



A zero-inflated multivariate peaks-over-threshold model for serial multi-stage processing time data

Hwi Jun Jung , Suk Joo Bae *

Department of Industrial Engineering, Hanyang University, Seoul, Republic of Korea

ARTICLE INFO

Keywords:

Automated monitoring
Extreme value theory
Generalized Pareto distribution
Multivariate max-stable distribution
Tail-dependent structure

ABSTRACT

Serial multi-stage manufacturing process involves a sequence of interconnected processing steps. The complexity of each stage introduces challenges tied to the accumulation of variations, along with highly dependent structure among stages. A delayed or tardy process significantly affects the throughput of overall production by cumulative delays encountered across sequential stages. Incorporating zero-inflated data within target processing time limit and heavy-tailed characteristics, this study proposes two types of zero-inflated multivariate peaks-over-threshold (PoT) models called a zero-inflated multivariate generalized Pareto distribution. The proposed models aim at extending the conventional multivariate PoT models by explicitly incorporating inflated zeros and dependence structure arising from evolutionary manufacturing environments. By leveraging the angular representation and Pickands dependence, MLEs of the model parameters are derived. The inference procedure is developed using the asymptotic consistency and normality of MLEs. The proposed multivariate models are validated through real-world battery manufacturing process data. Zero inflation provides further insight into the underlying characteristics of processing time distributions in PoT applications within manufacturing systems. The proposed method facilitates improved industrial process optimization, especially in detecting and alleviating bottlenecks associated with both nominal operating conditions and sporadic extreme delays.

1. Introduction

Continuing advances in the Internet of Things (IoTs) and smart sensing technology have facilitated the evolution of multi-sensor networks and monitoring techniques using them in intelligent manufacturing structure. Extensive and high-dimensional processing data from multi-sensor networks help process engineers continuously pursue the integration of systemic inspection schemes into key manufacturing pipelines for real-time monitoring. It can permit seamless and comprehensive screening that conventional inspection procedures requiring intermittent intervention by operators scarcely ever provide. In particular, the development of automated monitoring scheme for serial multi-stage manufacturing process is at the core of numerous modern production environments. Recently, devising accurate methods for analyzing multi-stage process data from ubiquitous monitoring practices has gained increasing interest in diverse industrial applications, e.g., semiconductors, displays, and lithium-ion batteries.

Serial multi-stage manufacturing process involves a sequence of interconnected processing steps through which raw materials or intermediate products pass, with the output of one stage serving as the input to the next. This structure facilitates large-scale production with

high consistency, cost-effectiveness, and the capacity to meet diverse market demands. However, the complexity of each manufacturing stage introduces challenges tied to the accumulation of variations, highly correlated quality characteristics between stages, and sensitivity to disturbances. These challenges necessitate rigorous monitoring and analytical methods for ensuring the quality of products and operational efficiency.

This research is mainly motivated by multi-stage manufacturing processes of lithium-ion batteries by a battery manufacturer. Following inherent nature of batch process, battery production is executed in a strictly ordered sequence whereby entire processes complete. In the considered production system, consecutive processes operate under a synchronous tact-time structure. Process completion times within the tact-time are treated as regular completions, whereas exceedances beyond the tact-time represent delayed operations that may propagate congestion to subsequent production stages. In such a serial multi-stage manufacturing process, each process introduces unique sources of variation, along with significant dependent structure among processes. In particular, a delayed or tardy job often referred to as the *bottleneck* job (or process) can significantly affect overall tact time by cumulative

* Corresponding author.

E-mail addresses: wjdglnwns20@psm.hanyang.ac.kr (H.J. Jung), sjbae@hanyang.ac.kr (S.J. Bae).

<https://doi.org/10.1016/j.cie.2026.112352>

Received 6 May 2026; Received in revised form 26 August 2026; Accepted 28 August 2026

Available online 4 September 2026

0360-8352/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

delays encountered across sequential stages of battery manufacturing. At the end of the bottleneck job, every job for a batch process has completed the current operation and the batch can move forward simultaneously to the subsequent process (Shi & Zhou, 2009). Although some processing jobs have finished their assigned operations much earlier, they must wait until the tardiest job in that stage completes its task for the entire batch to move to next stage. Eventually, the bottleneck job dictates entire processes at each stage, impacting overall productivity directly. By identifying and rigorously analyzing time-consuming operations within each stage, process engineers can be better equipped to implement timely interventions for minimizing tact times and enhancing the throughput of overall production.

Meanwhile, most jobs at each stage are required to finish their operations prior to target processing time for a batch (considered as the *threshold*). Processing time data will be zeros if the processes are finished within the threshold. The phenomenon of more zeros than a distribution can accommodate is generally called *zero-inflated*. Because the throughput of each stage is inherently constrained by its tardiest job, it is crucial to develop a robust model that can accurately quantify the delays associated with the bottleneck job (or process) and accommodate a number of zeros. Recent developments have placed an emphasis on advanced quality models for multivariate characteristics, e.g., zero-inflated models and generalized Pareto distribution (GPD) to model peaks-over-threshold (PoT) values, to handle heavy-tailed or zero-inflated outcomes.

To effectively address challenging issues in serial multi-stage processing data, this study proposes a new zero-inflated multivariate peaks-over-threshold (PoT) model; that is, the zero-inflated multivariate generalized Pareto distribution (called “ZIMGPD” hereinafter). A significant limitation of (multivariate) GPD is that it can only model observations that exceed a specified threshold, leaving uncertainty about the behaviors of observations below this threshold denoted as zeros. Various attempts have been made to model the entire range of data using the mixture of several models. In a similar vein, the ZIMGPD aims to extend the conventional multivariate GPD model by explicitly incorporating zero-inflated data. In addition, by explicitly capturing tail dependency, the ZIMGPD provides enhanced flexibility and interpretability in modeling complex structure of (serial) multi-stage manufacturing data. We propose the application of the ZIMGPD model as a robust and practical solution for the analysis of zero-inflated and dependent extremal data arising from evolutionary manufacturing environments.

2. Literature review

A key advantage of serial multi-stage production systems lies in their capacity for specialization, such that each stage can be separately optimized to perform a dedicated task by leveraging specialized equipment and expertise. While this modular approach can improve manufacturing efficiency, each stage also introduces distinct sources of variation that may propagate across and become correlated with subsequent stages. Continual improvement of serial multi-stage processes therefore requires a thorough understanding of parameter estimation and asymptotic properties under complex dependence structures. For example, (Zantek et al., 2002) measured the effect of each stage's performance on the output quality of subsequent stages and identified where investments in process quality improvement should be concentrated. Xiang and Tsung (2008) developed an engineering-model-based framework for statistical analysis of multi-stage processes. Lee et al. (2021) proposed a multi-response optimization method in a multistage process using a classification and regression tree (CART) method. Shi (2006) provided unified and systematic methodologies for variation management and reduction in multi-stage manufacturing processes.

Existing approaches have been developed mainly under the asymptotic normality of parameters or quality metrics for modeling multivariate process data. However, quality characteristics arising from

multi-stage processes may be non-Gaussian, so methods relying on Gaussian assumptions may not adequately describe their distributional behavior. Moreover, the conditions observed at individual stages can exhibit complex profiles and strong dependence across stages, as illustrated by machining process data with cutting-force profiles (Shi et al., 2007). Bacro et al. (2025) proposed hierarchical constructions using a latent random vector with gamma margins to model multivariate threshold exceedances. Nevertheless, statistical modeling methods for highly correlated multivariate extremal data remain underdeveloped.

The estimation of maxima or minima rather than average behavior is crucial in applications involving rare but potentially severe outcomes, such as minimum breaking stress, minimum material strength, or exceptionally long processing delays. Distributions for extremal values have been developed from the limiting properties of order statistics. Among them, the generalized extreme value (GEV) distribution arises as the limiting distribution of normalized block maxima under the Fisher–Tippett–Gnedenko theorem (de Haan & Ferreira, 2006). For a sequence of independent and identically distributed random variables, suitably normalized maxima converge to the Gumbel, Fréchet, or Weibull type, which are unified within the GEV framework. Developed from extreme value theory, the GPD provides a useful framework for modeling peaks over a high threshold and quantifying rare, heavy-tailed outcomes. Coles and Tawn (2005) incorporated seasonal effects into GPD-based models for extreme coastal surges, while Belzile and Davison (2022) developed likelihood-based inference for risk measures in univariate extreme-value problems. In an industrial setting, Boyce et al. (2017) used extreme-value statistics to identify rare failure-critical defects in additive manufacturing, demonstrating the practical importance of modeling tail behavior. Murphy et al. (2025) developed an automated threshold selection methodology for accounting for the uncertainty in the threshold estimation and propagating this uncertainty through to high quantile inference. Despite these developments, the application of EVT to multi-stage manufacturing processes remains limited.

As mentioned in the motivating example of a lithium-ion battery manufacturing process, processing-time exceedances contain many zeros because numerous jobs finish before the specified threshold. Such data require a model that allocates more probability mass at zero than standard continuous distributions can accommodate. For count data, zero-inflated Poisson (ZIP) and zero-inflated negative binomial models have frequently been used to represent excessive zeros through a mixture of a Bernoulli component and a count distribution. Lambert (1992) introduced ZIP regression for manufacturing defect data, and Angers and Biswas (2003) developed a Bayesian zero-inflated generalized Poisson model. In manufacturing applications, Bae et al. (2007) proposed a ZIP yield model for spatially nonhomogeneous wafer defects, and Yuan et al. (2011) developed a hierarchical Bayesian ZINB model for integrated-circuit yield. These approaches, however, primarily concern univariate count data. Couturier and Victoria-Feser (2010) proposed a zero-inflated truncated Pareto model for continuous observations, while Wang and Li (2021) developed generalized confidence intervals for a univariate zero-inflated Pareto distribution. Recently, Tang et al. (2025) proposed a zero-inflated multivariate tobit regression to model non-negative multivariate zero-inflated continuous data. However, most existing zero-inflated models therefore remain univariate and do not explicitly account for dependence among multivariate heavy-tailed observations. Main contributions of this work can be summarized as follows:

- We propose the development of a unified modeling framework that integrates zero inflation, marginal extreme-value behavior, and multivariate dependence. Rather than treating the separation of the discrete component as an independent methodological contribution, the proposed framework combines these components to characterize both within-tact completions and dependent process delays.

- By leveraging the angular representation and Pickands dependence, we specify regularity conditions to derive asymptotic distributions for the MLEs of the model parameters. Then, we develop inference procedure on all the parameters based on their asymptotic consistency and normality.
- We validate the proposed method via the analysis of real-world battery manufacturing process data in terms of modeling accuracy, particularly in the vicinity of zero where conventional multivariate PoT models exhibit systematic bias in capturing the probability mass at zero.

3. The zero-inflated multivariate peaks-over-threshold model

3.1. Theoretical background

In this section, we provide a formal definition of the ZIMGPD as the mixture of the MGPD and the probability mass at zero, then reformulate motivating characterizations. Suppose $\{\mathbf{X}_i, i \geq 1\} \equiv \{(X_i^{(1)}, \dots, X_i^{(d)}), i \geq 1\}$ are independent and identically distributed random vectors in $\mathbb{R}^d$ with a distribution function F and the tail function of a distribution $\bar{F} = 1 - F$. Let $\mathbf{M}_n$ be the vector of componentwise maxima as

$$\mathbf{M}_n \equiv (M_n^{(1)}, \dots, M_n^{(d)}) = \left(\bigvee_{i=1}^n X_i^{(1)}, \dots, \bigvee_{i=1}^n X_i^{(d)} \right).$$

It is assumed that there exist normalizing constants $\sigma_n^{(i)}(> 0)$, $u_n^{(i)} \in \mathbb{R}$, $1 \leq i \leq d$, $n \geq 1$, such that as $n \rightarrow \infty$,

$$P[(M_n^{(i)} - u_n^{(i)})/\sigma_n^{(i)} \leq x_n^{(i)}, 1 \leq i \leq d] = F^n(\sigma_n^{(1)}x_n^{(1)} + u_n^{(1)}, \dots, \sigma_n^{(d)}x_n^{(d)} + u_n^{(d)}) \rightarrow G(\mathbf{x}), \quad (1)$$

with the (multivariate) limiting distribution G with $0 < G(\mathbf{0}) < 1$ such that each marginal G_i is non-degenerate for $i = 1, \dots, d$. F is said to be in the (so-called) *max-domain of attraction* of a multivariate max-stable distribution G , that is, $F \in D(G)$, where G is a multivariate extreme value distribution. By setting all $\mathbf{x}$ except $x^{(i)}$ to $+\infty$, each marginal G_i of G becomes an extreme value distribution so that

$$G_i(x) = \exp \left\{ - \left(1 + \xi_i \frac{x - \mu_i}{\sigma_i} \right)_+^{-1/\xi_i} \right\}, \quad (2)$$

where μ_i is a location parameter, $\sigma_i(> 0)$ is a scale parameter, and ξ_i is a shape parameter. Here, $a_+ \equiv \max(a, 0)$. Note that when $\xi_i \rightarrow 0$, we have the limiting form as

$$G_i(x) = \exp \left\{ - \exp \left(- \frac{x - \mu_i}{\sigma_i} \right) \right\}.$$

For $\mathbf{u}_n = (u_n^{(1)}, \dots, u_n^{(d)})$, $\boldsymbol{\sigma}_n = (\sigma_n^{(1)}, \dots, \sigma_n^{(d)})$, we can represent a multivariate form as

$$P \left[\frac{\mathbf{M}_n - \mathbf{u}_n}{\boldsymbol{\sigma}_n} \leq \mathbf{x} \right] \rightarrow G(\mathbf{x}). \quad (3)$$

To define the model with exceedances, let $\mathbf{X}_u$ be the vector of normalized exceedances over the threshold levels $\mathbf{u}$ as $\mathbf{X}_u = (\mathbf{X} - \mathbf{u})/\boldsymbol{\sigma}(\mathbf{u})$, where $\{\mathbf{u}(t) | t \in [1, \infty)\}$ is a d -dimensional curve starting $\mathbf{u}(1) = \mathbf{0}$, and $\boldsymbol{\sigma}(\mathbf{u}(t))(> 0)$ is a function with values in $\mathbb{R}^d$. If $F \in D(G)$, there exists an increasing continuous curve $\mathbf{u}$ with $F(\mathbf{u}(t)) \rightarrow 1$ as $t \rightarrow \infty$, and a function $\boldsymbol{\sigma}(\mathbf{u}(t))$ such that

$$P[\mathbf{X}_u \leq \mathbf{x} | \mathbf{X}_u \not\leq \mathbf{0}] \rightarrow - \frac{1}{\log G(\mathbf{0})} \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})} \equiv H(\mathbf{x}), \quad (4)$$

where H corresponds to the MGPD having its non-degenerate marginals in $\mathbb{R}_+$ (Rootz'en & Tajvidi, 2006). Particularly, $H(\mathbf{x}) = 0$ for $\mathbf{x} < \mathbf{0}$, $H(\mathbf{x}) = 1 - \log G(\mathbf{x})/\log G(\mathbf{0})$, and $G(\mathbf{0}) = e^{-1}$, for $\mathbf{x} > \mathbf{0}$. The marginals of H are two-parameter GPD as

$$h(x; \xi, \sigma) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{\xi x}{\sigma} \right)_+^{-\frac{1}{\xi}-1}, & \text{if } \xi \neq 0, \\ \frac{1}{\sigma} \exp \left(-\frac{x}{\sigma} \right), & \text{if } \xi = 0, \end{cases} \quad (5)$$

where σ is the scale parameter and ξ is the shape parameter that determines the heaviness of the distribution's tail. x , which is the observation exceeding the threshold u , satisfies $0 < x < \infty$ if $\xi \geq 0$, and $0 < x < -\sigma/\xi$, if $\xi < 0$. The case of $\xi = 0$, which is the exponential distribution with the mean of σ , needs to be interpreted as the limit distribution as $\xi \rightarrow 0$. When $\xi = -1$, the GPD becomes a uniform distribution $U[0, \sigma]$. The GPD is a generalization of the Pareto distribution, which was studied extensively by Arnold (2015).

Developed under the umbrella of the EVT providing a robust framework for analyzing extremal events across diverse fields, the (multivariate) GPD is a fundamental tool for modeling extreme observations that exceed a defined threshold, particularly modeling tail behaviors of PoT (Coles, 2001). Davison and Smith (1990) demonstrated that exceedances over a sufficiently high threshold converge to the MGPD, providing a theoretical basis for statistical modeling of extremal events in multivariate setting.

3.2. The ZIMGPD model

As the mixture form of the probability mass function of zeros and the GPD, we express the marginals of a ZIMGPD as

$$f_x(x; \sigma, \xi, \pi) = \begin{cases} \pi, & \text{if } x = 0, \\ (1 - \pi) \cdot h(x; \xi, \sigma), & \text{if } x > 0, \xi \neq 0, \end{cases} \quad (6)$$

where the parameter $0 < \pi < 1$ represents the probability mass assigned to zero. The ZIMGPD is designed to accommodate excessive zeros frequently encountered in real-world applications. While the MGPD effectively models heavy-tailed behavior dominated by extremal events (Coles, 2001; Rootz'en et al., 2018), it assumes that observations belong to the positive exceedance region. This assumption can lead to inadequate inference when a substantial proportion of observations are recorded as zeros, as in the battery processing-time data (Kulkarni & Patil, 2018).

The proposed ZIMGPD combines two components: a zero-inflated one with probability π for representing the occurrence of structural zeros and an MGPD with probability $1 - \pi$ for modeling non-zero values having heavy-tailed nature. By integrating the zero-inflation component, the ZIMGPD aims to extend traditional PoT model to more generalized one where excessive zeros significantly affect inference procedure on the parameters or quantities of interest. Eastoe and Tawn (2012) emphasized the importance of modeling cluster maxima in extreme value analysis, validating the need for improved PoT modeling framework.

In this work, we propose two types of the ZIMGPD: Type I and Type II ZIMGPD models. With $\xi = 0$, the MGPD term in the ZIMGPD reduces to an exponential distribution and the estimation of σ in this case is relatively easy. Henceforth, we concentrate on the case where $\xi \neq 0$ for two types of ZIMGPD models in this study. Type I ZIMGPD assumes that each variate for exceedances over threshold follows an independent GPD process, then, for $\xi_i > -\frac{\sigma_i}{x_{i(n)}}$, where $x_{i(n)}$ is the largest order statistic for the i th variate, a generalization to the d -variate case becomes obvious and the joint density function can be written as

$$f(\mathbf{x}; \boldsymbol{\pi}, \boldsymbol{\xi}, \boldsymbol{\sigma}) = \prod_{i=1}^d \left\{ \pi_i \mathbb{I}_{\{0\}}(x_i) + (1 - \mathbb{I}_{\{0\}}(x_i)) \cdot \frac{1}{\sigma_i} \left(1 + \frac{\xi_i x_i}{\sigma_i} \right)_+^{-\frac{1}{\xi_i}-1} \right\}, \quad (7)$$

for $\mathbf{x} \equiv (x_1, \dots, x_d)^T$, $\boldsymbol{\pi} \equiv (\pi_1, \dots, \pi_d)^T$, $\boldsymbol{\xi} \equiv (\xi_1, \dots, \xi_d)^T$, and $\boldsymbol{\sigma} \equiv (\sigma_1, \dots, \sigma_d)^T$, where $\mathbb{I}_{\{0\}}(x)$ is a degenerate distribution concentrated at zero. The support is defined by $\{\mathbf{x} : 1 + \xi_i x_i/\sigma_i > 0\}$, where $\sigma_i > 0$ and $-\infty < \xi_i < \infty$. This formulation preserves the interpretability of each marginal distribution while providing a flexible framework for modeling multivariate data (Arnold, 1990).

Taking tail-dependent structure among extremal values into account, Type II ZIMGPD model is given by

$$f(\mathbf{x}; \boldsymbol{\pi}, \boldsymbol{\xi}, \boldsymbol{\sigma}, \boldsymbol{\alpha}) = \begin{cases} \prod_{i=1}^d \pi_i, & \text{if } x_i = 0, \\ \frac{\prod_{i=1}^d (1 - \pi_i)}{\prod_{i < j} (1 - \alpha_{ij}^2)} \cdot \frac{1}{\prod_{i=1}^d x_i} \cdot \left(\prod_{i=1}^d \left(\frac{\sigma_i}{x_i} \right)^{\xi_i} \right)^{\frac{1}{\prod_{i < j} (1 - \alpha_{ij}^2)}} & \text{if } x_i > 0, \\ \times \prod_{i < j} J_0 \left(\frac{2\sqrt{|\alpha_{ij} \xi_i \xi_j \log(x_i/\sigma_i) \log(x_j/\sigma_j)|}}{1 - \alpha_{ij}^2} \right), & \end{cases} \quad (8)$$

where $J_k(z) = \sum_{r=0}^{\infty} \{\Gamma(r+1)\Gamma(k+r+1)\}^{-1} (z/2)^{k+2r}$ is the modified Bessel function of the first kind of the k th order, which is needed to ensure that the dependency structure is properly reflected in the joint density function. The correlation term

$$\alpha_{ij} = \text{Corr} \left(\xi_i \log \left(\frac{x_i}{\sigma_i} \right), \xi_j \log \left(\frac{x_j}{\sigma_j} \right) \right) \quad (9)$$

quantifies the dependency between log-transformed exceedances of x_i and x_j . This correlation structure accounts for the likelihood that an extremal observation in one dimension is associated with an extremal observation in the other. Fig. 1 presents density contours for the Type II ZIMGPD model with a variety of correlation structures, showing its tail dependent characteristics. Along with higher values of correlation α_{ij} , larger data are highly scattered on tails of the Type II ZIMGPD model. Note that the Type II ZIMGPD model corresponds to the Type I ZIMGPD model when $\alpha_{ij} = 0$.

Along with the presence of structural zeros, the dependency between exceedances may be modeled using a copula-like formulation linking tail behaviors of multivariate data. However, we mainly express the dependence structure via correlation matrix that governs joint behaviors of extremal observations in this work. This kind of dependency structure has been often employed under the MGPD formula in financial portfolio risk analysis where extreme losses often propagate across multiple assets, and environmental applications in which extreme climatic events tend to occur in clusters.

4. Inference procedure

Inference procedure for extremal events (or values) in multivariate cases are quite complicated since there is no universal definition of an exceedance of multivariate threshold. Whatever the definition is selected, the family of distributions proposed by asymptotic theory may be no longer parametric. As the mixture of degenerate distribution at zero and the MGPD, excessive zeros make a precise estimation of the parameters in the ZIMGPD model more difficult, presenting unique challenges for statistical inference. In particular, the proportion of zeros corresponding to the values smaller than the threshold, π would be predicted through the pdf of the ZIMGPD $f(\mathbf{x})$ or observed number of zeros empirically.

4.1. Type I ZIMGPD model

The log-likelihood function for n multivariate observations, $\mathbf{X} \equiv (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, $\mathbf{x}_i \in \mathbb{R}^d$, under the Type I ZIMGPD model is

$$\begin{aligned} \ell(\mathbf{X}; \boldsymbol{\theta}_1) &= \sum_{i=1}^d \sum_{j=1}^n \left\{ \pi_i \mathbb{I}_{\{0\}}(x_{ij}) + \log(1 - \mathbb{I}_{\{0\}}(x_{ij})) - \log \sigma_i - \left(1 + \frac{1}{\xi_i} \right) \log \left(1 + \frac{\xi_i x_{ij}}{\sigma_i} \right) \right\}, \end{aligned} \quad (10)$$

for the parameter set $\boldsymbol{\theta}_1 \equiv (\boldsymbol{\pi}, \boldsymbol{\theta})$, where $\boldsymbol{\theta} \equiv (\boldsymbol{\xi}, \boldsymbol{\sigma})$. When the origins of zeros are known and $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$ are not functionally related, the log-likelihood is easy to maximize because $\mathbb{I}_{\{0\}}(x_{ij})$ and MGPD can be maximized separately. Let $A_i = \{x_{ij} : x_{ij} = 0, j = 1, \dots, n\}$ be the set of values equal to 0 and $N(A_i)$ the number of values equal

to 0 for the i th variate. Then, the elements of A_i come from the zero degenerate distribution. First, the proportion of zeros π_i can be estimated by $N(A_i)/n$, for $i = 1, \dots, d$. Then, MLEs of ξ_i and σ_i are obtained through the maximization of the log-likelihood (10) by setting its partial derivatives with respect to ξ_i and σ_i to zero as

$$\frac{\partial \ell}{\partial \sigma_i} = \sum_{j=1}^n \left\{ -\frac{1}{\sigma_i} + \left(1 + \frac{1}{\xi_i} \right) \frac{\xi_i x_{ij}}{\sigma_i^2 \left(1 + \frac{\xi_i x_{ij}}{\sigma_i} \right)} \right\}_{x_{ij} > 0} = 0, \quad (11)$$

$$\frac{\partial \ell}{\partial \xi_i} = \sum_{j=1}^n \left\{ -\frac{1}{\xi_i^2} \log \left(1 + \frac{\xi_i x_{ij}}{\sigma_i} \right) + \left(1 + \frac{1}{\xi_i} \right) \frac{x_{ij}}{\sigma_i \left(1 + \frac{\xi_i x_{ij}}{\sigma_i} \right)} \right\}_{x_{ij} > 0} = 0. \quad (12)$$

From Eqs. (11) and (12), it is straightforward to obtain the following ML equations:

$$\frac{1}{n} \sum_{j=1}^n \log \left(1 + \frac{\xi_i x_{ij}}{\sigma_i} \right)_{x_{ij} > 0} = \xi_i, \quad \text{and} \quad \frac{1}{n} \sum_{j=1}^n \left(1 + \frac{\xi_i x_{ij}}{\sigma_i} \right)^{-1}_{x_{ij} > 0} = \frac{1}{1 + \xi_i}. \quad (13)$$

To obtain the MLEs, no explicit nor exact solutions for Eq. (13) can be given. The log-likelihood may be made arbitrarily large by taking $\xi_i > 1$ and σ_i/ξ_i arbitrarily close to $\max(x_{ij})$, so the MLEs of the parameters are taken to be the values $\hat{\xi}_i$ and $\hat{\sigma}_i$ yielding a local maximum of ℓ . Finding the local maximum of ℓ requires numerical methods such as Newton–Raphson iteration because the minimal sufficient statistics for the Type I ZIMGPD are the order statistics. When the shape parameter of the GPD is larger than $-1/2$, Smith (1985) showed that, under certain regularity conditions, the MLEs of (univariate) GPD are asymptotically efficient and normal. In addition to ML estimation, alternative estimation methods such as the method of moments, probability-weighted moments, and L-moments can be employed to estimate the marginal GPD parameters. These estimators provide numerically stable alternatives in relatively small size of samples, although their admissible parameter ranges and finite-sample properties differ from those of the MLE.

After obtaining the MLEs of π_i and θ_i , $(\hat{\pi}_i, \hat{\theta}_i)$, we can derive the asymptotic distributions for the MLEs. Because the log-likelihood function of Type I ZIMGPD is the sum of individual likelihoods, the Fisher information matrix and the asymptotic variance of the MLEs are easily derived by using asymptotic properties from the univariate GPD. The asymptotic distribution for the proportion of excessive zeros for the i th variate can be derived from the limiting distribution of $N(A_i)$. The number of zeros for the i th variate follows a binomial distribution and its estimated proportion $\hat{\pi}_i = N(A_i)/n$ is asymptotically normal with the mean π_i and the variance $\pi_i(1 - \pi_i)/n$, that is, $\sqrt{n}(\hat{\pi}_i - \pi_i) \sim \mathcal{N}(0, \pi_i(1 - \pi_i))$. Under certain regularity conditions, the MLEs $\hat{\theta}_i \equiv (\hat{\xi}_i, \hat{\sigma}_i)$ are consistent and asymptotically efficient, and their asymptotic distributions are normal as

$$N_{\{\setminus 0\}} \text{Var} \begin{pmatrix} \hat{\xi}_i \\ \hat{\sigma}_i \end{pmatrix} \sim \mathcal{MN} \begin{pmatrix} 2\sigma_i^2(1 + \xi_i) & \sigma_i(1 + \xi_i) \\ \sigma_i(1 + \xi_i) & (1 + \xi_i)^2 \end{pmatrix}, \quad (14)$$

for $\xi_i > -\frac{1}{2}$, where $N_{\{\setminus 0\}}$ is the total number of values are not equal to zeros (Smith, 1985). Consequently, the asymptotic distribution of $\hat{\boldsymbol{\theta}}_1 \equiv (\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_d)^T$, where $\hat{\boldsymbol{\theta}}_i = (\hat{\xi}_i, \hat{\sigma}_i)^T$, $i = 1, \dots, d$, are (multivariate) normal as $\mathcal{MN}(\boldsymbol{\theta}_1, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_1})$, where the variance–covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_1}$ is of the block diagonal matrix form with its i th element equal to

$$\begin{pmatrix} \frac{\pi_i(1 - \pi_i)}{n}, 0, 0, 0, \frac{2\sigma_i^2(1 + \xi_i)}{N_{\{\setminus 0\}}}, \frac{\sigma_i(1 + \xi_i)}{N_{\{\setminus 0\}}}; 0, \frac{\sigma_i(1 + \xi_i)}{N_{\{\setminus 0\}}}, \frac{(1 + \xi_i)^2}{N_{\{\setminus 0\}}} \end{pmatrix}. \quad (15)$$

Note that the asymptotic variance–covariance matrix is the inverse of expected Fisher information matrix for $\boldsymbol{\theta}_1$. Alternatively the asymptotic variance–covariance matrix can be estimated using the observed

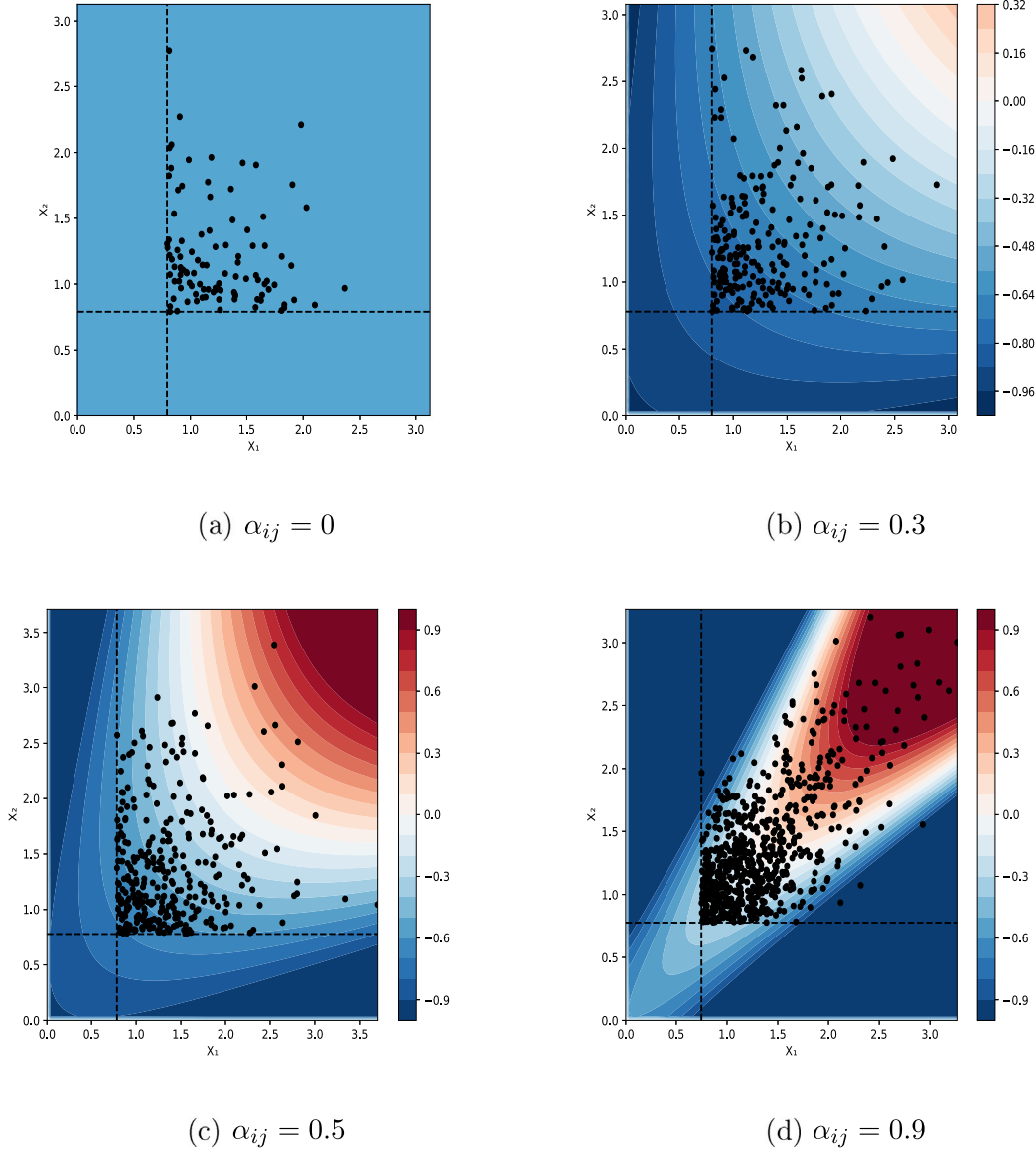


Fig. 1. Density contours of the Type II ZIMGPD model with different correlation values (dashed parallel and horizontal lines depict 90% percentile of the Type II ZIMGPD).

information matrix for Θ_1 as

$$\hat{I}_{\Theta_1} = \sum_{j=1}^n \left[-\frac{\partial^2 \ell_j(\Theta_1)}{\partial \Theta_1 \cdot \partial \Theta_1^T} \right]_{\Theta_1 = \hat{\Theta}_1},$$

for the total number of observations n . Based on the asymptotic normal distribution of the model parameters, asymptotic confidence intervals (CIs) for the MLEs in the Type I ZIMGPD model can be obtained. The unknown parameters in the asymptotic variance of the estimator can be replaced by the MLEs to obtain a consistent estimator of the variance. The approximate $100(1 - \alpha)\%$ CIs for the parameters are

$$\hat{\pi}_i \pm z_{\alpha/2} \sqrt{\frac{\hat{\pi}_i(1 - \hat{\pi}_i)}{n}}, \quad \hat{\xi}_i \pm z_{\alpha/2} \sqrt{\frac{2\hat{\sigma}_i^2(1 + \hat{\xi}_i)}{N_{\{0\}}}}, \quad \text{and} \quad \hat{\sigma}_i \pm z_{\alpha/2} \sqrt{\frac{(1 + \hat{\xi}_i)^2}{N_{\{0\}}}}, \quad (16)$$

where $z_{\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the standard normal distribution.

4.2. Type II ZIMGPD model

The log-likelihood function for n multivariate observations $\mathbf{X}$ under the Type II ZIMGPD model is

$$\begin{aligned} \ell(\mathbf{X}; \Theta_2) = & \sum_{k=1}^n \left\{ \mathbb{I}_{\{0\}}(x_i) \sum_{i=1}^d \log \pi_i + \left[\sum_{i=1}^d \log(1 - \mathbb{I}_{\{0\}}(x_i)) - \sum_{i < j} \log(1 - \alpha_{ij}^2) \right] \right. \\ & + \frac{1}{\prod_{i < j} (1 - \alpha_{ij}^2)} \sum_{i=1}^d \left(-\log \sigma_i - \left(\frac{1}{\xi_i} + 1 \right) \log \left(1 + \frac{\xi_i x_{ik}}{\sigma_i} \right) \right) \\ & \left. + \sum_{i < j} \log \mathcal{J}_0 \left(\frac{2 \sqrt{|\alpha_{ij} \xi_i \xi_j \log(x_{ik}/\sigma_i) \log(x_{jk}/\sigma_j)|}}{1 - \alpha_{ij}^2} \right) \right\}, \quad (17) \end{aligned}$$

for the parameter set $\Theta_2 \equiv (\pi, \sigma, \xi, \alpha) \in [0, 1]^d \times (0, \infty)^d \times \Xi^d \times (-1, 1)^{d(d-1)/2}$, $\Xi \equiv \{\xi_i \in \mathbb{R} : 1 + \frac{\xi_i x_i}{\sigma_i} > 0, \forall x_i > 0\}$, where $\alpha \equiv \{\alpha_{ij} : 1 \leq i < j \leq d\}$ are the correlation matrix with its correlation elements α_{ij} between variates x_i and x_j .

The correlation parameter α_{ij} specifies dependent structure between variates x_i and x_j in the Type II ZIMPD model. However, the interdependence between α_{ij} 's and marginal parameters σ_i and ξ_i create computational challenges in directly deriving MLEs $\hat{\alpha}_{ij}$ for all i and j . The parameter space for α_{ij} is constrained by $|\alpha_{ij}| \leq 1$, which further complicates the optimization process for solving score functions. In this work, we propose a reparameterization approach to resolve the computational difficulty as

$$q_i = \xi_i \log\left(\frac{x_i}{\sigma_i}\right), \text{ such that } \alpha_{ij} = \text{Corr}(q_i, q_j). \quad (18)$$

Inspired from seminal work by [Mardia \(1962\)](#) on multivariate Pareto distributions, the reparameterization significantly improves the robustness and efficiency of inference procedure in the Type II ZIMGPD model, particularly when dealing with heavy-tailed data with complex dependence structures. It permits to easily handle the zero-inflated structure by properly scaling non-zero observations, effectively decoupling the dependence structure from marginal distributions to accurately estimate the parameters of Type II ZIMGPD model. Furthermore, by separating the estimation of correlation parameters from that of marginal parameters, numerical optimization algorithms present better convergence properties when working with the reparameterization, increasing the possibility of reaching to a global optimum.

Lemma 1. *Under the reparameterization (18), the correlation measure α_{ij} between q_i and q_j are location and scale invariant with respect to the parameters σ_i and ξ_i .*

The proof of [Lemma 1](#) is relegated in [Appendix A](#). From the results of [Lemma 1](#), the dependence parameter α_{ij} is determined by the correlation of the transformed variables $q_i = \xi_i \log(x_i/\sigma_i)$ on the log scale. Because the logarithmic transformation is nonlinear, this correlation is not equal to the Pearson correlation between the original variables x_i and x_j ; therefore, the dependence structure of the proposed model is defined through the transformed variables. For notational convenience, define the variable

$$z_{ijk} \equiv \alpha_{ij} \xi_i \xi_j \log(x_{ik}/\sigma_i) \log(x_{jk}/\sigma_j) = \alpha_{ij} \cdot q_{ik} \cdot q_{jk}.$$

Using the chain rules for each parameter and through partial differential equations of each parameter for z_{ijk} and Bessel terms, MLEs of the parameters in the Type II ZIMPD are obtained by solving the following score functions:

$$\begin{aligned} \frac{\partial \ell}{\partial \sigma_i} &= \sum_{k \in K_+} \frac{1}{\prod_{i < j} (1 - \alpha_{ij}^2)} \left[-\frac{1}{\sigma_i} + \left(\frac{1}{\xi_i} + 1 \right) \frac{\xi_i x_{ik}}{\sigma_i^2 (1 + \xi_i x_{ik}/\sigma_i)} \right] \\ &\quad + \sum_{k \in K_+} \sum_{j \neq i} \frac{J_1(z'_{ijk})}{J_0(z'_{ijk})} \left[-\alpha_{ij} \cdot \left(\frac{\xi_i}{\sigma_i} \right) \cdot \xi_j \log\left(\frac{x_{jk}}{\sigma_j}\right) \right] = 0, \\ \frac{\partial \ell}{\partial \xi_i} &= \sum_{k \in K_+} \frac{1}{\prod_{i < j} (1 - \alpha_{ij}^2)} \left[\frac{1}{\xi_i^2} \log\left(1 + \frac{\xi_i x_{ik}}{\sigma_i}\right) - \left(\frac{1}{\xi_i} + 1 \right) \frac{x_{ik}/\sigma_i}{1 + \xi_i x_{ik}/\sigma_i} \right] \\ &\quad + \sum_{k \in K_+} \sum_{j \neq i} \frac{J_1(z'_{ijk})}{J_0(z'_{ijk})} \left[\alpha_{ij} \cdot \log\left(\frac{x_{ik}}{\sigma_i}\right) \cdot \xi_j \log\left(\frac{x_{jk}}{\sigma_j}\right) \right] = 0, \\ \frac{\partial \ell}{\partial \alpha_{ij}} &= \frac{2\alpha_{ij} N_{\setminus \{0\}}}{1 - \alpha_{ij}^2} + \sum_{k \in K_+} \frac{2\alpha_{ij}}{\prod_{i < j} (1 - \alpha_{ij}^2)} \sum_{l=1}^d \left(-\log \sigma_i - \left(\frac{1}{\xi_i} + 1 \right) \log\left(1 + \frac{\xi_i x_{ik}}{\sigma_i}\right) \right) \\ &\quad + \sum_{k \in K_+} \frac{J_1(z'_{ijk})}{J_0(z'_{ijk})} \left[\xi_i \log\left(\frac{x_{ik}}{\sigma_i}\right) \cdot \xi_j \log\left(\frac{x_{jk}}{\sigma_j}\right) \right] = 0, \end{aligned} \quad (19)$$

where K_+ denotes the set of non-zeros. Here, z'_{ijk} is the function of z_{ijk} , i.e., $z'_{ijk} \equiv 2\sqrt{|z_{ijk}|/(1 - \alpha_{ij}^2)}$. The derivations of the score functions are detailed in [Appendix B](#).

By invariant property of the correlation measure, MLEs obtained based on the reparameterization are equivalent to the MLEs derived by solving the score functions (19). Similar to the Type I ZIMGPD, after obtaining the MLE of π_i as $\hat{\pi}_i = N(A_i)/n$, the MLEs of ξ, σ , and α can be then derived. The likelihood can be factorized into a discrete component involving π and a continuous component involving the

marginal and dependence parameters. Consequently, the maximization with respect to π is separable from that with respect to σ, ξ , and α . Therefore, the plug-in estimator $\hat{\pi}_i = N(A_i)/n$ is identical to the corresponding component of the joint MLE. However, deriving the MLEs $\hat{\xi}, \hat{\sigma}$, and $\hat{\alpha}$ in the Type II ZIMGPD model is more complicated than the Type I ZIMGPD since the score functions (19) include the modified Bessel functions J_0 and J_1 . Finally after obtaining the MLEs $\hat{\Theta}_2$ by solving the score functions using numerical algorithms such as the quasi-Newton method, the inference on MLEs can be made using asymptotic properties of the MLEs in the Type II ZIMGPD model. By leveraging the angular representation and Pickands dependence function, [Michel \(2006\)](#) specified regularity conditions for proving asymptotic normality of the MLEs in the MGPD model.

We will specify regularity conditions to derive asymptotic distributions for the MLEs of the Type II ZIMGPD model, then make an inference on all the parameters based on their asymptotic distributions in the following.

4.2.1. Angular transformation and Pickands representation

Let $\mathbf{x} = (x_1, \dots, x_d)^T \in \mathbb{R}_+^d$ denote the observed positive vector, which follows the ZIMGPD. To analyze asymptotic properties of the Type II ZIMGPD, we apply the radial-angular decomposition $\mathbf{x} = r\mathbf{w}$, where $r \equiv \|\mathbf{x}\|_1$ is the radial component and angular-specified weights $\mathbf{w} = \{\mathbf{w} \in \mathbb{R}_+^d : \sum_{i=1}^d w_i = 1\}$ lie on the d -variate unit simplex S_d . For a large value of the radial component r , we can approximate the marginal term in the Type II ZIMGPD as

$$\left(1 + \frac{\xi_i x_i}{\sigma_i}\right)_+^{-\frac{1}{\xi_i}-1} = \left(1 + \frac{\xi_i r w_i}{\sigma_i}\right)_+^{-\frac{1}{\xi_i}-1} \approx (r w_i)^{-\frac{1}{\xi_i}-1}, \quad (20)$$

and

$$\begin{aligned} J_0\left(\frac{2\sqrt{|\alpha_{ij} \xi_i \xi_j \log(x_i/\sigma_i) \log(x_j/\sigma_j)|}}{1 - \alpha_{ij}^2}\right) &= J_0\left(\frac{2\sqrt{|\alpha_{ij} \xi_i \xi_j \log(w_i) \log(w_j)|}}{1 - \alpha_{ij}^2}\right) + \mathcal{O}(\log r) \\ &\approx J_0\left(\frac{2\sqrt{|\alpha_{ij} \xi_i \xi_j \log(w_i) \log(w_j)|}}{1 - \alpha_{ij}^2}\right), \end{aligned} \quad (21)$$

where $\mathcal{O}(\log r)$ represents the term that grows logarithmically with r . Along with the representation of the Pickands coordinate with regard to Fréchet margins having asymptotically uniformity for $r \rightarrow \infty$, the Type II ZIMGPD model is expressed as the angular density

$$g(\mathbf{w}; \Theta_2) \propto \prod_{i=1}^d (1 - \pi_i) \cdot w_i^{-\frac{1}{\xi_i}-1} \prod_{i < j} J_0\left(\frac{2\sqrt{|\alpha_{ij} \xi_i \xi_j \log(w_i) \log(w_j)|}}{1 - \alpha_{ij}^2}\right). \quad (22)$$

The corresponding Pickands dependence function is defined as ([Pickands I.I.I, 1975](#))

$$A(\mathbf{w}; \Theta_2) = (1 - \pi_i) \int_{S_d} \max_{i=1, \dots, d} \left(\frac{w_i}{v_i} \right) g(\mathbf{v}; \Theta_2) d\mathbf{v}, \quad (23)$$

where $\mathbf{v} = (v_1, \dots, v_d) \in S_d$ are the integration variables. Under the angular transformation and Pickands representation, the Type II ZIMGPD model satisfies the following regularity conditions:

- (1) The angular density function $g(\mathbf{w}; \Theta_2)$ is sufficiently smooth; that is, three times continuously differentiable.
- (2) The derivatives of $\log g(\mathbf{w}; \Theta_2)$ are dominated by integrable functions.
- (3) The Fisher information matrix exists and it is finite.
- (4) The Pickands dependence function satisfies necessary convexity conditions.

The proofs of the regularity conditions (1) to (4) as for the Type II ZIMGPD are relegated in [Appendix C](#) to [F](#), respectively.

4.2.2. Asymptotic properties of MLEs

Satisfying the four regularity conditions for the Type II ZIMGPD model in the previous subsection, the MLEs possess desirable asymptotic properties. Let us denote the MLEs of the Type II ZIMGPD

model parameters by $\hat{\Theta}_2$ based on a sample of size n . For the angular representation of the Type II ZIMGPD, we define the log-likelihood function as

$$\ell(\mathbf{w}_1, \dots, \mathbf{w}_n; \Theta_2) = \sum_{i=1}^n \log g(\mathbf{w}_i; \Theta_2) - n \log d_2^*, \quad (24)$$

where $d_2^* = \int_{S_d} g(\mathbf{w}; \Theta_2) d\mathbf{w}$ is the normalizing constant. The score function is then given by

$$A_n = \frac{1}{n} \sum_{i=1}^n \nabla_{\Theta_2} \log g(\mathbf{w}_i; \Theta_2). \quad (25)$$

Under the regularity conditions established above, we can apply the Taylor expansion to the score equation $\nabla_{\Theta_2} \ell(\hat{\Theta}_2) = 0$, yielding

$$\sqrt{n}(\hat{\Theta}_2 - \Theta_2) = -\sqrt{n}G_n^{-1}A_n + \mathcal{O}_d(1), \quad (26)$$

where $G_n = A_n + B_n + \frac{1}{n} \sum_{i=1}^n R(\mathbf{w}_i; \hat{\Theta}_2, \Theta_2)$ with $B_n = \frac{1}{n} \sum_{i=1}^n \nabla_{\Theta_2}^2 \log g(\mathbf{w}_i; \Theta_2)$ and $R(\cdot)$ representing the Taylor remainder term. By the law of large numbers, $G_n \xrightarrow{p} -I(\Theta_2)$ as $n \rightarrow \infty$, where $I(\Theta_2)$ is the Fisher information matrix evaluated at the true parameter values of the Type II ZIMGPD model Θ_2 . By the central limit theorem, $\sqrt{n}A_n \xrightarrow{d} \mathcal{MN}(\mathbf{0}, I(\Theta_2))$. When the regularity conditions are satisfied, by Slutsky's theorem, the MLEs are consistent and asymptotically normally distributed as

$$\sqrt{n}(\hat{\Theta}_2 - \Theta_2) \xrightarrow{d} \mathcal{MN}(\mathbf{0}, I^{-1}(\Theta_2)). \quad (27)$$

The Fisher information matrix, $I(\Theta_2)$, has a block structure due to the invariance properties of α_{ij} , with zero blocks corresponding to the covariances between (ξ_i, σ_i) and α_{jk} . This structure leads to asymptotic independence between these parameter estimators, which presents a desirable property for statistical inference on the Type II ZIMGPD model. This structural property holds for $\xi_i > -\frac{1}{2}$, following the regularity conditions for the asymptotic normality of the MLEs. The results provide a rigorous statistical foundation for the inference on the Type II ZIGPD model, permitting asymptotic efficiency of the MLEs, hypothesis tests, and the constructions of CIs for the parameters.

Asymptotic CIs for the MLEs in the Type II ZIMGPD model can be then obtained based on the asymptotic normality of the model parameters. The estimated Fisher information matrix of the Type II ZIMGPD model can be expressed as $\hat{I}(\hat{\Theta}_2) = -\frac{\partial^2 \ell(\Theta_2)}{\partial \Theta_2 \partial \Theta_2^T} \Big|_{\Theta_2 = \hat{\Theta}_2}$. The unknown parameters in the asymptotic variance of the estimator can be replaced by the MLEs to obtain a consistent estimator of the variance. However, the construction of CIs for the Type II ZIMGPD presents additional challenges due to the complexity introduced by the modified Bessel function of the first kind, J_0 . Because higher-order derivatives of J_0 rely on additional terms involving the modified Bessel functions of increasing order, recursively dependent on $J_1(x)$, $J_2(x)$, and so on, such as

$$\frac{d}{dx} \log J_0(x) = \frac{J_1(x)}{J_0(x)},$$

which leads to the expressions that cannot be reduced to elementary algebraic functions. As a result, traditional analytical solutions for the MLE parameters ξ and σ are infeasible. We re-formulate the modified Bessel function J_0 as an infinite series

$$J_0(x) = \sum_{m=0}^{\infty} \frac{(x/2)^{2m}}{(m!)^2}. \quad (28)$$

Besides, since J_0 includes additional terms arising from the dependence structure, we approximate them using numerical optimization methods to iteratively refine parameter estimates. After ensuring that the Hessian matrix is positive definite and the determinant is positive, i.e., $\left(\frac{\partial^2 \ell}{\partial \xi_i^2}\right) \left(\frac{\partial^2 \ell}{\partial \sigma_i^2}\right) > \left(\frac{\partial^2 \ell}{\partial \xi_i \partial \sigma_i}\right)^2$, the approximate $100(1-\alpha)\%$ CI for each parameter can then be constructed as

$$\hat{\xi}_i \pm z_{\alpha/2} \sqrt{\hat{I}_{ii}^{-1}(\hat{\Theta}_2)|_{i=1}}, \quad \text{and} \quad \hat{\sigma}_i \pm z_{\alpha/2} \sqrt{\hat{I}_{ii}^{-1}(\hat{\Theta}_2)|_{i=2}}, \quad (29)$$

where $\hat{I}_{ii}^{-1}(\hat{\Theta}_2)$ represents the i th diagonal element of the inverse of the estimated Fisher information matrix.

5. Simulation study

A simulation study was conducted to examine the sensitivity of the proposed model to different zero-inflation levels and evaluate the finite-sample performance of alternative estimators for the marginal GPD parameters. The simulation study consists of two parts: First, the zero-inflation probability was varied to assess whether the proposed model can appropriately represent different proportions of zero observations. Second, ML estimation was compared with the method of moments, probability-weighted moments, and L-moment estimation under different sample sizes and shape-parameter settings.

5.1. Sensitivity to the zero-inflation probability

The effect of the zero-inflation probability was examined under $\pi \in \{0.1, 0.3, 0.6\}$, representing low, moderate, and high zero-inflation regimes, respectively. The remaining model parameters were fixed across the three scenarios so that the effect of π could be evaluated separately.

As shown in Fig. 2, the proposed model appropriately reflects the different proportions of zero observations. The estimated zero-inflation probabilities were $\hat{\pi} = 0.099$, $\hat{\pi} = 0.312$, and $\hat{\pi} = 0.596$ when the true values were $\pi = 0.1$, $\pi = 0.3$, and $\pi = 0.6$, respectively. The estimated probability mass at zero is close to the true value in all three scenarios. The fitted positive distributions also remain close to the simulated distributions, indicating that changes in the zero-inflation level do not significantly affect the estimation of the continuous component. In contrast, the conventional MGPD assigns $\Pr(X_i = 0) = 0$, thus it cannot directly represent a point mass at zero. The proposed model explicitly accounts for the probability of zero observations through the zero-inflation component, while the MGPD characterizes the magnitude and dependence of positive exceedances. These results demonstrate that the proposed model can accommodate a wide range of zero-inflation levels without changing its underlying continuous-tail structure.

5.2. Finite-sample performance of marginal estimators

The finite-sample performance of the MLE, the method-of-moments estimator (MOM), the probability-weighted-moment estimator (PWM), and the L-moment estimator was also investigated. The estimators were compared in terms of mean residual life (MRL) and root mean squared error (RMSE) under different sample sizes and GPD shape-parameter settings.

Let the positive exceedances follow a GPD with scale parameter σ and shape parameter ξ . The MOM estimator is obtained from the first two ordinary moments of the GPD. The GPD mean and variance are given by $E(X) = \sigma/(1-\xi)$ and $\text{Var}(X) = \sigma^2/(1-\xi)^2(1-2\xi)$, provided that the corresponding moments exist. Therefore, MOM requires $\xi < 1/2$ because the GPD variance does not exist when $\xi \geq 1/2$. PWM and L-moment estimators are constructed from weighted combinations of ordered observations. Compared with ordinary sample moments, these estimators are generally less sensitive to a small number of extremely large observations. For the two-parameter GPD, the PWM and L-moment estimators based on the first two probability-weighted moments are algebraically equivalent. They are therefore presented together as PWM/L-moment estimators in the simulation results.

The small-sample performance was first examined over sample sizes from $n = 20$ to $n = 200$, using 1000 Monte Carlo replications for each setting. Fig. 3 compares the bias and RMSE of the estimators under the representative setting $\xi = 0.2$. The moment-based estimators showed competitive finite-sample performance when the number of

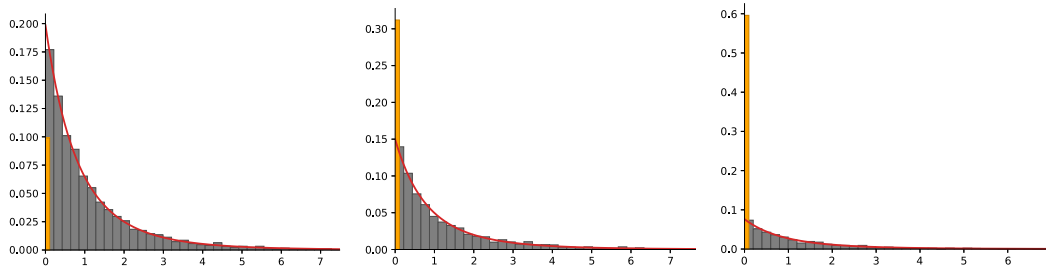


Fig. 2. Sensitivity analysis for the zero-inflation ratio (π) (Orange bars represent the densities of zeros).

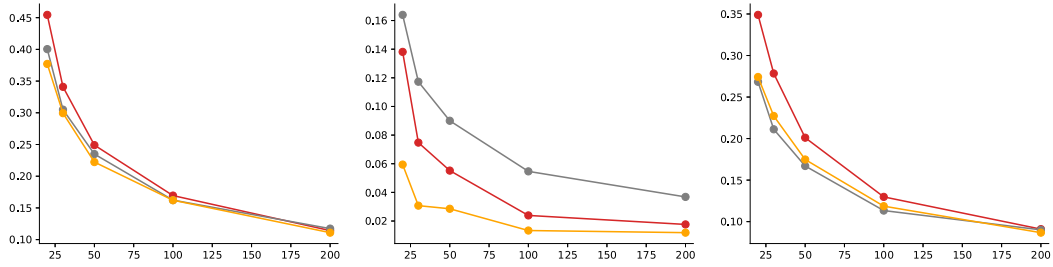


Fig. 3. Comparison of RMSE for MLE, MOM, and PWM (•: MLE, •: MOM, •: PWM).

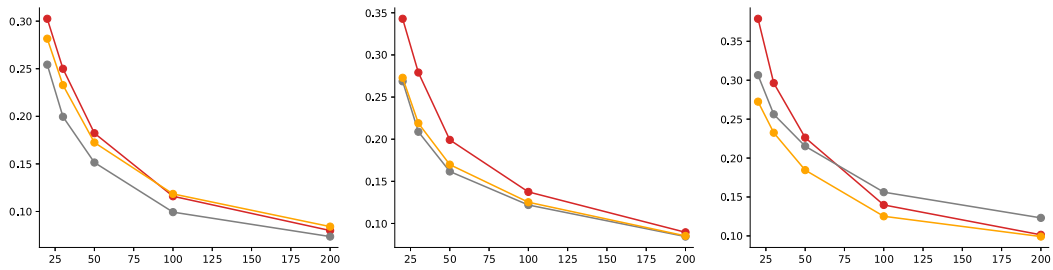


Fig. 4. Comparison of MRL for MLE, MOM, and PWM (•: MLE, •: MOM, •: PWM).

positive exceedances is small. However, the differences among the estimators decrease as n increases, and the performance of MLE becomes increasingly stable.

The sensitivity of the estimators to tail heaviness was additionally evaluated under $\xi \in \{0.01, 0.2, 0.4\}$. The results are presented in Fig. 4. When $\xi = 0.01$, the GPD is close to an exponential distribution, and MOM provides competitive estimation performance. As ξ increases, the positive tail becomes heavier and the finite-variance condition becomes more restrictive. Consequently, the performance of MOM deteriorates as ξ approaches $1/2$. In contrast, the PWM/L-moment estimator remains relatively stable in moderately heavy-tailed and small-sample settings because it is less strongly affected by the largest observations.

Overall, the simulation results indicate that moment-based estimators provide useful alternatives when the number of positive exceedances is limited. MOM is particularly suitable for relatively light-tailed distributions satisfying $\xi < 1/2$, whereas PWM/L-moment estimation provides greater stability for moderately heavy-tailed and small-sample data. MLE becomes increasingly competitive as the sample size increases and remains advantageous for large samples because of its asymptotic efficiency and compatibility with likelihood-based dependence estimation.

In the empirical application, the dependence parameters are estimated using 4456 jointly exceeding observations. This represents a sufficiently large size of samples for which MLE is expected to provide stable and efficient estimates. Therefore, MLE is retained for the main application, while MOM and PWM/L-moment estimators are

considered practical alternatives when a small or moderate size of exceedances is available.

6. Application: Serial multi-stage processing time data of Li-ion batteries

In this section, we analyze serial multi-stage processing time data of li-ion batteries to validate the proposed zero-inflated multivariate model. This analysis was mainly performed in the context of process monitoring for battery manufacturing systems characterized by serial multi-stage production processes. The sequential dependency inherent in such manufacturing systems necessitates robust statistical frameworks for accurately capturing distributional characteristics of process delays, particularly when these delays exhibit both extreme values and frequent sub-threshold occurrences. The identification of bottleneck jobs in each process and their quantification represents a critical aspect of productivity, where the tardiest job dictates overall tact time for entire batch processes. For this purpose, statistical models for processing time data need to demonstrate the accuracy in quantifying both the frequency and the magnitude of the tardiness.

The battery manufacturing process under investigation consists of three sequential processing stages: cell assembly, electrolyte injection, and tab welding. Each stage consists of multiple individual jobs that must be completed prior to moving to the subsequent stage. Individual process follows a synchronous completion paradigm where the entire stage is considered to be complete right when the tardiest job

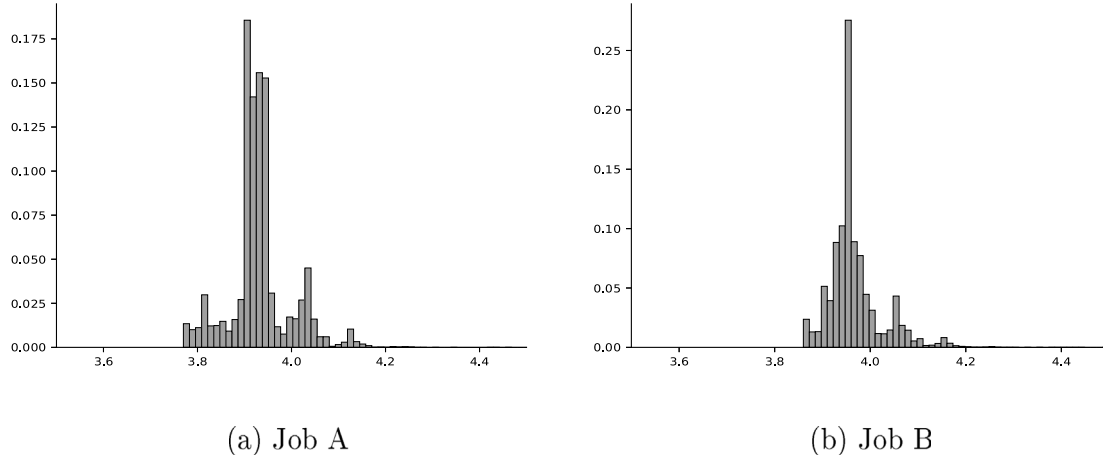


Fig. 5. Histograms of processing times for two critical tardy jobs in a cell assembly line.

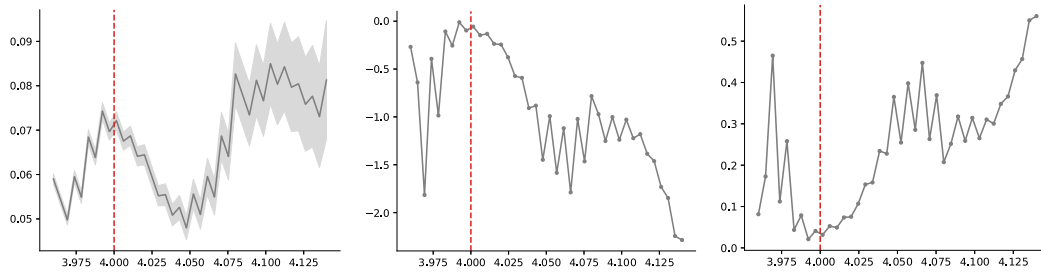


Fig. 6. MRL plot (left) and parameters sensitivity plots ($\hat{\sigma}_i$: middle, $\hat{\xi}_i$: right) for job A.

within that stage finishes its operations. For example, Fig. 5 presents histograms of processing times for most critical two jobs identified through process analysis in a cell assembly line. Most jobs at each stage are required to finish their operations before target processing time for a batch, e.g., 4 s, for productivity management. A great number of processing times of the two jobs exceed 4 s and they seem to be critical bottleneck jobs in the cell assembly line. Processing times less than the threshold are recorded as zeros and they are generally ignored in the PoT analysis. In addition, the processing time data exhibits non-Gaussian and heavy-tailed characteristics, showing that they are significant deviated from traditional Gaussian assumptions for multivariate data. These characteristics appear across multiple stages of the battery manufacturing process.

The threshold was fixed at $u = 4$ seconds because it corresponds to the tact-time of the production line at the data-providing company. Thus, the threshold is not a tuning parameter selected through a data-driven statistical procedure, but an operationally meaningful boundary that distinguishes within-tact completions from delayed processes. To examine the statistical behavior around the operational threshold, the plots of MRL and parameter-stability are provided in Figs. 6 and 7, respectively. In addition, a threshold-sensitivity analysis was conducted over $u \in \{3.90, 3.95, 4.00, 4.05, 4.10\}$. The results are depicted in Fig. 8. The estimated shape parameter is sensitive to the threshold near the tact-time. For example, for Process A, $\hat{\xi}$ changes from approximately -0.04 at $u = 3.90$ to 0.24 at $u = 4.10$, and the MRL plots are not strictly linear above $u = 4.00$. This behavior indicates that exceedances near the tact-time may include a mixture of within-tact completions and delayed right-tail observations.

A threshold-sensitivity analysis was conducted over $u \in \{3.90, 3.95, 4.00, 4.05, 4.10\}$. The results are presented in Fig. 8 and also summarized in Table 1. The estimates, particularly the shape parameter $\hat{\xi}$, vary across the candidate thresholds. For job A, $\hat{\xi}$ changes from

approximately -0.04 at $u = 3.90$ to 0.24 at $u = 4.10$. The threshold was retained at 4 s because changing it through statistical optimization would redefine the operational set of delayed processes that the monitoring framework is intended to manage. All of the analyses are conducted based on the threshold of 4 s hereinafter.

First, to compare modeling performance of the ZIMGPD proposed model, we fitted the PoT data over 4 s to the MGPD model, ignoring a number of zeros by eliminating processing times under 4 s in the model construction. We considered two types of MGPD models similar to the ZIMGPD models: Type I MGPD model and Type II MGPD model. By assuming the independence between variables as a conventional parametric structure for modeling extreme values, the Type I MGPD model is given by

$$f(\mathbf{x}; \xi, \sigma) = \prod_{i=1}^d \frac{1}{\sigma_i} \left(1 + \frac{\xi_i x_i}{\sigma_i} \right)^{-\frac{1}{\xi_i} - 1}, \quad x_i > 0$$

and considering tail-dependent structure among extremal values, the Type II MGPD model is given by

$$f(\mathbf{x}; \xi, \sigma, \alpha) = \left(\prod_{i < j} (1 - \alpha_{ij}^2) \prod_{i=1}^d x_i \right)^{-1} \left(\prod_{i=1}^d \left(\frac{\sigma_i}{x_i} \right)^{\xi_i} \right)^{\frac{1}{\prod_{i < j} (1 - \alpha_{ij}^2)}} \\ \times \prod_{i < j} \mathcal{J}_0 \left(\frac{2\sqrt{|\alpha_{ij} \xi_i \xi_j \log(x_i/\sigma_i) \log(x_j/\sigma_j)|}}{1 - \alpha_{ij}^2} \right), \quad x_i > 0.$$

The MLEs of the scale and shape parameters in the Type I MGPD models are: $\hat{\sigma} = (0.0717, 0.0493)^\top$ and $\hat{\xi} = (0.0256, 0.0942)^\top$, respectively. The MLEs of the parameters in the Type II MGPD models are: $\hat{\sigma} = (0.0846, 0.0454)^\top$, $\hat{\xi} = (0.0112, 0.1467)^\top$, and $\hat{\alpha}_{12} = 0.5977$. Fitting results of Type I MGPD and Type II MGPD models to each job are presented in Fig. 9. To quantitatively evaluate the modeling performance, we calculated root mean square error (RMSE), whose metric is defined as: $\text{RMSE} = \{(\mathbf{y} - \hat{\mathbf{y}})^2 / N\}^{1/2}$, where N is the total number of data, and $\mathbf{y}$ and

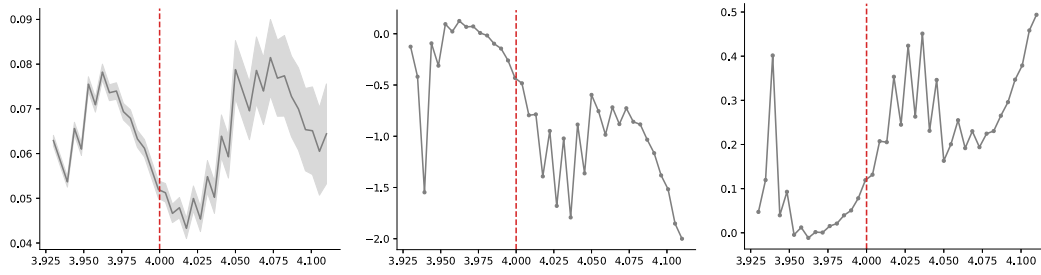
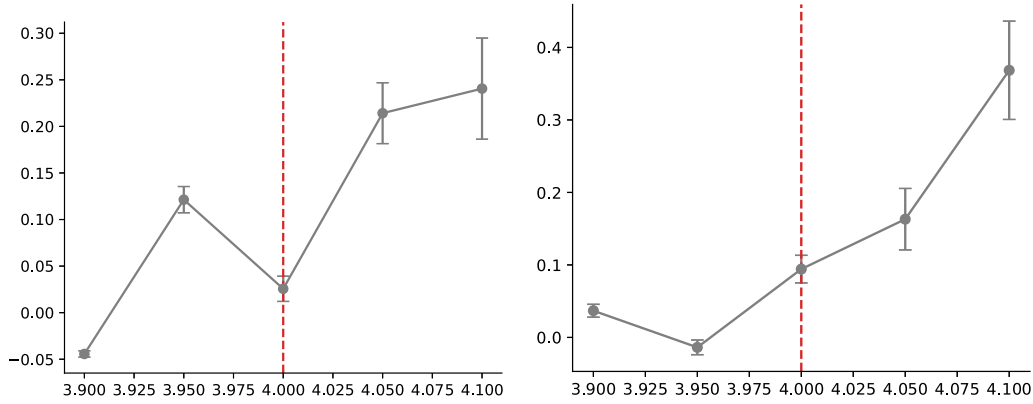
Fig. 7. MRL plot (left) and parameters sensitivity plots ($\hat{\sigma}_i$: middle, $\hat{\xi}_i$: right) for job B.

Fig. 8. Threshold sensitivity analysis for job A (left) and job B (right).

Table 1
Threshold sensitivity analysis for job A and job B.

u	Proc.	n_{exc}	$\hat{\sigma}$ (SE)	$\hat{\xi}$ (SE)	$\hat{\alpha}$
3.90	A	34,813	0.0784 (0.0004)	-0.0443 (0.0017)	0.616
3.90	B	28,458	0.0533 (0.0004)	0.0368 (0.0045)	
3.95	A	20,437	0.0460 (0.0005)	0.1213 (0.0072)	0.703
3.95	B	7796	0.0797 (0.0010)	-0.0137 (0.0052)	
4.00	A	6180	0.0717 (0.0010)	0.0256 (0.0069)	0.607
4.00	B	5273	0.0494 (0.0008)	0.0942 (0.0097)	
4.05	A	3352	0.0442 (0.0010)	0.2142 (0.0167)	0.800
4.05	B	1364	0.0643 (0.0022)	0.1631 (0.0216)	
4.10	A	1013	0.0641 (0.0026)	0.2406 (0.0277)	0.728
4.10	B	830	0.0374 (0.0018)	0.3685 (0.0346)	

$\hat{\mathbf{y}}$ are multivariate observations and their predicted values, respectively. RMSE values from the Type I and the Type II MGPD models are 1.2955 and 1.2742, respectively. The Type II MGPD model shows better performance by incorporating correlation structure between two jobs in the model. To statistically assess the significance of correlation term α_{12} in the Type II MGPD model, we employed the likelihood ratio test (LRT) procedure. Denote $\mathcal{L}_2$ as the likelihood for the Type II MGPD model, and $\mathcal{L}_1$ as the likelihood for the Type I MGPD model. Then under the null hypothesis that the Type I MGPD model is adequate, the likelihood ratio test (LRT) statistic $\Lambda = 2 \log(\mathcal{L}_2/\mathcal{L}_1) = 2(\ell_2 - \ell_1)$ will approximately follow the χ^2 distribution with 1 degree of freedom. Resulting LRT statistic is 2817.28 with the p -value smaller than 0.0001, confirming the Type II MGPD model in terms of the goodness-of-fit.

However, a number of zeros may provide additional information about the manufacturing process, thus the MGPD models suffer from informational loss by ignoring the zero data. To evaluate the effects of a number of zeros in multivariate PoT model, we fitted the two ZIMGPD models for the PoT data including zeros denoting process times less than 4 s to compare their modeling performance. The MLEs of the proportion of zeros, scale and shape parameters in the Type I

ZIMGPD models are: $\hat{\boldsymbol{\pi}} = (0.1649, 0.1403)^\top$, $\hat{\boldsymbol{\sigma}} = (0.0717, 0.0494)^\top$ and $\hat{\xi} = (0.0256, 0.0942)^\top$, respectively. The MLEs of the parameters in the Type II ZIMGPD models are: $\hat{\boldsymbol{\pi}} = (0.1649, 0.1403)^\top$, $\hat{\boldsymbol{\sigma}} = (0.0761, 0.0412)^\top$, $\hat{\xi} = (0.0186, 0.1609)^\top$, and $\hat{\alpha}_{12} = 0.6007$. Fitting results from Type I ZIMGPD and Type II ZIMGPD models are presented in Fig. 10. To test the significance of correlation term α_{12} in the Type II ZIMGPD model, we calculated the LRT statistic for the two ZIMGPD models. Similarly, by denoting $\mathcal{L}_2$ as the likelihood for the Type II ZIMGPD model, and $\mathcal{L}_1$ as the likelihood for the Type I ZIMGPD model, resulting LRT statistic is $\Lambda = 12345.40$ with the p -value smaller than 0.0001, recommending the Type II ZIMGPD model as the better candidate model.

To compare the MGPD models with the ZIMGPD models, we fitted the MGPD and ZIMGPD models to processing time data including zeros. Their fitting results are plotted in Fig. 11 for Type I model and Fig. 12 for Type II model, respectively. Figs. 11 and 12 show that the MGPD models tend to highly over-estimate near zero density regions by ignoring zero-inflation effects. RMSE values from the Type I MGPD and the Type I ZIMGPD models are 6.4624 and 0.2025, respectively. RMSE values were also calculated for the Type II MGPD and the Type II ZIMGPD models, with their respective values of 6.4715 and 0.1970.

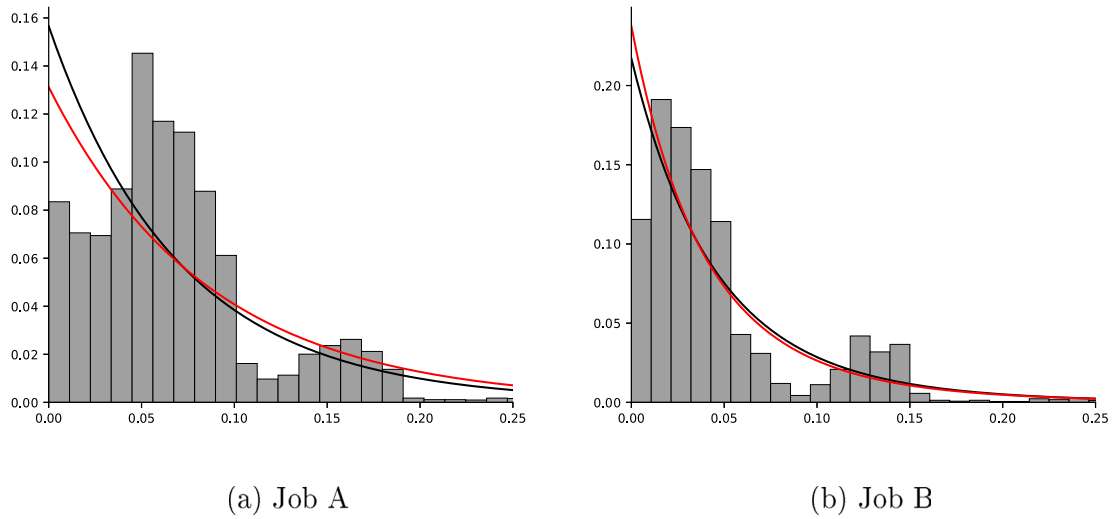


Fig. 9. Fitting results of the Type I MGPD model (black line) and the Type II MGPD model (red line) to each job.

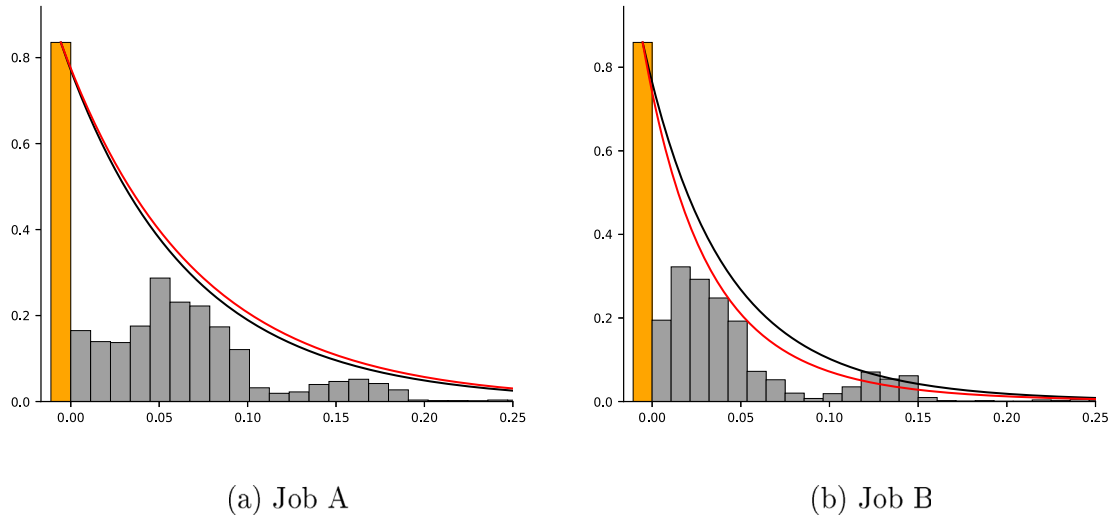


Fig. 10. Fitting results for the Type I ZIMGPD model (black line) and the Type II ZIMGPD model (red line) (Orange bars represent the densities of zeros).

Overall, when compared with the MGPD models in terms of RMSE, the ZIMGPD models significantly improve modeling performance for manufacturing process data by incorporating a number of zeros into the model. The zero-inflation provides additional information about the characteristic of processing times in PoT applications to manufacturing systems. It is also noted that the Type II ZIMGPD model has the best results in fitting processing times by simultaneously capturing the dependency structure and zero-inflated nature of processing time data. The ZIMGPD models enhance the identification and quantification of bottleneck jobs in manufacturing process, enabling more targeted interventions aimed at reducing tact times and optimizing production flow.

To quantitatively access the modeling accuracy, we constructed asymptotic CIs for the MLEs of ZIMGPD model parameters and the results are given in Table 2. For example, the CIs from the MGPD and ZIMGPD models of Type II are plotted in Fig. 13. The Type II MGPD CIs are constructed based on the processing time data excluding zeros for comparison. The Type II ZIMGPD model shows narrower confidence bands than the Type II MGPD model near zero domain, indicating enhanced accuracy in capturing the zero-inflated structure inherent in processing time data. Wider confidence bands near zero values are much apparent when we fit the Type II MGPD models to processing

times including zeros. We observed similar patterns in the MGPD and ZIMGPD models of Type I.

To further evaluate the fitting performance of the proposed model, the comparison was extended to include the independent MGPD and the Clayton copula-MGPD models. For a fair likelihood-based comparison, the competing models were evaluated on the common support consisting of the $n = 4456$ observations for which both processes exceeded the 4-second threshold. The log-likelihood, Akaike information criterion (AIC), and Bayesian information criterion (BIC) are reported in Table 3.

The likelihood-ratio test comparing the independent MGPD and the Clayton copula-MGPD produced a statistic of 2594.8 with 1 degree of freedom, yielding $p < 0.001$. The Vuong statistic was $z = 30.0$, also favoring the dependent model. The estimated Clayton dependence corresponded to Kendall's $\tau = 0.58$. A semi-parametric multivariate extreme model was also considered. In this model, the marginal distributions were estimated using empirical rank-based distributions, while only the extremal dependence structure was modeled parametrically. Because the effective number of parameters associated with the non-parametric margins is not uniquely defined, its AIC and BIC are not directly comparable with those of the fully parametric models. The parametric and semi-parametric approaches were therefore compared

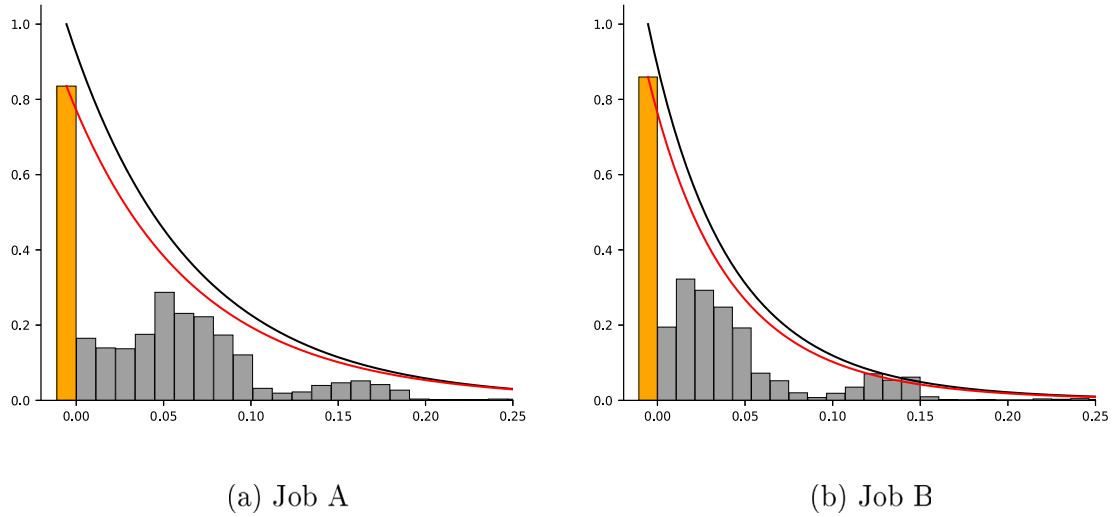


Fig. 11. Fitting results for the Type I MGPD model (black line) and the Type I ZIMGPD model (red line).

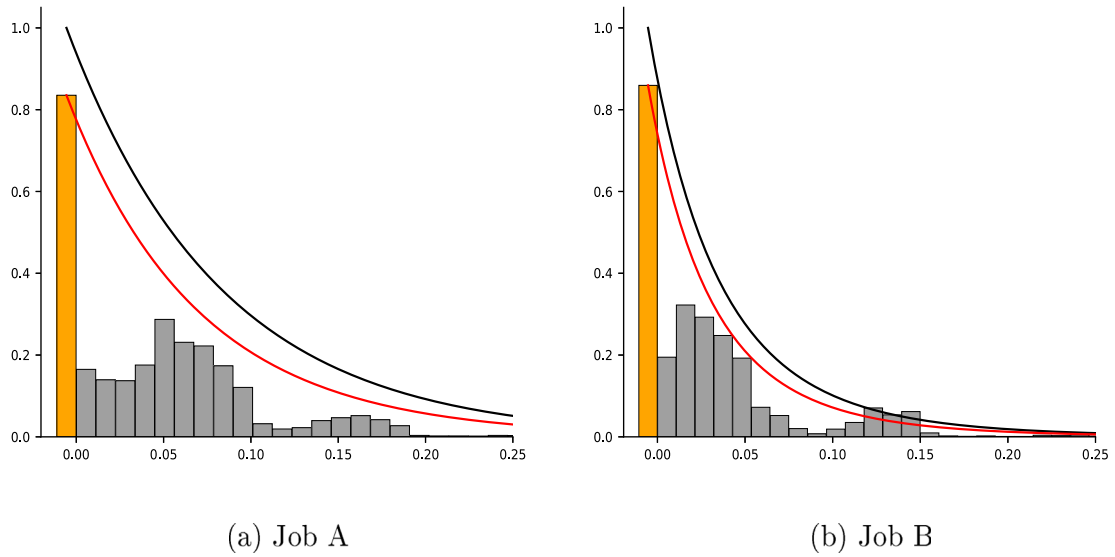


Fig. 12. Fitting results for the Type II MGPD model (black line) and the Type II ZIMGPD model (red line).

Table 2
95% confidence intervals for estimated parameters.

Model	Job	$\hat{\pi}$	$\hat{\sigma}$	$\hat{\xi}$
ZiMGPD Type I	A	(0.1639, 0.1685)	(0.0697, 0.0738)	(0.0121, 0.0392)
	B	(0.1370, 0.1440)	(0.0477, 0.0510)	(0.0752, 0.1133)
ZiMGPD Type II	A	(0.1639, 0.1685)	(0.0740, 0.0782)	(0.0050, 0.0322)
	B	(0.1370, 0.1440)	(0.0399, 0.0426)	(0.1368, 0.1849)

Table 3
Goodness of fit test for MGPD, Clayton Copula-based MGPD, and ZIMGPD.

Model	log-Lik	k	AIC	BIC
MGPD	15,246.9	4	-30,485.9	-30,460.3
Copula-MGPD (Clayton)	16,544.3	5	-33,078.6	-33,046.6
ZIMGPD	21,388.7	5	-42,767.5	-42,735.3

using Kendall's τ , which is invariant to monotone marginal transformations. The estimated Kendall's τ values were 0.45 and 0.40 for the parametric and semi-parametric models, respectively, compared with the empirical value of 0.45. The semi-parametric model provides greater robustness to marginal misspecification, whereas the proposed

model yields interpretable marginal and dependence parameters that are directly useful for process monitoring.

Fig. 14 presents the excess quantile–quantile plots for Processes A and B. The fitted distributions adequately capture the central portion of the exceedance distributions, whereas departures are observed in the

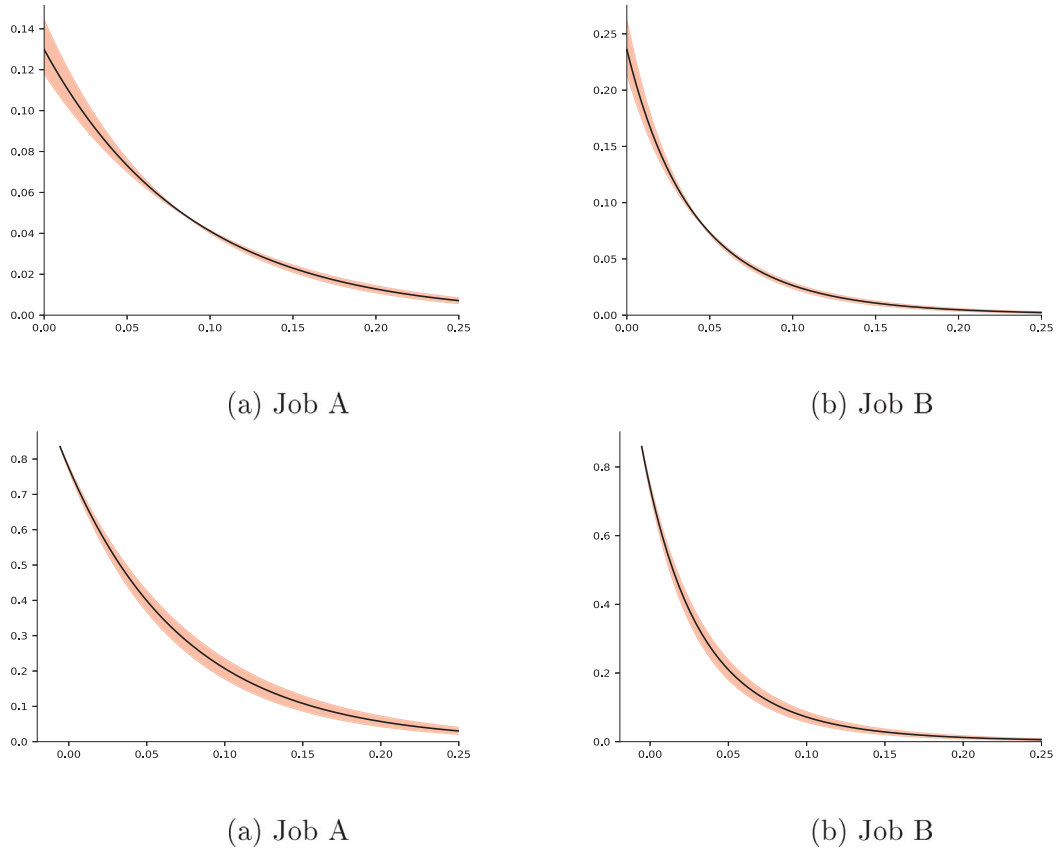


Fig. 13. 95% CIs of the Type II MGPD model (upper) and the Type II ZIMGPD model (lower).

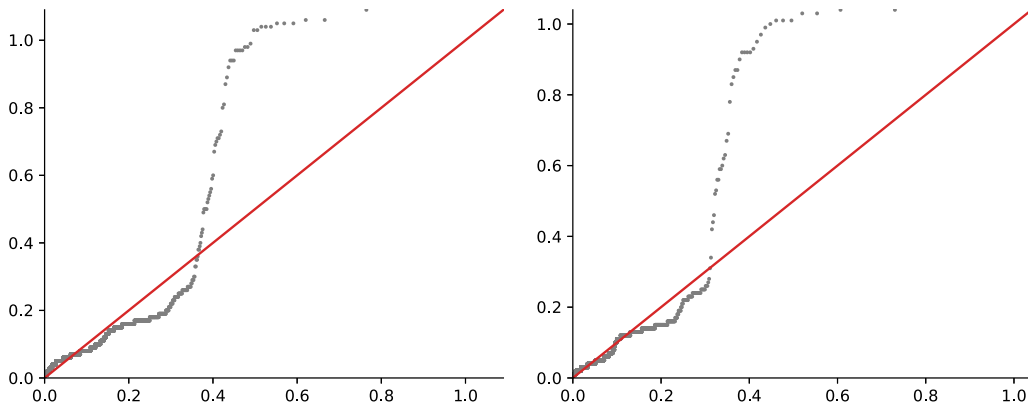


Fig. 14. Excess QQ-plot for job A (left) and job B (right). The straight lines indicate GPD.

most extreme tail region. Exceedances greater than 0.3 s account for less than 1% of the positive observations, indicating that the apparent tail deviation is based on a highly sparse region of the data.

The proposed model reproduces the observed zero counts exactly because the zero-inflation probabilities are estimated directly from the discrete likelihood component. In contrast, the MGPD assigns zero probability to observations at zero. Consequently, its full-sample log-likelihood and Vuong statistic are not well defined when the zero observations are retained. The competing models were therefore additionally evaluated using proper scoring rules that remain finite for mixed discrete-continuous distributions.

The residual serial dependence was examined using the autocorrelation function and the Ljung–Box statistic, defined as $Q(m) = n(n +$

$2) \sum_{h=1}^m \hat{\rho}_h^2 / (n - h)$, where $\hat{\rho}_h$ denotes the residual autocorrelation at lag h . As shown in Fig. 15, several short-lag autocorrelations remain statistically significant, indicating that consecutive process delays may be affected by temporary production congestion. The Ljung–Box portmanteau test to the probability-integral-transform residuals $\{r_t\}_{t=1}^m$ of the fitted model was conducted. Let

$$\hat{\rho}_k = \frac{\sum_{t=1}^{m-k} (r_t - \bar{r})(r_{t+k} - \bar{r})}{\sum_{t=1}^m (r_t - \bar{r})^2} \quad (30)$$

denote the lag- k sample autocorrelation of the residuals. The Ljung–Box statistic for the first h lags is

$$Q(h) = m(m+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{m-k} \stackrel{H_0}{\sim} \chi_h^2, \quad (31)$$

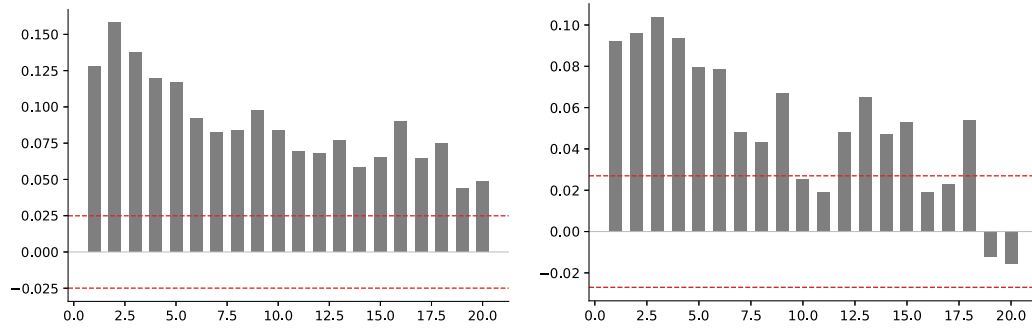


Fig. 15. Residual Autocorrelation ACF plots for job A (left) and job B (right).

where the null hypothesis H_0 is that the residuals are white noise (no serial dependence up to lag h). A large value of $Q(h)$ relative to the χ_h^2 distribution indicates that the residuals retain significant autocorrelation that the model has not captured. The test decisively rejects H_0 : with $h = 10$ obtained $Q(10) = 785.2$ for process A and $Q(10) = 311.6$ for process B, both values significantly exceed the critical value of $\chi_{10,0.95}^2 \approx 18.3$ ($p < 0.001$). The rejection remains consistent when $h = 20$, with values of $Q(20) = 1064.2$ and 396.9 , respectively. The residual autocorrelations in Fig. 15 are positive at short lags (lag-1 $\hat{\rho}_1 = 0.13$ for process A and 0.09 for process B), indicating that a delay in one cycle tends to be followed by further delays. It represents a physically plausible congestion effect in a synchronized line, where a disturbance propagates over several subsequent cycles rather than dissipating immediately.

Overall, the diagnostic results indicate that the proposed model adequately represents the zero-inflated marginal behavior, the main body of the positive exceedance distributions, and the dependence between the two processes. The remaining departures are concentrated in the sparsely observed extreme-tail region and in short-range serial dependence.

7. Conclusion & future work

In the environments where the progression to subsequent processes occurs only when every job has completed the current operation, accurate modeling of both extreme delays and sub-threshold process times is essential for comprehensive process understanding and optimization. This study proposes the ZIMGPD models as a robust extension to the traditional PoT models for data scenarios that exhibit a significant proportion of zeros alongside extreme values. By incorporating a zero-inflation component, the ZIMGPD models address notable challenge of excessive zero observations commonly observed in modern manufacturing processes, such as high-frequency sub-threshold production times or equipment downtime. Two distinct variants of the ZIMGPD model, namely, Type I and Type II models, are developed to capture independent and dependent extreme behaviors, respectively. Type I ZIMGPD model extends the MGPD by incorporating zero-inflation independently into each marginal distribution, whereas Type II ZIMGPD model further accounts for tail dependence through a modified Bessel function. We develop theoretical derivation of MLEs for the two ZIMGPD model parameters based on their asymptotic consistency and normality.

The proposed multivariate zero-inflated models are empirically validated via real-world battery manufacturing process data. The ZIMGPD models significantly outperforms conventional MGPD models in the situations with structural zero-inflation. In particular, the ZIMGPD models substantially improve modeling accuracy in the vicinity of zero, whereas the MGPD models exhibit systematic bias in capturing the probability mass at zero. By more accurately characterizing the full distribution of sub-threshold process times, the proposed method offers tangible advantages for industrial process optimization, particularly in

identifying and mitigating bottlenecks arising from both nominal operating conditions and intermittent extreme delays. This study aims to bridge a critical gap in extreme value theory by offering a robust framework for process control applications in manufacturing environments characterized by serial dependency and batch processing constraints. The demonstrated flexibility of the ZIMGPD models, combined with its solid theoretical foundation and empirical efficacy, underscores its potential in improving both reliability and efficiency in next-generation manufacturing systems.

Notwithstanding its demonstrated advantages, the proposed approach still presents certain limitations and opportunities for future research. For high-dimensional cases, the numerical burden of estimating the tail dependency parameters in the presence of Bessel function terms can become computationally intensive, suggesting the need for scalable algorithms or approximate inference techniques. Moreover, the residual serial dependence observed in the application indicates that the current static formulation may not fully capture temporary congestion and delay-propagation effects, motivating extensions to dynamic or time-varying dependence structures and more intricate zero-inflation mechanisms. Finally, integrating these modeling insights with real-time monitoring or predictive maintenance can further enhance automated quality control and throughput optimization in advanced manufacturing environments.

CRedit authorship contribution statement

Hwi Jun Jung: Writing – original draft, Software, Methodology, Formal analysis, Data curation. **Suk Joo Bae:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00407803) and Korea Environment Industry & Technology Institute (KEITI) through Technology R&D project for Recycling Core Components of Hydrogen Fuel Cell Vehicles, funded by Korea Ministry of Climate, Energy and Environment (MCEE) (grant number: RS-2026-25506226).

Appendix A. Proof of Lemma 1

Proof. Following the results from Kibble (1941) and Krishnamoorthy and Parthasarathy (1951), we can express the dependence parameter α_{ij} in terms of the transformed variables as

$$\alpha'_{ij} = \frac{\text{Cov}(q_{ik}, q_{jk})}{\sqrt{\text{Var}(q_{ik}) \cdot \text{Var}(q_{jk})}}.$$

When σ_i changes by a small amount $\Delta\sigma_i$, using the partial derivative $\frac{\partial q_{ik}}{\partial \sigma_i} = -\frac{\xi_i}{\sigma_i}$, the first-order Taylor approximation of q_{ik} gives:

$$q_{ik}(\sigma_i + \Delta\sigma_i) \approx q_{ik} \cdot \sigma_i - \frac{\xi_i}{\sigma_i} \Delta\sigma_i.$$

This transformation affects all observations for the i th variates by a factor proportional to $\Delta\sigma_i$. Although this is not a simple additive shift, it preserves correlation structure since for any two random variables, adding a constant to one variable does not change their correlation. Specifically:

$$\begin{aligned} \text{Cov}(q_{ik}(\sigma_i + \Delta\sigma_i), q_{jk}) &= \text{Cov}\left(q_{ik} \cdot \sigma_i - \frac{\xi_i}{\sigma_i} \Delta\sigma_i, q_{jk}\right) \\ &= \text{Cov}(q_{ik} \cdot \sigma_i, q_{jk}). \end{aligned}$$

Similarly, for the variance term

$$\begin{aligned} \text{Var}(q_{ik}(\sigma_i + \Delta\sigma_i)) &= \text{Var}\left(q_{ik} \cdot \sigma_i - \frac{\xi_i}{\sigma_i} \Delta\sigma_i\right) \\ &= \text{Var}(q_{ik} \cdot \sigma_i). \end{aligned}$$

When ξ_i changes by a small amount $\Delta\xi_i$, the transformation of q_{ik} is multiplicative as

$$q_{ik}(\xi_i + \Delta\xi_i) = (\xi_i + \Delta\xi_i) \log\left(\frac{x_{ik}}{\sigma_i}\right) = \left(1 + \frac{\Delta\xi_i}{\xi_i}\right) \cdot q_{ik}(\xi_i).$$

This represents a scale transformation with factor $c = \left(1 + \frac{\Delta\xi_i}{\xi_i}\right)$. The correlation coefficient is invariant under such scale transformations because $\text{Var}(c \cdot q_{ik}) = c^2 \cdot \text{Var}(q_{ik})$ and $\text{Cov}(c \cdot q_{ik}, q_{jk}) = c \cdot \text{Cov}(q_{ik}, q_{jk})$. Substituting these into the correlation formula:

$$\alpha'_{ij} = \frac{c \cdot \text{Cov}(q_{ik}, q_{jk})}{\sqrt{c^2 \cdot \text{Var}(q_{ik})} \cdot \sqrt{\text{Var}(q_{jk})}} = \frac{\text{Cov}(q_{ik}, q_{jk})}{\sqrt{\text{Var}(q_{ik})} \cdot \sqrt{\text{Var}(q_{jk})}} = \alpha_{ij}.$$

Therefore, the correlation parameter α_{ij} remains invariant under changes in both σ_i and ξ_i . ■

Appendix B. Derivations of score functions

Before calculating the score function of each parameter, we must first know the partial derivatives of each parameter with respect to $z_{ijk} \cdot \sigma_i$ and ξ_i must be calculated based on the chain rule using the substituted q_{ik} . The partial derivatives of z_{ijk} with respect to σ_i and ξ_i calculated through the chain rule applied to each parameter are as follows:

$$\begin{aligned} \frac{\partial z_{ijk}}{\partial \sigma_i} &= \frac{\partial z_{ijk}}{\partial q_{ik}} \cdot \frac{\partial q_{ik}}{\partial \sigma_i} = -q_{jk} \cdot \alpha_{ij} \cdot \frac{\xi_i}{\sigma_i} = -\alpha_{ij} \xi_j \left(\frac{\xi_i}{\sigma_i}\right) \log\left(\frac{x_{jk}}{\sigma_j}\right), \\ \frac{\partial z_{ijk}}{\partial \xi_i} &= \frac{\partial z_{ijk}}{\partial q_{ik}} \cdot \frac{\partial q_{ik}}{\partial \xi_i} = q_{jk} \cdot \alpha_{ij} \cdot \log\left(\frac{x_{ik}}{\sigma_i}\right) = \alpha_{ij} \xi_j \log\left(\frac{x_{ik}}{\sigma_i}\right) \log\left(\frac{x_{jk}}{\sigma_j}\right). \end{aligned}$$

The score function of $\frac{\partial \ell}{\partial \alpha_{ij}}$ is derived by applying the chain rule to differentiate the log-likelihood function. This involves that the differentiations of $\log(1 - \alpha_{ij}^2)$ and the normalization term $\{\prod_{i < j} (1 - \alpha_{ij}^2)\}^{-1}$, and the application of the chain rule to the Bessel function terms $\log J_0(z_{ijk})$. We use the relationship $\frac{d}{dx} J_0(x) = -J_1(x)$ and carefully differentiate z_{ijk} with respect to α_{ij} as

$$\begin{aligned} \frac{\partial z_{ijk}}{\partial \alpha_{ij}} &= \frac{\partial}{\partial \alpha_{ij}} (\alpha_{ij} \cdot q_{ik} \cdot q_{jk}) = q_{ik} \cdot q_{jk} \\ &= \xi_i \log\left(\frac{x_{ik}}{\sigma_i}\right) \cdot \xi_j \log\left(\frac{x_{jk}}{\sigma_j}\right). \end{aligned}$$

Appendix C. Differentiability condition

The marginal terms $w_i^{-\frac{1}{\xi_i}-1}$ are three times continuously differentiable with respect to ξ_i for $w_i > 0$ such that

$$\frac{\partial}{\partial \xi_i} w_i^{-\frac{1}{\xi_i}-1} = w_i^{-\frac{1}{\xi_i}-1} \cdot \frac{\log w_i}{\xi_i^2},$$

$$\begin{aligned} \frac{\partial^2}{\partial \xi_i^2} w_i^{-\frac{1}{\xi_i}-1} &= w_i^{-\frac{1}{\xi_i}-1} \cdot \left(\frac{(\log w_i)^2}{\xi_i^4} - \frac{2 \log w_i}{\xi_i^3} \right), \\ \frac{\partial^3}{\partial \xi_i^3} w_i^{-\frac{1}{\xi_i}-1} &= w_i^{-\frac{1}{\xi_i}-1} \cdot \left(\frac{(\log w_i)^3}{\xi_i^6} - \frac{6(\log w_i)^2}{\xi_i^5} + \frac{6 \log w_i}{\xi_i^4} \right). \end{aligned}$$

These derivatives exist and are continuous for all w_i 's and $\xi_i \neq 0$. In addition, the modified Bessel function J is analytic, making the dependence structure terms infinitely differentiable with respect to α_{ij} , ξ_i , and ξ_j , $\forall i, j$, since $\frac{d}{dz} \log J_0(z) = \frac{J_1(z)}{J_0(z)}$. Finally, the angular density $g(\mathbf{w}; \boldsymbol{\theta}_2)$ is three times continuously differentiable with respect to the parameter set $\boldsymbol{\theta}_2$.

Appendix D. Dominated derivatives condition

We will show that the derivatives of $\log g(\mathbf{w}; \boldsymbol{\theta}_2)$ with respect to each component of $\boldsymbol{\theta}_2$ are bounded by integrable functions over the unit simplex S_k . Specifically,

$$\frac{\partial \log g(\mathbf{w}; \boldsymbol{\theta}_2)}{\partial \xi_i} = \frac{1}{\xi_i^2} \log w_i + \sum_{j \neq i} \frac{\partial}{\partial \xi_i} \log J_0 \left(\frac{2\sqrt{|\alpha_{ij} \xi_i \xi_j \log(w_i) \log(w_j)|}}{1 - \alpha_{ij}^2} \right).$$

For the modified Bessel term, using the property that $\frac{d}{dz} \log J_0(z) = \frac{J_1(z)}{J_0(z)}$ and the bound $\frac{J_1(z)}{J_0(z)} \leq \frac{z}{2} + \sqrt{\frac{z^2}{4} + 1}$ for $z > 0$, we obtain

$$\left| \frac{\partial \log g(\mathbf{w}; \boldsymbol{\theta}_2)}{\partial \xi_i} \right| \leq \frac{|\log w_i|}{\xi_i^2} + C_1 \cdot \sum_{j \neq i} \sqrt{|\log w_i| \cdot |\log w_j|}.$$

This is dominated by an integrable function on S_k because

$$\int_{S_d} |\log w_i| \cdot w_i^{-\frac{1}{\xi_i}-1} dw < \infty, \text{ and } \int_{S_d} \sqrt{|\log w_i| \cdot |\log w_j|} w_i^{-\frac{1}{\xi_i}-1} w_j^{-\frac{1}{\xi_j}-1} dw < \infty,$$

for $\xi_i, \xi_j > -\frac{1}{2}$. For the dependence parameters α_{ij} , we have

$$\frac{\partial \log g(\mathbf{w}; \boldsymbol{\theta}_2)}{\partial \alpha_{ij}} = \frac{2\alpha_{ij}}{1 - \alpha_{ij}^2} + \frac{\partial}{\partial \alpha_{ij}} \log J_0 \left(\frac{2\sqrt{|\alpha_{ij} \xi_i \xi_j \log(w_i) \log(w_j)|}}{1 - \alpha_{ij}^2} \right),$$

and the derivatives of J_0 are bounded as

$$\left| \frac{\partial}{\partial \alpha_{ij}} \log J_0 \left(\frac{2\sqrt{|\alpha_{ij} \xi_i \xi_j \log(w_i) \log(w_j)|}}{1 - \alpha_{ij}^2} \right) \right| \leq C_2 \cdot \frac{\sqrt{|\log w_i| \cdot |\log w_j|}}{1 - \alpha_{ij}^2}.$$

Similar bounds can be established for the derivatives with respect to other parameters as

$$\left| \frac{\partial \log g(\mathbf{w}; \boldsymbol{\theta}_2)}{\partial \theta_{2,d}} \right| \leq C (|\log w_i| + |\log w_j| + 1),$$

which is integrable on S_d due to the polynomial decay of $w_i^{-\frac{1}{\xi_i}-1}$ and controlled growth of the modified Bessel function terms.

Appendix E. Fisher information condition

The Fisher information matrix $\mathcal{I}(\boldsymbol{\theta}_2)$ is defined as

$$\begin{aligned} \mathcal{I}(\boldsymbol{\theta}_2) &= \mathbb{E}_{\boldsymbol{\theta}_2} \left[\nabla_{\boldsymbol{\theta}_2} \log g(\mathbf{w}; \boldsymbol{\theta}_2) \nabla_{\boldsymbol{\theta}_2}^T \log g(\mathbf{w}; \boldsymbol{\theta}_2) \right] \\ &= \int_{S_d} \nabla_{\boldsymbol{\theta}_2} \log g(\mathbf{w}; \boldsymbol{\theta}_2) \nabla_{\boldsymbol{\theta}_2}^T \log g(\mathbf{w}; \boldsymbol{\theta}_2) \cdot g(\mathbf{w}; \boldsymbol{\theta}_2) d\mathbf{w}. \end{aligned}$$

For the Fisher information matrix to be finite, we need the condition that second moments of the score functions are integrable. The diagonal elements of the Fisher information matrix for ξ_i are

$$\begin{aligned} \mathcal{I}_{ii}(\boldsymbol{\theta}_2) &= \mathbb{E}_{\boldsymbol{\theta}_2} \left[\left(\frac{\partial}{\partial \xi_i} \log g(\mathbf{w}; \boldsymbol{\theta}_2) \right)^2 \right] \\ &= \mathbb{E}_{\boldsymbol{\theta}_2} \left[\left(\frac{\log w_i}{\xi_i^2} + \sum_{j \neq i} \frac{\partial}{\partial \xi_i} \log J_0(\cdot) \right)^2 \right]. \end{aligned}$$

For $\xi_i > -\frac{1}{2}$, the integrability of $\int_{S_d} \left| \frac{\partial}{\partial \xi_i} \log g(\mathbf{w}; \boldsymbol{\Theta}_2) \right|^2 \cdot g(\mathbf{w}; \boldsymbol{\Theta}_2) d\mathbf{w} < \infty$ ensures that the diagonal elements of $I(\boldsymbol{\Theta}_2)$ are finite. For the off-diagonal elements, we have

$$\begin{aligned} I_{ij}(\boldsymbol{\Theta}_2) &= \mathbb{E}_{\boldsymbol{\Theta}_2} \left[\frac{\partial}{\partial \xi_i} \log g(\mathbf{w}; \boldsymbol{\Theta}_2) \frac{\partial}{\partial \xi_j} \log g(\mathbf{w}; \boldsymbol{\Theta}_2) \right] \\ &= \int_{S_d} \frac{\partial}{\partial \xi_i} \log g(\mathbf{w}; \boldsymbol{\Theta}_2) \frac{\partial}{\partial \xi_j} \log g(\mathbf{w}; \boldsymbol{\Theta}_2) \cdot g(\mathbf{w}; \boldsymbol{\Theta}_2) d\mathbf{w}. \end{aligned}$$

The cross-derivatives involve the terms such as $\frac{\log w_i}{\xi_i^2} \cdot \frac{\log w_j}{\xi_j^2} +$ Bessel cross-terms. These are integrable when ξ_i and $\xi_j > -\frac{1}{2}$ since

$$\int_{S_d} |\log w_i| \cdot |\log w_j| w_i^{-\frac{1}{\xi_i}-1} w_j^{-\frac{1}{\xi_j}-1} d\mathbf{w} < \infty.$$

Finally, the Fisher information matrix $I(\boldsymbol{\Theta}_2)$ exists and is positive definite, which ensures that the polynomial decay in w_i terms guarantees finite second moments of derivatives; the sub-exponential growth of the modified Bessel function J_0 controls the dependence structure contributions; and $I(\boldsymbol{\Theta}_2)$ is invertible due to the identifiability of the parameters in the Type II ZIMGPD model.

Appendix F. Convexity of Pickands function

To ensure the validity of the Pickands dependence function $A(\mathbf{w}; \lambda)$, we need to verify that $A(\mathbf{w}; \boldsymbol{\Theta}_2)$ is convex on S_d and $A(\mathbf{w}; \boldsymbol{\Theta}_2) \geq \max w_i$ for all $\mathbf{w} \in S_d$. For the Type II ZIMGPD model with the Bessel-type dependence, we can establish the convexity of $A(\mathbf{w}; \boldsymbol{\Theta}_2)$ by considering the second derivatives with respect to w_i

$$\frac{\partial^2 A(\mathbf{w}; \boldsymbol{\Theta}_2)}{\partial w_i \partial w_j} = \int_{S_d} \frac{\partial^2}{\partial w_i \partial w_j} \max_{k=1, \dots, d} \left(\frac{w_k}{v_k} \right) g(\mathbf{v}; \boldsymbol{\Theta}_2) d\mathbf{v}.$$

For the bivariate case, we can show that

$$\frac{d^2 A(w)}{dw^2} = \frac{2}{B(\xi_1, \xi_2, \alpha_{12})} w^{-\frac{1}{\xi_1}-2} (1-w)^{-\frac{1}{\xi_2}-2} J_0(\cdot),$$

$$\text{where } B(\xi_1, \xi_2, \alpha_{12}) = \int_0^1 v^{-\frac{1}{\xi_1}-1} (1-v)^{-\frac{1}{\xi_2}-1} J_0 \left(\frac{2\sqrt{\alpha_{12}\xi_1\xi_2}\log(v)\log(1-v)}{1-\alpha_{12}^2} \right) dv$$

is the normalizing constant satisfying $\int_0^1 \dots \int_0^1 g(\mathbf{v}; \boldsymbol{\Theta}_2) d\mathbf{v} = 1$. This is non-negative for all i and j under the constraints:

- Second order differentiation of $A(\mathbf{w}; \boldsymbol{\Theta}_2)$ must be possible, so $\xi_i > -\frac{1}{2}$ for all i .
- Proper dependence correlation satisfies $|\alpha_{ij}| < 1$.
- The modified Bessel function $J_0(z) > 0$ ensures the positivity.

By extending the bivariate case to multivariate case, we can easily show that the Hessian matrix of $A(\mathbf{w}; \boldsymbol{\Theta}_2)$ is positive semi-definite when $\xi_i > -\frac{1}{2}$ for all i 's, ensuring the convexity on S_d . The lower bound $A(\mathbf{w}; \boldsymbol{\Theta}_2) \geq \max_i w_i$ follows from the definition of the Pickands function and the properties of the max operation in the integral representation.

Data availability

The authors do not have permission to share data.

References

- Angers, J.-F., & Biswas, A. (2003). A Bayesian analysis of zero-inflated generalized Poisson model. *Computational Statistics & Data Analysis*, 42(1–2), 37–46. [https://doi.org/10.1016/S0167-9473\(02\)00154-8](https://doi.org/10.1016/S0167-9473(02)00154-8).
- Arnold, B. C. (1990). A flexible family of multivariate Pareto distributions. *Journal of Statistical Planning and Inference*, 24(2), 249–258. [https://doi.org/10.1016/0378-3758\(90\)90045-V](https://doi.org/10.1016/0378-3758(90)90045-V).
- Arnold, B. C. (2015). *Pareto distributions* (2nd ed.). Boca Raton, FL: CRC Press, <https://doi.org/10.1201/b18141>.
- Bacro, J.-N., Gaetan, C., Opitz, T., & Toulemonde, G. (2025). Multivariate peaks-over-threshold with latent variable representations of generalized Pareto vectors. *Extremes*, 28(2), 273–300. <https://doi.org/10.1007/s10687-024-00503-2>.
- Bae, S. J., Hwang, J. Y., & Kuo, W. (2007). Yield prediction via spatial modeling of clustered defect counts across a wafer map. *IEEE Transactions*, 39(12), 1073–1083. <https://doi.org/10.1080/07408170701275335>.
- Belzile, L. R., & Davison, A. C. (2022). Improved inference on risk measures for univariate extremes. *The Annals of Applied Statistics*, 16(3), 1524–1549. <https://doi.org/10.1214/21-AOAS1555>.
- Boyce, B. L., Salzbrenner, B. C., Rodelas, J. M., Swiler, L. P., Madison, J. D., Jared, B. H., & Shen, Y. L. (2017). Extreme-value statistics reveal rare failure-critical defects in additive manufacturing. *Advanced Engineering Materials*, 19(8), Article 1700102. <https://doi.org/10.1002/adem.201700102>.
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London, UK.: Springer, <https://doi.org/10.1007/978-1-4471-3675-0>.
- Coles, S. G., & Tawn, J. A. (2005). Seasonal effects of extreme surges. *Stochastic Environmental Research and Risk Assessment*, 19(6), 417–427. <https://doi.org/10.1007/s00477-005-0008-3>.
- Couturier, D.-L., & Victoria-Feser, M.-P. (2010). Zero-inflated truncated generalized Pareto distribution for the analysis of radio audience data. *The Annals of Applied Statistics*, 4(4), 1824–1846. <https://doi.org/10.1214/10-AOAS358>.
- Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 52(3), 393–442. <https://doi.org/10.1111/j.2517-6161.1990.tb01796.x>.
- de Haan, L., & Ferreira, A. (2006). *Extreme value theory: An introduction*. New York, NY: Springer, <https://doi.org/10.1007/0-387-34471-3>.
- Eastoe, E. F., & Tawn, J. A. (2012). Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. *Biometrika*, 99(1), 43–55. <https://doi.org/10.1093/biomet/asr078>.
- Kibble, W. F. (1941). A two-variate gamma type distribution. *Sankhya*, 5, 137–150.
- Krishnamoorthy, A. S., & Parthasarathy, M. (1951). A multivariate gamma-type distribution. *The Annals of Mathematical Statistics*, 22(4), 549–557. <https://doi.org/10.1214/aoms/1177729544>.
- Kulkarni, H. V., & Patil, K. P. (2018). Two sample comparisons including zero-inflated continuous data: a parametric approach with applications to microarray experiment. *Mathematical Biosciences*, 298, 19–28. <https://doi.org/10.1016/j.mbs.2018.01.009>.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1–14. <https://doi.org/10.1080/00401706.1992.10485228>.
- Lee, D.-H., Kim, S.-H., & Kim, K.-J. (2021). Multistage MR-CART: multiresponse optimization in a multistage process using a classification and regression tree method. *Computers & Industrial Engineering*, 159, Article 107513. <https://doi.org/10.1016/j.cie.2021.107513>.
- Mardia, K. V. (1962). Multivariate Pareto distributions. *The Annals of Mathematical Statistics*, 33(3), 1008–1015. <https://doi.org/10.1214/aoms/1177704468>.
- Michel, R. (2006). *Simulation and estimation in multivariate generalized Pareto models*. (Doctoral dissertation), Universität Würzburg..
- Murphy, C., Tawn, J. A., & Varty, Z. (2025). Automated threshold selection and associated inference uncertainty for univariate extremes. *Technometrics*, 67(2), 215–224. <https://doi.org/10.1080/00401706.2024.2421744>.
- Pickands I.I., J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1), 119–131. <https://doi.org/10.1214/aos/1176343003>.
- Rootz'en, H., Segers, J., & Wadsworth, J. L. (2018). Multivariate generalized Pareto distributions: parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165, 117–131. <https://doi.org/10.1016/j.jmva.2017.12.003>.
- Rootz'en, H., & Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12(5), 917–930. <https://doi.org/10.3150/bj/1161614952>.
- Shi, J. (2006). *Stream of variation modeling and analysis for multistage manufacturing processes*. Boca Raton, FL: CRC Press, <https://doi.org/10.1201/9781420003901>.
- Shi, D., Axinte, D. A., & Gindy, N. N. (2007). Development of an online machining process monitoring system: a case study of the broaching process. *International Journal of Advanced Manufacturing Technology*, 34(1), 34–46. <https://doi.org/10.1007/s00170-006-0588-1>.
- Shi, J., & Zhou, S. (2009). Quality control and improvement for multistage systems: a survey. *IEEE Transactions*, 41(9), 744–753. <https://doi.org/10.1080/07408170902966344>.
- Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1), 67–90. <https://doi.org/10.1093/biomet/72.1.67>.
- Tang, B., Frye, H. A., Silander, J. A., & Jr. Gelfand, A. E. (2025). Zero-inflated multivariate tobit regression modeling. *Journal of Statistical Planning and Inference*, 236, Article 106229. <https://doi.org/10.1016/j.jspi.2024.106229>.
- Wang, X., & Li, X. (2021). Generalized confidence intervals for zero-inflated Pareto distribution. *Mathematics*, 9(24), 3272. <https://doi.org/10.3390/math9243272>.
- Xiang, L., & Tsung, F. (2008). Statistical monitoring of multi-stage processes based on engineering models. *IEEE Transactions*, 40(10), 957–970. <https://doi.org/10.1080/07408170701880845>.
- Yuan, T., Ramadan, S. Z., & Bae, S. J. (2011). Yield prediction for integrated circuits manufacturing through hierarchical Bayesian modeling of spatial defects. *IEEE Transactions on Reliability*, 60(4), 729–741. <https://doi.org/10.1109/TR.2011.2161698>.
- Zantek, P. F., Wright, G. P., & Plante, R. D. (2002). Process and product improvement in manufacturing systems with correlated stages. *Management Science*, 48(5), 591–606. <https://doi.org/10.1287/mnsc.48.5.591.7804>.