

Then we can characterize the joint extreme behavior of $(X^{(i)}, X^{(j)})$ via the following tail dependence condition.

C(R): For any (i, j) with $i, j = 1, 2, \cdots, N$ and $i \neq j$, there exists a function R_{ij} such that

$$R_{ij}(x_i, x_j) = \lim_{t \downarrow 0} \frac{1}{t} \overline{C}_{ij}(tx_i, tx_j), \quad (x_i, x_j) \in [0, \infty]^2 \backslash \{(\infty, \infty)\},$$

where $\overline{C}_{ij}(\cdot, \cdot)$ denotes the bivariate survival copula of $(X^{(i)}, X^{(j)})$.

Under condition **C**(R), the function $R_{ij}(\cdot, \cdot)$ is referred to as the (upper) tail copula of $(X^{(i)}, X^{(j)})$ in the literature. This pair-wise upper tail dependence assumption, along with the second-order condition for each marginal distribution, is sufficient to derive the following joint asymptotic distribution of $\widehat{\boldsymbol{\gamma}}_N := \left(\widehat{\gamma}_1, \widehat{\gamma}_2, \ldots, \widehat{\gamma}_N\right)^\top$, which is the key to the subsequent work in this article.

Proposition 2. *(Daouia et al., 2024)*

Assume that the condition $\mathbf{C}(R)$ *holds, and the marginal distribution* F_j *satisfies condition* $\mathbf{C}_2(\gamma_j, \rho_j, A_j)$ *for* $j = 1, \cdots, N$. *Then, under condition* $\mathbf{C}(n,k)$,

$$\left(\sqrt{k}(\widehat{\gamma}_1 - \gamma_1), \sqrt{k}(\widehat{\gamma}_2 - \gamma_2), \cdots, \sqrt{k}(\widehat{\gamma}_N - \gamma_N)\right)^\top \xrightarrow{d} \mathcal{N}(\boldsymbol{\mu}, \Sigma), \tag{5.11}$$

where

$$\boldsymbol{\mu} = \left(\frac{\sqrt{c_1^{-1}}\lambda_1}{1-\rho_1}, \cdots, \frac{\sqrt{c_N^{-1}}\lambda_N}{1-\rho_N}\right)^\top, \text{ and } \Sigma_{ij} = \begin{cases} c_i^{-1}\gamma_i^2, & i = j, \\ c_i^{-1}\frac{\pi_j}{\max(\pi_i,\pi_j)} R_{ij}\left(\frac{c_i}{c_j}, \frac{\pi_i}{\pi_j}\right)\gamma_i\gamma_j, & i < j. \end{cases}$$

and $\Sigma_{ji} = \Sigma_{ij}$ *for* $i < j$.

Proposition 2 indicates that the asymptotic distribution of the vector $\widehat{\boldsymbol{\gamma}}_N$ is a multivariate normal distribution. Notice that the result contains the independent case, which corresponds to $R_{ij}(\cdot,\cdot) \equiv 0$ for any $i < j$. For a detailed discussion about Proposition 2, please refer to Daouia et al. (2024).

We know that for any $j \in \mathbb{I}_{\mathcal{C}}$, the EVI γ_j satisfies $k^{1/2}(\gamma_j - \gamma_1) \xrightarrow{p} 0$. Combining this fact with the result in Proposition 2, and assuming that the membership of $\mathcal{C}$ is completely known, we propose the following procedure to derive the oracle estimator under the tail-dependent context.

Step 1. Obtain the joint distribution of $\widehat{\boldsymbol{\gamma}}_c$.

Replacing γ_j with γ_1 for all $j \in \mathbb{I}_{\mathcal{C}}$ in Eq.(5.11) and extracting the corresponding sub-vector, we obtain the joint asymptotic normality of $\widehat{\boldsymbol{\gamma}}_c := \left(\widehat{\gamma}_j\right)_{j\in\mathbb{I}_{\mathcal{C}}}^\top$ under the same conditions in Proposition 2, that is,

$$\sqrt{k}\Big((\widehat{\gamma}_j - \gamma_1)\Big)_{j\in\mathbb{I}_{\mathcal{C}}}^\top \xrightarrow{d} \mathcal{N}(\boldsymbol{\mu}_c, \Sigma_c),$$

where $\boldsymbol{\mu}_c$ is a vector and Σ_c is a symmetric matrix with

$$\mu_j = \frac{\sqrt{c_j^{-1}}\lambda_j}{1-\rho_j},\ j \in \mathbb{I}_{\mathcal{C}},\ \text{and}\ \Sigma_{ij} = \begin{cases} c_i^{-1}\gamma_i^2, & i = j \in \mathbb{I}_{\mathcal{C}}, \\ c_i^{-1}\frac{\pi_j}{\max(\pi_i,\pi_j)} R_{ij}\left(c_i/c_j, \pi_i/\pi_j\right)\gamma_i\gamma_j, & i < j \in \mathbb{I}_{\mathcal{C}}. \end{cases}$$

Step 2. Estimate the asymptotic covariance matrix.

For constructing the working likelihood only, we use the bias-negligible approximation $\lambda_j = 0$ for any $j \in \mathbb{I}_{\mathcal{C}}$. We approximate $\widehat{\boldsymbol{\gamma}}_c$ by a multivariate normal distribution $\mathcal{N}(\boldsymbol{\gamma}_1, k^{-1}\widehat{\Sigma}_c)$, where $\boldsymbol{\gamma}_1$ is the $N_{\mathcal{C}}$-dimensional vector whose entries are all equal to γ_1. Moreover, $\widehat{\Sigma}_c := \left(\widehat{\Sigma}_{ij}\right)_{i,j\in\mathbb{I}_{\mathcal{C}}}$ denotes the estimated covariance matrix with

$$\widehat{\Sigma}_{ij} = \begin{cases} (k/k_i)\widehat{\gamma}_i^2, & i = j \\ (k/k_i)\frac{n_j}{\max(n_i,n_j)}\widehat{R}_{ij}(k_i/k_j, n_i/n_j)\widehat{\gamma}_i\widehat{\gamma}_j, & i < j. \end{cases}$$

and $\widehat{\Sigma}_{ji} = \widehat{\Sigma}_{ij}$ for $i < j$. Here, $\widehat{R}(\cdot,\cdot)$ denotes an estimator of the tail dependence function $R(\cdot,\cdot)$, which is obtained using the standard approach (Drees and Huang, 1998), by

$$\widehat{R}(x,y) = \frac{1}{k^*}\sum_{i=1}^{n^*} I(X_i \geq X_{n^*-[k^*x]+1}, Y_i \geq Y_{n^*-[k^*y]+1}), \quad x,y \geq 0, \tag{5.12}$$

where $n^* = \min\{n_i, n_j\}$ and $k^* = k_i$ if $n_i < n_j$, $k^* = k_j$ otherwise.

Step 3. Obtain the target estimator.

Let $V_c = \Sigma_c^{-1}$ and $\widehat{V}_c = \widehat{\Sigma}_c^{-1}$. Maximizing the approximate likelihood of the single observation $\widehat{\boldsymbol{\gamma}}_c$ with respect to γ_1 yields the following oracle estimator:

$$\widetilde{\gamma}_1^{(o)} = \frac{\sum_{j\in\mathbb{I}_{\mathcal{C}}}\left(\sum_{i\in\mathbb{I}_{\mathcal{C}}}\widehat{V}_{ij}\right)\widehat{\gamma}_j}{\sum\sum_{s,t\in\mathbb{I}_{\mathcal{C}}}\widehat{V}_{st}} = \frac{\mathbf{1}^\top\widehat{V}_c\widehat{\boldsymbol{\gamma}}_c}{\mathbf{1}^\top\widehat{V}_c\mathbf{1}}.$$

Similar to $\widehat{\gamma}_1^{(o)}$ in Eq.(4.7), $\widetilde{\gamma}_1^{(o)}$ also can be rewritten as $\widetilde{\gamma}_1^{(o)} = \sum_{j\in\mathbb{I}_{\mathcal{C}}}\zeta_j^{(o)}\widehat{\gamma}_j$ with

$$\zeta_j^{(o)} = \frac{\sum_{i\in\mathbb{I}_{\mathcal{C}}}\widehat{V}_{ij}}{\sum\sum_{s,t\in\mathbb{I}_{\mathcal{C}}}\widehat{V}_{st}} = \frac{\sum_{i\in\mathbb{I}_{\mathcal{C}}}\widehat{V}_{ij}}{\mathbf{1}^\top\widehat{V}_c\mathbf{1}}.$$

Moreover, there exists some link between the two estimators. Under the independent data setup described in Section 3.1, the covariance matrix Σ_c will reduce to a diagonal matrix, which

directly leads to the reduction of $\zeta_j^{(o)}$ to $\eta_j^{(o)}$, and finally, the estimator $\widetilde{\gamma}_1^{(o)}$ corresponds to the estimator $\widehat{\gamma}_1^{(o)}$. Back to the dependent case, we can find that the estimator $\widetilde{\gamma}_1^{(o)}$ incorporates the tail dependence structure through the estimator of R-function. The following theorem elucidates the proposed strategy by establishing the asymptotic result for the estimator.

Theorem 5. *Suppose that the set $\mathbb{I}_{\mathcal{C}}$ is known, the condition $\mathbf{C}(R)$ holds for the dataset $\mathbb{S}$, and the marginal distribution F_j satisfies condition $\mathbf{C}_2(\gamma_j, \rho_j, A_j)$ for any $j \in \mathbb{I}_{\mathcal{C}}$. Then, under condition $\mathbf{C}(n,k)$,*

$$\sqrt{k}(\widetilde{\gamma}_1^{(o)} - \gamma_1) \xrightarrow{d} N\big(\nu_1^{(o)}, \Sigma_1^{(o)}\big) \text{ with } \nu_1^{(o)} = \frac{\mathbf{1}^\top V_c \boldsymbol{\mu}_c}{\mathbf{1}^\top V_c \mathbf{1}},\ \Sigma_1^{(o)} = \frac{1}{\mathbf{1}^\top V_c \mathbf{1}}.$$

This result generalizes Theorem 2 to the tail-dependent setting. Note that $\widetilde{\gamma}_1^{(o)}$ reduces to $\widehat{\gamma}_1^{(o)}$ when $R_{ij}(\cdot,\cdot) \equiv 0$, and, correspondingly, the asymptotic variance $\Sigma_1^{(o)}$ simplifies to $\Delta_1^{(o)}$ from Theorem 2 in this case, which is straightforward to verify. It is also noteworthy that, like Eq.(4.7), $\widetilde{\gamma}_1^{(o)}$ can be expressed as a pooling estimator with weight vector $\boldsymbol{\zeta} = \widehat{V}_c\mathbf{1}/\mathbf{1}^\top\widehat{V}_c\mathbf{1}$. This matches the estimated version of variance-optimal weight vector $\widehat{\omega}^{(\mathrm{Var})}$ proposed in Daouia et al. (2024), although our estimation strategies are different from that work. Most importantly, it highlights that the oracle estimator $\widetilde{\gamma}_1^{(o)}$ outperforms both the classical Hill estimator $\widehat{\gamma}_1$ with weight vector $\boldsymbol{\zeta}_{\mathrm{H}} = (1, 0, \cdots, 0)^\top$, and the distributed Hill estimator $\sum_{j\in\mathbb{I}_{\mathcal{C}}} \widehat{\gamma}_j / N_{\mathcal{C}}$ with weight vector $\boldsymbol{\zeta}_{\mathrm{DH}} = (1/N_{\mathcal{C}}, \cdots, 1/N_{\mathcal{C}})^\top$.

Let the set $\mathbb{I}_{\mathcal{BC}} = \mathbb{I}_{\mathcal{B}} \cup \mathbb{I}_{\mathcal{C}}$. Heuristically, by combining the oracle estimator $\widetilde{\gamma}_1^{(o)}$ in tail-dependent case and the estimator $\widehat{\gamma}_1^{(c)}$ defined in Eq.(3.5), we propose an iFusion estimator that incorporates the weights adapted to the tail-dependent setting:

$$\widetilde{\gamma}_1^{(c)} = \frac{\sum_{j\in\mathbb{I}_{\mathcal{BC}}} \big(\sum_{i\in\mathbb{I}_{\mathcal{BC}}} \widehat{V}_{ij}\big)\widehat{\gamma}_j}{\sum\sum_{s,t\in\mathbb{I}_{\mathcal{BC}}} \widehat{V}_{st}} = \frac{\mathbf{1}^\top \widehat{V}_{bc}\widehat{\boldsymbol{\gamma}}_{bc}}{\mathbf{1}^\top \widehat{V}_{bc}\mathbf{1}},$$

where $\widehat{\boldsymbol{\gamma}}_{bc} = \big(\widehat{\gamma}_j\big)_{j\in\mathbb{I}_{\mathcal{BC}}}^\top$.

Similar to Theorems 3 and 4, we consider the cases that (i) the membership of $\mathcal{C}$ is unknown but the separation condition holds; (ii) the membership of $\mathcal{C}$ is unknown and the separation

condition does not hold. The following corollary presents the related results.

Corollary 1. *Suppose that the condition* $\mathbf{C}(R)$ *holds for the dataset* $\mathbb{S}$*, and the marginal distribution* F_j *satisfies condition* $\mathbf{C}_2(\gamma_j, \rho_j, A_j)$ *for any* $j \in \{1, \cdots, N\}$*. The following results hold.*

(i) If the separation condition $\mathbf{C}(d)$ *holds and the screening weights satisfy condition* $\mathbf{C}_1(w)$*, then, under condition* $\mathbf{C}(n,k)$*,*

$$\sqrt{k}\big(\widetilde{\gamma}_1^{(c)} - \gamma_1\big) \xrightarrow{d} N\Big(\nu_1^{(o)}, \Sigma_1^{(o)}\Big),$$

where $\nu_1^{(o)}$ *and* $\Sigma_1^{(o)}$ *are defined in Theorem 5.*

(ii) If the separation condition $\mathbf{C}(d)$ *does not hold, the screening weights satisfy condition* $\mathbf{C}_2(w)$*, for* $j = 1, \ldots, N$*. Then, under condition* $\mathbf{C}(n,k)$*,*

$$\sqrt{k}\big(\widetilde{\gamma}_1^{(c)} - \gamma_1\big) - T_1 \xrightarrow{d} N\Big(\nu_1, \Sigma_1\Big), \text{ where } T_1 = \frac{\mathbf{1}^\top V_{bc}\mathbf{B}_{bc}}{\mathbf{1}^\top V_{bc}\mathbf{1}}, \quad \nu_1 = \frac{\mathbf{1}^\top V_{bc}\boldsymbol{\mu}_{bc}}{\mathbf{1}^\top V_{bc}\mathbf{1}}, \quad \Sigma_1 = \frac{1}{\mathbf{1}^\top V_{bc}\mathbf{1}}.$$

Here, $\boldsymbol{\mu}_{bc} = \left(\frac{\sqrt{c_j^{-1}}\lambda_j}{1-\rho_j}\right)_{j\in\mathbb{I}_{\mathcal{BC}}}$ *and* $\boldsymbol{B}_{bc} = \left(B_j^\dagger\right)_{j\in\mathbb{I}_{\mathcal{BC}}}$*, with* $B_j^\dagger = \begin{cases} 0, & j \in \mathbb{I}_{\mathcal{C}}, \\ B_j, & j \in \mathbb{I}_{\mathcal{B}}. \end{cases}$

6. Simulation Study

In this section, we conduct simulation studies to evaluate the finite-sample performance of the iFusion method under both independent and tail-dependent cases. To demonstrate the advantages of the proposed method, we compare its performance with several competing estimators in both simulations and the real-data analysis. We first describe the different competitors in the following simple example.

Example 2. Consider the case where the dataset $\mathbb{S}$ consists of five data sources, i.e., $N = 5$. For the j-th individual, let γ_j and $\widehat{\gamma}_j$ denote the true EVI and its corresponding estimator, respectively. Assume that the index sets are given by $\mathbb{I}_{\mathcal{C}} = \{1,2,3\}, \mathbb{I}_{\mathcal{B}} = \{4\}$ and $\mathbb{I}_{\mathcal{D}} = \{5\}$. Then the estimators based on various methods are summarized in the following table.

Table 2: Estimation methods for target extreme value index γ_1

Methods		Estimator	Weights					Asymptotic Variance
			$\widehat{\gamma}_1$	$\widehat{\gamma}_2$	$\widehat{\gamma}_3$	$\widehat{\gamma}_4$	$\widehat{\gamma}_5$	
Classical Hill		$\widehat{\gamma}_H = \widehat{\gamma}_1$	1	0	0	0	0	γ_1^2/k_1
Distributed_O		$\widehat{\gamma}_{\mathrm{D}}^{(o)} = \sum_{j\in\mathbb{I}_{\mathcal{C}}} \frac{1}{3}\widehat{\gamma}_j$	1/3	1/3	1/3	0	0	$\sum_{j\in\mathbb{I}_{\mathcal{C}}} \frac{\gamma_j^2}{k_j}/N_{\mathcal{C}}^2$
Distributed_All		$\widehat{\gamma}_{\mathrm{D}} = \sum_{j=1}^{N} \frac{1}{5}\widehat{\gamma}_j$	1/5	1/5	1/5	1/5	1/5	$\sum_{j=1}^{5} \frac{\gamma_j^2}{k_j}/N^2$
iFusion_O	ind.	$\widehat{\gamma}_1^{(o)} = \sum_{j\in\mathbb{I}_{\mathcal{C}}} \eta_j^{(o)}\widehat{\gamma}_j$	$\eta_1^{(o)}$	$\eta_2^{(o)}$	$\eta_3^{(o)}$	0	0	$\Delta_1^{(o)}/k$
	dep.	$\widetilde{\gamma}_1^{(o)} = \sum_{j\in\mathbb{I}_{\mathcal{C}}} \zeta_j^{(o)}\widehat{\gamma}_j$	$\zeta_1^{(o)}$	$\zeta_2^{(o)}$	$\zeta_3^{(o)}$	0	0	$\Sigma_1^{(o)}/k$
iFusion	ind.	$\widehat{\gamma}_1^{(c)} = \sum_{j=1}^{N} \eta_j^{(c)}\widehat{\gamma}_j$	$\eta_1^{(c)}$	$\eta_2^{(c)}$	$\eta_3^{(c)}$	$\eta_4^{(c)}$	$\eta_5^{(c)}$	Δ_1/k
	dep.	$\widetilde{\gamma}_1^{(c)} = \sum_{j\in\mathbb{I}_{\mathcal{BC}}} \zeta_j^{(c)}\widehat{\gamma}_j$	$\zeta_1^{(c)}$	$\zeta_2^{(c)}$	$\zeta_3^{(c)}$	$\zeta_4^{(c)}$	$\zeta_5^{(c)}$	Σ_1/k

Notes: For these notations, refer to the previous sections.

6.1 The independent case

We consider two distributions, the Fréchet and the absolute Student-t distribution, which both belong to the MDA of Fréchet distribution and satisfy the second-order condition. The simulation studies on the absolute Student-t distribution are deferred to Section S3 in Supplementary Material.

- Fréchet(α) : $F(x) = \exp(-x^{-\alpha})$, $x > 0$ with the EVI $\gamma = 1/\alpha$ and second-order parameter $\rho = -1$.

Then we consider the data-generating process as follows. For each setting, we repeat the simulation 1000 times for $N = 20$ data sources. The full sample size $n = 10000$ and the total effective sample size $k = 10\% \cdot n = 1000$. Specifically, we consider the sample allocation scheme:

Balanced: $n_1 = \cdots = n_N = 500$ and $k_1 = \cdots = k_N = 50$.

Unbalanced: $(n_1, \ldots, n_N) = (200, 300, 400, 400, 500, 500, 600, 600, 700, 800, 200, 400, 500, 600, 800, 300, 400, 500, 600, 700)$ and $(k_1, \cdots, k_N) = 10\% \cdot (n_1, \ldots, n_N)$.

Case I: Generate random data: $X_\ell^{(j)} \sim$ Fréchet(α_j), for $\ell = 1, \ldots, n_j$, $j = 1, \cdots, N$, where

$\alpha_j = 1/\gamma_j$. Assume that $U_j \overset{iid}{\sim} \text{Unif}(1/2, 1)$ and γ_j take values as follows: (i) $\gamma_1 = \gamma^\star$ and $\gamma_j = \gamma^\star + U_j/k^2$ for $j = 2, \ldots, 10$, this forms the clique group, whose parameters are asymptotically close to the target; (ii) $\gamma_j = \gamma^\star + U_j/\sqrt{k}$ for $j = 11, \ldots, 15$; (iii) $\gamma_j = \gamma^\star + U_j/k^{1/10}$ for $j = 16, \ldots, 20$, where $\gamma^\star \in \{0.4,\ 0.6,\ 0.8,\ 1.0\}$. Hence, the group sizes are $N_{\mathcal{C}} = 10, N_{\mathcal{B}} = N_{\mathcal{D}} = 5$.

Our goal is to estimate the EVI for individual-1, i.e., γ_1. We implement four estimators, (i) the initial estimator $\widehat{\gamma}_1$ (denoted as Class._Hill in the table); (ii) the distributed Hill estimator (including Distri._O $\widehat{\gamma}_{\text{D}}^{(o)}$ and Distri._All $\widehat{\gamma}_{\text{D}}$); (iii) the oracle estimator $\widehat{\gamma}_1^{(o)}$ (i.e., iFusion_O); (iv) the iFusion estimator $\widehat{\gamma}_1^{(c)}$, further categorized into iFusion_C when $\mathcal{B} = \varnothing$ and iFusion_BC otherwise. For the screening weights $w_{1j}, j = 1, \ldots, N$, used in the iFusion estimation, we adopt the uniform kernel-based form introduced in Section S1 in the Supplementary Material.

The results are summarized in Tables 3-4, which report the bias, variance (Var), as well as mean squared error (MSE) of the corresponding estimators. The main findings are outlined as follows.

Table 3: Estimation results of γ_1 for the Fréchet distributions with balanced samples

	$\gamma^\star$	Class._Hill	Distri._All	Distri._O	iFusion_O	iFusion_C	iFusion_BC
BIAS	0.4	0.0102	0.1131	0.0103	-0.0042	-0.0001	0.0055
	0.6	0.0153	0.1185	0.0155	-0.0063	0.0060	0.0087
	0.8	0.0203	0.1239	0.0206	-0.0083	0.0135	0.0137
	1.0	0.0254	0.1294	0.0258	-0.0104	0.0227	0.0205
Var ($\times 10^{-2}$)	0.4	0.3241	0.0356	0.0322	0.0347	0.0481	0.0341
	0.6	0.7293	0.0606	0.0724	0.0780	0.1802	0.1291
	0.8	1.2965	0.0939	0.1287	0.1386	0.4429	0.3488
	1.0	2.0258	0.1356	0.2010	0.2166	0.7217	0.5849
MSE ($\times 10^{-2}$)	0.4	0.3345	1.3146	0.0428	0.0364	0.0481	0.0370
	0.6	0.7526	1.4653	0.0963	0.0819	0.1838	0.1367
	0.8	1.3379	1.6302	0.1713	0.1456	0.4613	0.3676
	1.0	2.0905	1.8093	0.2676	0.2275	0.7734	0.6271

Notes: $\gamma^\star$, the true value of the target parameter.

- We first analyze the results of the bias. Comparing the classical and distributed estimators shows that the distributed Hill estimators (Distri._All and even Distri._O) fail to serve the purpose of bias mitigation. In contrast, the oracle estimator (iFusion_O)

Table 4: Estimation results of γ_1 for the Fréchet distribution with unbalanced samples

	$\gamma^\star$	Class._Hill	Distri._All	Distri._O	iFusion_O	iFusion_C	iFusion_BC
BIAS	0.4	0.0070	0.1126	0.0096	-0.0052	0.0047	0.0090
	0.6	0.0104	0.1178	0.0144	-0.0078	0.0137	0.0153
	0.8	0.0139	0.1230	0.0193	-0.0104	0.0227	0.0218
	1.0	0.0174	0.1283	0.0241	-0.0131	0.0295	0.0275
Var ($\times 10^{-2}$)	0.4	0.8739	0.0401	0.0371	0.0355	0.1906	0.1442
	0.6	1.9662	0.0684	0.0834	0.0799	0.6732	0.5648
	0.8	3.4955	0.1060	0.1482	0.1421	1.5099	1.3162
	1.0	5.4617	0.1529	0.2316	0.2221	2.3305	2.0672
MSE ($\times 10^{-2}$)	0.4	0.8787	1.3075	0.0463	0.0383	0.1928	0.1523
	0.6	1.9771	1.4561	0.1042	0.0861	0.6919	0.5883
	0.8	3.5149	1.6196	0.1853	0.1530	1.5615	1.3638
	1.0	5.4920	1.7980	0.2896	0.2391	2.4180	2.1429

demonstrates superior performance. Compared to the classical Hill estimator, the iFusion estimators (iFusion_C and iFusion_BC) have a slight advantage in terms of bias results under the balanced case.

- Secondly, we discuss the results of variance (Var). The results show that the distributed Hill estimator, as well as the estimators under the iFusion method, are significantly better than the classical Hill estimator. This phenomenon reflects the necessity of the idea of data fusion. The Distri._All, which averages the estimators from all individuals, exhibits the smallest variance. Next, we explore the results under two scenarios. (i) For the case of balanced sample sizes among multiple data sources (see Table 3), the variance of the Distri._O as well as iFusion_O is close, although the latter is slightly larger. It is interesting to note that, as expected, the variance is approximately one-tenth that of the classical Hill estimator, consistent with the fact that $N_{\mathcal{C}} = 10$ and aligns with the discussion in Remark 2. In addition, the relationship between the size of the variance terms of the estimators under iFusion roughly follows the order iFusion_O < iFusion_BC < iFusion_C. This can be explained by the fact that the oracle estimator only considers the nearby data sources. As the number of fused data sources increases, the variance decreases to some extent, albeit at the cost of increased bias. (ii) For the

case of unbalanced sample sizes (see Table 4), the classical Hill estimator performs worst in terms of variance. The variance of the Distri._O and iFusion_O is close to each other, and the latter is smaller.

- Finally, we report the MSE for the estimators to provide a comprehensive evaluation. In most rows, the oracle estimator (iFusion_O) achieves the smallest MSE, with the Distri._O performing comparably. The iFusion_C and iFusion_BC estimators, while slightly inferior to these two, consistently outperform the classical Hill estimator. This suggests that the iFusion learning method improves estimation accuracy. The oracle estimator outperforms other estimators in terms of small MSE across different cases, especially for relatively unbalanced samples. The improvements of the oracle and iFusion estimators over the classical Hill estimator are more significant under unbalanced sample settings, which is reasonable. Further discussion can be found in Section S3.1 of the Supplementary Material.

6.2 The tail-dependent case

To study the performance of iFusion under a tail-dependent setting described in Section 5, we generate data from the multivariate distributions with heavy-tailed marginals. Here, we present the simulation experiments for Multivariate Burr distributions.

- Multivariate Burr distribution $\mathcal{MB}(\alpha, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma})$ with survival function

$$\overline{F}(x_1, \ldots, x_d) = \left[1 + \sum_{i=1}^{d} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^{1/\beta_i}\right]^{-\alpha}, x_i > \mu_i, i = 1, 2, \ldots, d$$

with marginal extreme value indices $\gamma_i = \beta_i/\alpha, i = 1, 2, \ldots, d$.

Similar to the independent case in the previous subsection, we consider the following setting for the above multivariate distributions.

Case II: Generate random data: $(X_1, X_2, \ldots, X_{20}) \sim \mathcal{MB}(\alpha, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma})$ with $\alpha = 1, \boldsymbol{\beta} = (\gamma_1, \gamma_2, \cdots, \gamma_d), \boldsymbol{\mu} = \mathbf{0}_d$, and $\boldsymbol{\sigma} = \mathbf{1}_d$, $d = 20$. Assume that $U_j \overset{iid}{\sim} \text{Unif}(1/2, 1)$ and γ_j take values as follows: (i)$\gamma_1 = \gamma^\star$

and $\gamma_j = \gamma^\star + U_j/k^2$ for $j = 2, \ldots, 10$, this forms the clique group, whose parameters are asymptotically close to the target; (ii) $\gamma_j = \gamma^\star + U_j/\sqrt{k}$ for $j = 11, \ldots, 15$; (iii) $\gamma_j = \gamma^\star + U_j/k^{1/10}$ for $j = 16, \ldots, 20$, where $\gamma^\star \in \{0.4,\ 0.6,\ 0.8,\ 1.0\}$.

Regarding Case II, we design comparisons from two perspectives. Of course, the first one is the comparison of estimation methods, like that for the independent cases in Section 6.1. The second is a comparison in terms of whether dependent structures are considered when implementing the iFusion method. It is worth mentioning that incorporating the dependence structure into the estimation process requires sharing the original data among all individuals. When the entire dataset $\mathbb{S}$ is available, it is not a bad idea to first perform data fusion before implementing Hill estimation. Here, we refer to such a method as a benchmark and add it to the list of comparisons.

The simulation results are presented in Tables 5-6. Note that a modification is made here, i.e., only the data sources corresponding to the subscript $j \in \mathbb{I}_{\mathcal{C}}$ are fused to obtain the benchmark Hill estimator (denoted as Bench._O). The setting of Case II is similar to Case I in Section 6.1. We can draw the following conclusions.

- Likewise, the bias under each method is compared first. The results are similar for both the balanced and unbalanced sample settings. With the exception of the Distri._All and iFusion_BC (ind.), the bias of the other estimators is smaller than that of the classical Hill estimator. The bias of the Distri._O is close to that of the Bench._O estimator, although the former is slightly larger. However, neither of these two takes into account the dependence of different data sources. The independent-version iFusion estimators, especially iFusion_C (ind.) and iFusion_BC (ind.), ignore the dependence among source-specific estimates and treat their estimation errors as independent. As a result, their absolute deviations may sometimes exceed those of the original estimator. This indicates that ignoring the dependence structure may adversely affect estimation accuracy. Nevertheless, after considering the dependency structure, the bias tends to be reduced.

Table 5: Estimation results of γ_1 for Multivariate Burr distributions with balanced samples

	$\gamma^\star$	Class._Hill	Bench._O	Distri._O	Distri._All	iFusion_O		iFusion_C		iFusion_BC	
						ind.	dep.	ind.	dep.	ind.	dep.
BIAS	0.4	0.0211	0.0178	0.0198	0.1247	0.0123	0.0094	0.0159	0.0106	0.0222	0.0065
	0.6	0.0316	0.0266	0.0298	0.1348	0.0184	0.0141	0.0296	0.0195	0.0337	0.0030
	0.8	0.0421	0.0355	0.0397	0.1449	0.0246	0.0187	0.0449	0.0301	0.0464	-0.0032
	1.0	0.0526	0.0444	0.0496	0.1550	0.0307	0.0234	0.0590	0.0419	0.0580	-0.0021
Var ($\times 10^{-2}$)	0.4	0.3502	0.1708	0.1724	0.2600	0.1729	0.1772	0.1690	0.1771	0.1682	0.1935
	0.6	0.7880	0.3843	0.3880	0.5029	0.3891	0.3987	0.3984	0.4112	0.3849	0.4614
	0.8	1.4009	0.6832	0.6898	0.8271	0.6917	0.7088	0.7456	0.7659	0.7093	0.9366
	1.0	2.1890	1.0675	1.0778	1.2326	1.0807	1.1074	1.2067	1.2528	1.1320	1.4573
MSE ($\times 10^{-2}$)	0.4	0.3946	0.2024	0.2118	1.8157	0.1880	0.1860	0.1942	0.1882	0.2177	0.1977
	0.6	0.8877	0.4553	0.4766	2.3201	0.4230	0.4185	0.4857	0.4492	0.4986	0.4623
	0.8	1.5782	0.8094	0.8473	2.9260	0.7519	0.7439	0.9473	0.8564	0.9245	0.9376
	1.0	2.4659	1.2647	1.3238	3.6336	1.1749	1.1624	1.5550	1.4287	1.4688	1.4577

Table 6: Estimation results of γ_1 for Multivariate Burr distributions with unbalanced samples

	$\gamma^\star$	Class._Hill	Bench._O	Distri._O	Distri._All	iFusion_O		iFusion_C		iFusion_BC	
						ind.	dep.	ind.	dep.	ind.	dep.
BIAS	0.4	0.0193	0.0189	0.0201	0.1244	0.0111	0.0094	0.0175	0.0117	0.0222	0.0048
	0.6	0.0290	0.0284	0.0302	0.1344	0.0167	0.0141	0.0336	0.0238	0.0353	0.0076
	0.8	0.0387	0.0378	0.0402	0.1444	0.0222	0.0188	0.0514	0.0385	0.0494	0.0120
	1.0	0.0483	0.0473	0.0503	0.1543	0.0278	0.0235	0.0644	0.0515	0.0608	0.0215
Var ($\times 10^{-2}$)	0.4	0.8302	0.1375	0.1527	0.2326	0.1380	0.1408	0.1961	0.1865	0.1654	0.1801
	0.6	1.8679	0.3093	0.3436	0.4488	0.3104	0.3169	0.5761	0.5541	0.5049	0.5745
	0.8	3.3207	0.5499	0.6108	0.7374	0.5518	0.5634	1.2355	1.1864	1.0637	1.1629
	1.0	5.1886	0.8592	0.9544	1.0984	0.8622	0.8803	2.1841	2.0816	1.9184	1.9767
MSE ($\times 10^{-2}$)	0.4	0.8676	0.1732	0.1931	1.7799	0.1503	0.1497	0.2268	0.2000	0.2149	0.1824
	0.6	1.9520	0.3897	0.4345	2.2544	0.3382	0.3367	0.6870	0.6110	0.6293	0.5802
	0.8	3.4703	0.6929	0.7725	2.8213	0.6013	0.5987	1.4993	1.3346	1.3080	1.1772
	1.0	5.4223	1.0826	1.2070	3.4804	0.9395	0.9354	2.5990	2.3466	2.2884	2.0227

- For both Var and MSE, it is clear that the classical Hill estimator and Distri._All yield the larger values, while the variances of the Bench._O, Distri._O, and iFusion_O estimators (including ind. and dep.) are comparable. Additionally, in contrast to the results of bias, the variance of the iFusion estimator, which accounts for dependencies, is larger than that of the independent treatment. This difference is related to the specific form of the estimator. The MSE results highlight the importance of considering dependencies when using the iFusion method, even though this approach may increase the variance.
- Finally, if the original observations of multiple data sources are available, the study of tail dependence structures is indispensable. This can be verified by comparing the performance of the benchmark and the oracle estimators. On the contrary, if only summary statistics are available, such as the specific estimate and the corresponding effective sample size, the benchmark estimator and the dependence-adjusted iFusion estimators are not applicable. But the initial estimator is less effective than the distributed Hill estimator and the iFusion estimators without dependence information. In this case, we recommend the independent-version iFusion estimator because it achieves a smaller MSE.

7. Real Data Analysis

As a major air pollutant, surface ozone (O_3) refers to ozone found near the Earth's surface. Excessive surface ozone concentrations can endanger human health and negatively affect agriculture, natural ecosystems, and the global climate. A detailed analysis of the extreme behavior of surface ozone concentration is essential for forecasting high ozone levels and issuing timely public health warnings. In this section, we collect the daily maximum concentrations of surface ozone in London from the UK Air Information Resource (https://uk-air.defra.gov.uk/). The dataset covers the period from January 1, 2020, to December 31, 2023.

We apply our methodology to the London area, focusing on six air-quality monitoring sites located within its boundaries. These sites, labeled S1 through S6, are indicated with pins

on the map in Figure 1. Detailed information about the monitoring sites is summarized in Table 7. Notice that the sample sizes across the sites are unbalanced due to differences in the establishment dates of the monitoring stations.

Figure 1: Map of London along with its air-quality monitoring sites.

Table 7: Summary of information for air-quality monitoring sites in London

Site Name	Label	Latitude	Longitude	Observation Size
London Bloomsbury	S1	51.52229	-0.12589	1396
London Eltham	S2	51.45258	0.07077	791
London Harlington	S3	51.48879	-0.44161	1432
London Hillingdon	S4	51.49633	-0.46086	1401
London Marylebone Road	S5	51.52253	-0.15461	1370
London Westminster	S6	51.49467	-0.13193	319

As a preliminary step, we analyze the distribution of the daily maximum ozone concentrations at each site through histograms, as shown in Figure 2. The histograms across different sites display a certain degree of similarity, while also exhibiting a pronounced peak and heavy tails, indicating the presence of extremely high values in the distributions. It highlights the importance of studying the EVI, which provides a quantitative measure to capture the tail behavior of the ozone concentration distributions. Accordingly, we proceed to estimate the EVI for the six monitoring sites using the proposed methodology.

Notably, unlike the simulation settings, certain information is unavailable in real data ap-

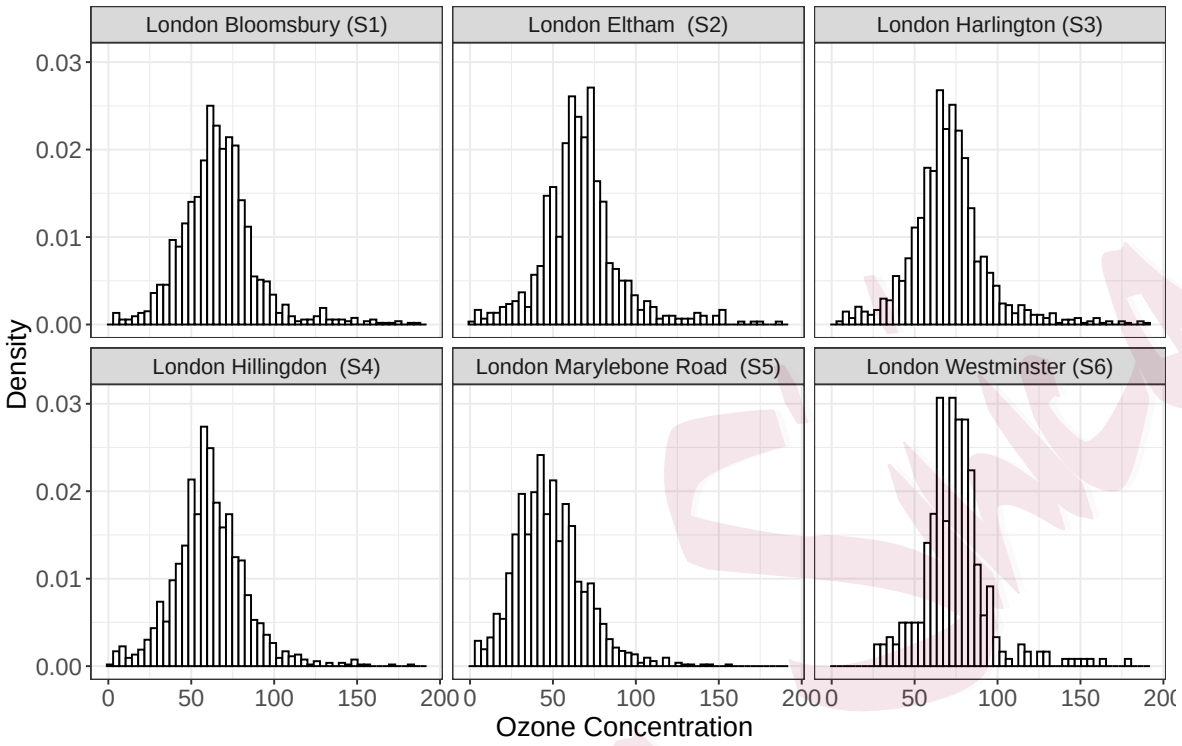

Figure 2: Histogram of the daily maximum ozone concentration for each site.

plications. For example, the index sets $\mathbb{I}_{\mathcal{C}}$, $\mathbb{I}_{\mathcal{B}}$, and $\mathbb{I}_{\mathcal{D}}$ are unknown, and the separation condition $\mathbf{C}(d)$ cannot be verified. Consequently, we do not impose any explicit subdivision when implementing the iFusion estimators. Instead, the estimator adaptively calibrates the contribution of each data source through a fully data-driven weighting scheme.

Table 8 summarizes the estimation results of the EVI of each site via different methods. The Class._Hill method consistently yields the highest variance across sites, whereas the Distri._All method produces the lowest. Although the latter achieves lower variance by aggregating data from all sites, this does not necessarily imply superior estimation performance, as it results in identical estimates for all sites and ignores local heterogeneity. In contrast, the Distri._C and iFusion methods incorporate information from similar sites to estimate the EVI for each location. The results of the screening weight vector $\boldsymbol{w}$ in Table 8 indicate which sites are selected by the two methods to contribute to the EVI estimation for each target site. Although both methods utilize the same set of data sources, the iFusion method yields estimates with smaller

variances due to its more refined, data-adaptive weighting scheme. These results confirm the advantage of the iFusion approach in achieving more stable and precise estimation.

Table 8: Estimation of γ by all methods, with the corresponding variances shown in brackets ($\times 10^{-3}$).

Site	Class._Hill	Distri._All	Distri._C	iFusion	Screening Weight Vector($\boldsymbol{w}$)
S1	0.2277 (0.7428)	0.1916 (0.1507)	0.2203 (0.4726)	0.2219 (0.4509)	$(1,1,0,0,0,0)$
S2	0.2130 (1.1474)	0.1916 (0.1507)	0.2203 (0.4726)	0.2219 (0.4509)	$(1,1,0,0,0,0)$
S3	0.1827 (0.4663)	0.1916 (0.1507)	0.1818 (0.3495)	0.1789 (0.2033)	$(0,0,1,1,0,1)$
S4	0.1733 (0.4291)	0.1916 (0.1507)	0.1772 (0.2209)	0.1735 (0.1335)	$(0,0,1,1,1,1)$
S5	0.1632 (0.3890)	0.1916 (0.1507)	0.1683 (0.2045)	0.1680 (0.2041)	$(0,0,0,1,1,0)$
S6	0.1894 (2.2501)	0.1916 (0.1507)	0.1818 (0.3495)	0.1789 (0.2033)	$(0,0,1,1,0,1)$

Notes: Distri._All averages the estimators from all sites, while Distri._C averages those from sites selected via screening weights.

To provide a more intuitive illustration of the asymptotic variance under the different methods, we take site S1 as an example and investigate the effect of the deletion ratio on the estimated variance by randomly deleting the observations from site S1. Specifically, for a given deletion ratio, we randomly remove a corresponding proportion of observations from the original dataset and repeat this process for 100 independent replications. For each replication, we compute the theoretical asymptotic variance for each competing estimator based on the remaining observations; the average of these 100 asymptotic variances is then used to quantify the estimation precision in the sparse-data setting. For each deletion, the effective sample size k_1 is set to a fixed proportion (10% in our study) of the remaining sample size, and the screening weights are calculated using the kernel-based approach detailed in Section S1 of the Supplementary Material. The left panel in Figure 3 suggests that as the deletion ratio increases, the variance of the Class._Hill method increases sharply and remains the highest among all methods. In contrast, the estimated variance of the Distri._All method increases slowly and remains the lowest. The variances of the Distri._C and iFusion estimators follow

similar trends, with the iFusion method consistently yielding lower variance. The results align well with the theoretical results of this article. To evaluate the generalizability of the observed stability, we extend the sensitivity analysis to the remaining five monitoring sites (S2–S6). The results for sites S2–S6 (detailed in Figure S4.6 of the Supplementary Material) consistently mirror the findings at site S1. Specifically, while the variance of the Class._Hill estimator exhibits an exponential-like escalation as data becomes sparser, the iFusion estimator maintains a remarkably stable and lower variance profile, second only to the Distri._All method, which sacrifices local heterogeneity. This robust performance across all sites reinforces that iFusion effectively borrows strength from auxiliary sources to mitigate the instability inherent in sparse-data estimation.

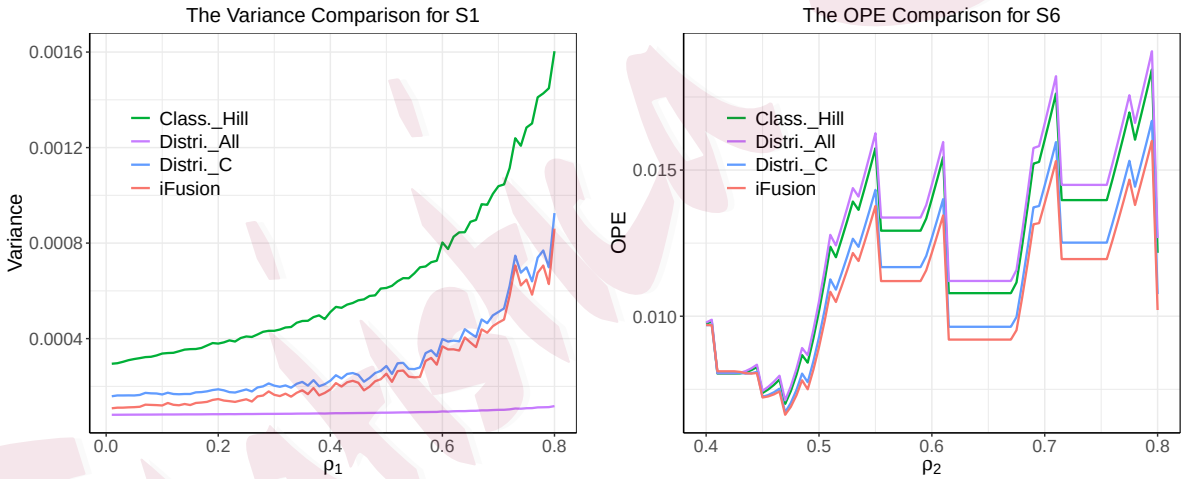

Figure 3: Left panel: variance comparison, where ρ_1 denotes the deletion ratio. Right panel: out-of-sample error comparison, where ρ_2 denotes the fraction of the sample used for testing.

It is worth noting that site S6 has relatively few observations due to its closure during 2021 and 2022. In such cases, the iFusion approach may enhance estimation performance by effectively leveraging information from other sites. To further validate the predictive performance, we conduct an out-of-sample forecasting study to compare the performance of various methods. Specifically, the last 50 observations from site S6, denoted by $\{Y_i^\star\}_{i=1}^{50}$, are reserved for out-of-sample testing. To evaluate prediction quality, we adopt the out-of-sample prediction error (OPE) metric proposed by Wang and Tsai (2009). In the current setting, the OPE is

defined as follows:

$$\mathrm{OPE}(\rho_2) = \Big(\sum_{i=1}^{50} \mathrm{I}\big(Y_i^\star > \tilde{\omega}_N\big)\Big)^{-1} \times \left(\sum_{i=1}^{50} \big\{\log(Y_i^\star/\tilde{\omega}_N) - \widehat{\gamma}\big\}^2 \cdot \mathrm{I}\big(Y_i^\star > \tilde{\omega}_N\big)\right),$$

where $\tilde{\omega}_N$ is a pre-specified cut-off value to differentiate the normal and extreme situations, $\rho_2 = \sum_{i=1}^{50} \mathrm{I}\big(Y_i^\star > \tilde{\omega}_N\big)/50$ is the sample fraction used in the test data, and $\mathrm{I}(\cdot)$ is the indicator function. The right panel in Figure 3 shows that the OPE of the iFusion estimator is consistently lower than those of the other estimators across all values of ρ_2, indicating that the iFusion method improves the estimation of the target EVI. Additional out-of-sample analyses for high-quantile forecasting, evaluated using the quantile loss (QL) function, are reported in Section S4.1 of the Supplementary Material; the results provide additional support for the advantage of the iFusion estimator in extreme-event prediction.

8. Conclusions and Further Discussion

In this paper, we develop an iFusion approach for estimating the EVI in multi-source settings. The key insight is that incorporating estimators from relevant data sources can significantly improve the efficiency of an initial estimator that relies solely on the target source. We establish the iFusion estimators and show their consistency and asymptotic normality under both independent and tail-dependent cases. The advantages of our approach are demonstrated through extensive simulation studies and real data applications, where it consistently outperforms several existing methods, especially in variance reduction and out-of-sample prediction, highlighting its potential to advance the field.

While the proposed iFusion estimators demonstrate robust performance both theoretically and empirically, several directions remain for further investigation. First, the fused EVI estimator can be combined with standard target-specific tail extrapolation methods, such as Weissman's estimator, to estimate extreme quantiles and tail risk measures. Second, the proposed framework can be extended beyond the classical Hill estimator. Other EVI estimators, such as moment, Pickands-type, or bias-corrected Hill estimators, may be incorporated by construct-

ing suitable source-wise inference objects and applying the same screening-and-fusion principle. Finally, this paper establishes asymptotic results under a fixed number of data sources N. An important direction for future research is to extend the theory to regimes where N may diverge with the overall sample size, a setting that is increasingly relevant in modern systems such as federated learning, sensor networks, and cloud computing. Such an extension is conceptually feasible when source-level sample sizes also grow, but it typically requires strengthened regularity conditions to ensure that N-dependent remainder terms remain asymptotically negligible. In contrast, when each source contributes only a limited number of observations, substantially stronger structural assumptions – such as the presence of many sources lying in a small neighborhood of the target – or alternative methodological frameworks may be necessary; see, for example, the discussion in Shen et al. (2020).

Supplementary Materials

The Supplementary Material contains the proofs of all theoretical results, auxiliary information, as well as additional simulation studies and real data analyses.

Acknowledgements

The authors acknowledge the Editor, the Associate Editor, and two anonymous reviewers for their very helpful comments that led to a greatly improved version of this paper. The work is supported by the National Natural Science Foundation of China (Nos. 12501657, 12371279, 12231017, 71991471, 72442027) and the Innovation Project (E5820801) of the Ministry of Education of China.

References

Ahmed, H. and J. H. Einmahl (2019). Improved estimation of the extreme value index using related variables. *Extremes 22*(4), 553–569.

Ahmed, H., J. H. Einmahl, and C. Zhou (2025). Extreme value statistics in semi-supervised models. *Journal of the American Statistical Association 120*(549), 291–304.

Caeiro, F., M. I. Gomes, and D. Pestana (2005). Direct reduction of bias of the classical Hill estimator. *REVSTAT-Statistical Journal 3*(2), 113–136.

Cai, C., R. Chen, and M.-g. Xie (2020). Individualized inference through fusion learning. *Wiley Interdisciplinary Reviews: Computational Statistics 12*(5), e1498.

Chen, L., D. Li, and C. Zhou (2022). Distributed inference for the extreme value index. *Biometrika 109*(1), 257–264.

Cui, Y. and M.-g. Xie (2023). Confidence distribution and distribution estimation for modern statistical inference. In *Springer Handbook of Engineering Statistics*, pp. 575–592. Springer.

Daouia, A., S. A. Padoan, and G. Stupfler (2024). Optimal weighted pooling for inference about the tail index and extreme quantiles. *Bernoulli 30*(2), 1287–1312.

de Haan, L. and A. Ferreira (2006). *Extreme value theory: an introduction*. Springer.

de Haan, L. and C. Zhou (2021). Trends in extreme value indices. *Journal of the American Statistical Association 116*(535), 1265–1279.

Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *The Annals of Applied Probability 10*(4), 1274–1301.

Drees, H. and X. Huang (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis 64*(1), 25–46.

Duan, J., M. Pelger, and R. Xiong (2024). Target PCA: Transfer learning large dimensional panel data. *Journal of Econometrics 244*(2), 105521.

Einmahl, J. H., L. de Haan, and C. Zhou (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology 78*(1), 31–51.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics 3*(5), 1163–1174.

Hsing, T. (1991). On tail index estimation using dependent data. *The Annals of Statistics 19*(3), 1547–1569.

Jin, J., J. Yan, R. H. Aseltine, and K. Chen (2024). Transfer learning with large-scale quantile regression. *Technometrics 66*(3), 381–393.

Kulik, R. and P. Soulier (2020). *Heavy-tailed time series*. Springer.

Li, S., T. T. Cai, and H. Li (2022). Transfer learning for high-dimensional linear regression: Prediction, estimation and minimax optimality. *Journal of the Royal Statistical Society Series B: Statistical*

Methodology 84(1), 149–173.

Li, S. and A. Luedtke (2023). Efficient estimation under data fusion. *Biometrika 110*(4), 1041–1054.

Li, Y., L. Chen, D. Li, and H. Wang (2024). Estimating extreme value index by subsampling for massive datasets with heavy-tailed distributions. *Statistics and Its Interface 17*(4), 605–622.

Shen, J., R. Y. Liu, and M.-g. Xie (2020). iFusion: Individualized fusion learning. *Journal of the American Statistical Association 115*(531), 1251–1267.

Shi, X., Z. Pan, and W. Miao (2023). Data integration in causal inference. *Wiley Interdisciplinary Reviews: Computational Statistics 15*(1), e1581.

Sun, B. and W. Miao (2022). On semiparametric instrumental variable estimation of average treatment effects through data fusion. *Statistica Sinica 32*, 569–590.

Tian, Y. and Y. Feng (2023). Transfer learning under high-dimensional generalized linear models. *Journal of the American Statistical Association 118*(544), 2684–2697.

Wang, H. and C.-L. Tsai (2009). Tail index regression. *Journal of the American Statistical Association 104*(487), 1233–1240.

Wu, P., S. Luo, and Z. Geng (2025). On the comparative analysis of average treatment effects estimation via data combination. *Journal of the American Statistical Association 120*(552), 2250–2261.

Xie, M.-g. and K. Singh (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review 81*(1), 3–39.

Zhang, Y. and Z. Zhu (2025). A data fusion method for quantile treatment effects. *Statistica Sinica 35*, 981–1002.

Hongfang Sun, School of Mathematical Sciences, Ministry of Education Key Laboratory of NSLSCS, Nanjing Normal University

E-mail: hfsun@njnu.edu.cn

Yu Chen, School of Public Affairs, University of Science and Technology of China

E-mail: cyu@ustc.edu.cn

Tao Xu, College of Computer and Information Sciences, Fujian Agriculture and Forestry University

E-mail: xutao@fafu.edu.cn

Zhengjun Zhang, School of Data Science and Artificial Intelligence, Beijing University of Chinese Medicine

School of Economics and Management, University of Chinese Academy of Sciences

Department of Statistics, University of Wisconsin

E-mail: zjz@stat.wisc.edu