Pak.j.stat.oper.res. Vol.22 No.1 2026 pp 45-24 DOI: http://dx.doi.org/10.18187/pjsor.v22i1.5093

Pakistan Journal of Statistics and Operation Research

Parameter Estimation for the Bivariate Compound Zero-Truncated Poisson-Gamma Model under Different Data Scenarios

Amal D. Alhejaili[1*], Ateq Ahmed AlGhamedi[2]

*Corresponding author

1. Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia
aalhejaili0005@stu.kau.edu.sa

2. Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia
ateq@kau.edu.sa

Abstract

The bivariate compound zero-truncated Poisson-gamma distribution represents the sum of a random number of bivariate Gamma variables, with the count governed by a zero-truncated Poisson distribution. This formulation makes the model particularly suitable for applications in actuarial science, climatology, and reliability engineering, where zero outcomes are structurally absent. However, owing to the intractable form of the probability density function, which involves an infinite series, direct maximum likelihood estimation becomes computationally demanding. In this study, we use standard (exact) maximum likelihood estimation when event counts are observed (complete data and Scenario A) and employ the saddle-point approximation only when counts are latent (Scenario B). We developed a stable maximum likelihood estimation based on the saddle-point approximation. We derived the cumulative distribution function from the cumulant generating function and obtained the probability density function using numerical differentiation. Detailed derivations, implementation guidelines in the R programming language, and a parameter initialization strategy using the method of moments estimation are provided. A simulation study using various sample sizes demonstrated the accuracy, consistency, and superiority of this method over the moment-based estimators. Computational challenges and limitations are discussed, along with potential extensions to model the dependence structures using copulas. In addition, we develop a likelihood ratio test and a formal symmetry test (for example, $H_0 : \alpha_1 = \alpha_2,\ \beta_1 = \beta_2$) to compare nested specifications, enabling principled inference on symmetry and overall model adequacy.

KeyWords: Bivariate Compound Distribution; Zero-Truncated Poisson-Gamma Distribution; Method of Moments Estimation; Maximum Likelihood Estimation; Saddle-Point Approximation; Symmetry Test; Likelihood Ratio Test.

Mathematical Subject Classification: 62F10, 60E10, 62P05.

1. Introduction

Compound distributions, which model the sum of a random number of independent and identically distributed (i.i.d.) random variables, are vital in stochastic modeling. In the bivariate setting, where each summand is a pair (X_{ji}, Y_{ji}), the aggregate vector for observation j is defined as $\mathbf{S}_j = (S_{1j}, S_{2j}) = \left(\sum_{i=1}^{N_j} X_{ji}, \sum_{i=1}^{N_j} Y_{ji}\right)$. Where N_j denotes the number of events for observation j, and X_{ji} and Y_{ji} represent independent gamma-distributed random variables associated with each event $i = 1, \ldots, N_j$.

The bivariate compound zero-truncated Poisson–gamma (BCZTPG) distribution, introduced by Alhejaili and Al-Ghamedi (2024a), specializes in structure by assuming that the event count N_j follows a zero-truncated Poisson (ZTP) distribution, $N_j \sim \text{ZTP}(\lambda)$, whereas $X_{ji} \sim \text{Gamma}(\alpha_1, \beta_1)$ and $Y_{ji} \sim \text{Gamma}(\alpha_2, \beta_2)$ are independent for each i and j. This formulation is particularly relevant in scenarios where zero observations are structurally excluded.

The BCZTPG model can be applied in various domains. Actuarial science supports the modeling of joint claim totals from two insurance lines, conditional on the occurrence of at least one claim, as follows: Such compound models have been widely applied in risk theory and quantitative finance McNeil et al. (2015). In climatology, it represents the total rainfall at two nearby stations on days with guaranteed precipitation. In reliability engineering, it offers a framework for assessing the cumulative stress or degradation from multiple sources, given that some impacts have been observed. These examples demonstrate the practical relevance of a model that handles non-zero bivariate event data. For further theoretical development and general application of saddle-point approximations (SPAs), see Butler (2007) and Huzurbazar (1956).

Despite its usefulness, BCZTPG distribution presents considerable challenges for parameter estimation because of the infinite sum involved in its probability density function (PDF). This complexity makes direct maximum likelihood estimation (MLE) computationally expensive and often impractical, as it typically requires truncation or numerical approximation, which may sacrifice accuracy. To overcome these limitations, this study proposes a novel estimation approach based on SPA. Originating from Daniels (1954) and further refined by Lugannani and Rice (1980), SPA offers highly accurate approximations of probability distributions through their cumulant generating functions (CGFs), particularly in the tails. It is significantly more efficient than Monte Carlo simulations or numerical integration in dealing with complex compound distributions. We developed an SPA for the cumulative distribution function (CDF) of the BCZTPG model and obtained the PDF via numerical differentiation. This enables stable and accurate computation of the likelihood function, facilitating the MLE of the parameter vector $\theta = (\lambda, \alpha_1, \alpha_2, \beta_1, \beta_2)^\top$. Our work includes full mathematical derivations of the CGF, saddle-point equations, and associated approximations, as well as a detailed discussion of numerical implementation strategies, including parameter initialization using the method of moments (MoMs).

The effectiveness of the proposed approach was validated through a comprehensive simulation study, which demonstrated its accuracy, consistency, and computational feasibility compared with moment-based methods. Finally, we address the potential limitations of the method, such as the assumption of independence between components, and discuss possible extensions, including the incorporation of dependence structures through copulas. In this study, we introduced the BCZTPG distribution, a compound model that combines a ZTP distribution with gamma-distributed claim amounts. We propose estimation procedures for the parameters of the BCZTPG distribution using both the MoM and MLE. Two estimation scenarios were considered. In the first scenario, we assumed complete data, where the full structure $\left(N_j, (X_{ji}, Y_{ji})_{i=1}^{N_j}\right)_{j=1}^{m}$ was observed. The second scenario addresses partially observed data and consists of two subcases: one in which $(N_j, S_{1j}, S_{2j})_{j=1}^{m}$ are available, and the other in which only $(S_{1j}, S_{2j})_{j=1}^{m}$ are observed, with N_j unobserved.

Throughout this study, we adopted the following notation: A BCZTPG model, denoted as $\text{ZTP}(\lambda) \bigvee \{\text{Gamma}(\alpha_1, \beta_1), \text{Gamma}(\alpha_2, \beta_2)\}$, has $\lambda > 0$ as the Poisson rate, $\alpha_1, \alpha_2 > 0$ as the gamma shape parameters, and $\beta_1, \beta_2 > 0$ as the gamma scale parameters. The BCZTPG variable represents the sum of a random number of gamma-distributed terms, where the number of terms follows a ZTP distribution.

The remainder of this paper is organized as follows. Section 1 introduces the problem and its contributions. Section 2 reviews related literature, and Section 3 describes the methodology. Sections 4 and 5 present the parameter estimation methods. Section 4 (complete data) describes the MoM and MLE. Section 5 (incomplete data) covers Scenarios A (5.1; observed N_j) and Scenario B (5.2; latent N_j), with scenario-specific derivations. Section 6 develops the likelihood-based tests (likelihood ratio test (LRT) and symmetry). Section 7 presents the simulation results. Section 8 presents real data, and Section 9 concludes the paper.

2. Literature Review

Compound distributions, in which a random number of random variables is summed, have been extensively studied in statistics and applied mathematics. In actuarial science, the compound Poisson distribution is a cornerstone for modeling aggregate claims, capturing the total claim amount over a random number of insurance events (Klugman et al., 2012). The Gamma distribution is frequently employed to model claim severity owing to its flexibility, positive support, and mathematical tractability in many scenarios (Bowers, 1997). Extensions to bivariate settings, such as bivariate Poisson–gamma models, offer the ability to represent joint claim events; however, these models often assume independence between components (Kocherlakota and Kocherlakota, 1992).

ZTP distributions become relevant when modeling scenarios in which zero observations are excluded, such as insurance claims conditioned on at least one recorded incident (Cameron and Trivedi, 2013). These models are also crucial in econometric applications, where truncation arises from the data collection process (Grogger and Carson, 1991). Recently, Nascimento et al. (2023) explored a univariate compound version, the compound truncated Poisson–gamma (CTPG) distribution, for climatological applications. The BCZTPG distribution introduced by Alhejaili and AlGhamedi (2024a) unifies these modeling features by incorporating bivariate gamma marginals, compound summation, and ZTP counts. Although their work laid the foundation for deriving the CGF and SPA for the CDF, it did not provide a complete estimation strategy for the model parameters.

To address the challenges posed by the analytically intractable PDF of the BCZTPG distribution, this study employs the SPA [for a recent comprehensive review, see Alhejaili and AlGhamedi (2025)]. Originating from the seminal work of Daniels (1954), saddle-point methods provide accurate approximations of distributions that lack closed-form densities. Lugannani and Rice (1980) extended the approach to the sums of independent random variables, while multivariate extensions were explored by Wang (1990) and Kolassa (2003a). Related advances tailored to compound settings include conditional SPAs for bivariate compound distributions (Alhejaili and AlGhamedi, 2024b) and stopped-sum SPA results (Alhejaili and Abd-Elfattah, 2013), both of which are directly relevant to our mixture-over-count construction. Kolassa (2003b) provided a detailed account of bivariate SPA, which directly supports the BCZTPG framework. Reid (1988) emphasized the power of these techniques in statistical inference, particularly for MLE in complex models. Asmussen (2000) specifically demonstrates the value of SPA in insurance risk models involving compound sums. Their application in risk theory (Jensen, 1991), climatology (Nascimento et al., 2023), and other fields underscores their practical relevance. Willmot and Lin (2007) examined compound distributions with exponential claims, providing insights into the flexibility of compound Poisson–gamma frameworks.

Compared with traditional alternatives, the SPA offers a compelling balance between computational efficiency and statistical accuracy. Direct series summation suffers from slow convergence and truncation bias, particularly for large values of λ. While flexible, Monte Carlo simulations can be computationally expensive when used for repeated likelihood evaluations. Although simpler, the MoM generally yields less efficient estimators than MLE and is often used solely for initialization. Numerical integration techniques, including the Fast Fourier Transform (FFT), can provide efficient approximations but are prone to discretization and stability issues. This study builds on the contributions of Alhejaili and AlGhamedi (2024a) by developing a practical SPA-based framework for MLE in the BCZTPG model. See Alhejaili and AlGhamedi (2024b) for the conditional SPA for the bivariate compound case. By leveraging the SPA to derive a numerically stable PDF through the differentiation of the CDF, we enable accurate and tractable parameter estimation. This approach bridges a notable gap in the literature and opens the door for broader applications of the BCZTPG distribution in the statistical modeling of nonzero bivariate random sums.

3. Methodology

The BCZTPG distribution is constructed as follows: For each observation j, let N_j denote the number of events, and let X_{ji} and Y_{ji} be independent random variables for each event $i = 1,\ldots,N_j$, with $X_{ji} \sim \text{Gamma}(\alpha_1, \beta_1)$ and $Y_{ji} \sim \text{Gamma}(\alpha_2, \beta_2)$. The bivariate compound sum for observation j is defined as

$$\mathbf{S}_j = \begin{pmatrix} \sum_{i=1}^{N_j} X_{ji} \\ \sum_{i=1}^{N_j} Y_{ji} \end{pmatrix} = \begin{pmatrix} S_{1j} \\ S_{2j} \end{pmatrix}, \tag{1}$$

Then, $\mathbf{S}_j$ is said to follow the BCZTPG distribution for observation j, where N_j is independent of (X_{ji}, Y_{ji}).

We further assume that the number of events N_j follows a ZTP distribution with parameter λ, denoted by ZTP(λ), which has the probability mass function (PMF)

$$P(N_j) = \frac{\lambda^{N_j} e^{-\lambda}}{N_j!(1-e^{-\lambda})}, \quad N_j = 1,2,\ldots, \ \lambda > 0, \tag{2}$$

This ensures that the summation involves at least one observation. This condition is particularly relevant in fields such as insurance and climatology, where the process of interest is conditional on an event. Given N_j, the aggregated variables $S_{1j} = \sum_{i=1}^{N_j} X_{ji}$ and $S_{2j} = \sum_{i=1}^{N_j} Y_{ji}$ follow gamma distributions with shape parameters $N_j\alpha_1$ and $N_j\alpha_2$, and scale parameters β_1 and β_2, respectively. The conditional joint density function of (S_{1j}, S_{2j}) given N_j is then given by the product of two independent gamma densities:

$$f_{S|N_j}(S_{1j}, S_{2j}) = \frac{S_{1j}^{N_j\alpha_1-1} e^{-S_{1j}/\beta_1}}{\Gamma(N_j\alpha_1)\beta_1^{N_j\alpha_1}} \cdot \frac{S_{2j}^{N_j\alpha_2-1} e^{-S_{2j}/\beta_2}}{\Gamma(N_j\alpha_2)\beta_2^{N_j\alpha_2}}, \quad S_{1j} > 0, \ S_{2j} > 0.$$

We assume the conditional independence of (S_{1j}, S_{2j}) given N_j, which induces dependence via the shared N_j but does not capture the residual cross-component correlation. Applications with stronger cross-component dependence may require extensions (e.g., copula-based conditioning), which will be addressed in future work. The estimation of the parameter vector $\theta = (\lambda, \alpha_1, \alpha_2, \beta_1, \beta_2)^\top$ is crucial for enabling the practical application of the BCZTPG distribution. This model arises from the sum of a random number $N_j \sim$ ZTP(λ) of i.i.d. bivariate gamma vectors (X_{ji}, Y_{ji}), where $X_{ji} \sim$ Gamma(α_1, β_1) and $Y_{ji} \sim$ Gamma(α_2, β_2). It is assumed that independence between X_{ji} and Y_{ji} leads to a tractable bivariate structure while preserving the flexibility in the marginal distributions.

The unconditional joint density of (S_{1j}, S_{2j}) is obtained by summing all possible values of N_j, resulting in an infinite series representation as follows:

$$f_{S_{1j},S_{2j}}(s_{1j}, s_{2j}; \theta) = \sum_{N_j=1}^{\infty} \left[\frac{\lambda^{N_j} e^{-\lambda}}{N_j!(1-e^{-\lambda})} \cdot \frac{s_{1j}^{N_j\alpha_1-1} e^{-s_{1j}/\beta_1}}{\beta_1^{N_j\alpha_1}\Gamma(N_j\alpha_1)} \cdot \frac{s_{2j}^{N_j\alpha_2-1} e^{-s_{2j}/\beta_2}}{\beta_2^{N_j\alpha_2}\Gamma(N_j\alpha_2)} \right].$$

Although this formulation is theoretically sound, the direct use of the likelihood function based on the density above presents significant computational challenges owing to the infinite summation and presence of gamma functions with shape parameters dependent on N_j.

We consider two estimation scenarios. In the first scenario, we assume complete data, where the full structure $\left(N_j, (X_{ji}, Y_{ji})_{i=1}^{N_j}\right)_{j=1}^{m}$ is observed. The second scenario involves partially observed data with two subcases: one where $(N_j, S_{1j}, S_{2j})_{j=1}^{m}$ are known, and another where only $(S_{1j}, S_{2j})_{j=1}^{m}$ are observed while N_j remains unobserved. To estimate the parameters under these settings, we employ the MoM and MLE, using SPA when N_j was not observed.

4. Parameter Estimation of the BCZTPG Distribution under Complete Data

In this section, we explore the complete data setting where $(N_j, (X_{ji}, Y_{ji})_{i=1}^{N_j})_{j=1}^{m}$ are observed. We estimate the parameters λ, α_1, α_2, β_1, and β_2 using the MoM and MLE. Since the counts N_j are observed, λ is directly informed by the data and can be estimated using the standard ZTP-MLE, providing both computational efficiency and accurate inference.

In contrast, when N_j is unobserved, the estimation relies solely on (S_{1j}, S_{2j}), which makes it significantly harder to separate λ, α_1, β_1, α_2, and β_2. In this scenario, constrained numerical optimization methods (e.g., L-BFGS-B) are typically used for the marginal likelihood, or one may resort to SPA as a stable surrogate; large samples and, when appropriate, Bayesian priors can further enhance identifiability. Thus, observing N_j is highly beneficial whenever

feasible. Below, we present a MoM procedure for estimating λ, α_1, α_2, β_1, and β_2 using complete data, followed by the associated MLE formulation. Our derivations build upon those of Nascimento et al. (2023), providing additional details for clarity.

4.1. Method of Moments Estimation

The MoM equates sample moments to their theoretical counterparts in order to solve for unknown parameters (Sorenson, 1980). With a sufficiently large m, the sample moments serve as consistent estimators of the population moments; in smaller samples, bias may occur, particularly when the sample means exceed the sample variances. Thus, basic diagnostics are advisable.

In the complete data case, we observe N_j and $(X_{ji}, Y_{ji})_{i=1}^{N_j}$ for $j = 1, \ldots, m$, where X_{ji} and Y_{ji} are independent gamma distributed random variables. The required theoretical moments are as follows: For the ZTP distribution,

$$\mathbb{E}[N_j] = \frac{\lambda}{1-e^{-\lambda}}, \qquad \mathrm{Var}(N_j) = \mathbb{E}[N_j]\big(1+\lambda-\mathbb{E}[N_j]\big). \tag{3}$$

For the Gamma distributions,

$$\mathbb{E}[X_{ji}] = \alpha_1\beta_1, \qquad \mathrm{Var}(X_{ji}) = \alpha_1\beta_1^2, \tag{4}$$

with similar expressions for Y_{ji}.

Define the empirical summaries

$$\bar{N} = \frac{1}{m}\sum_{j=1}^{m} N_j, \qquad \bar{X} = \frac{1}{\sum_{j=1}^{m} N_j}\sum_{j=1}^{m}\sum_{i=1}^{N_j} X_{ji}, \qquad s_X^2 = \frac{1}{\sum_{j=1}^{m} N_j - 1}\sum_{j=1}^{m}\sum_{i=1}^{N_j}(X_{ji}-\bar{X})^2,$$

with equivalent definitions for Y_{ji}.

To estimate λ, we utilize

$$\bar{N} = \frac{\hat{\lambda}}{1-e^{-\hat{\lambda}}}, \tag{5}$$

which is solved using numerical methods.

For the parameters α_1 and β_1, moment matching provides

$$\hat{\beta}_1 = \frac{s_X^2}{\bar{X}}, \qquad \hat{\alpha}_1 = \frac{\bar{X}^2}{s_X^2},$$

and the same formulas are applied to Y_{ji} to derive $(\hat{\alpha}_2, \hat{\beta}_2)$. In practice, ensure that $\bar{X}, \bar{Y}, s_X^2, s_Y^2 > 0$ and, when outliers are present, consider using robust dispersion measures (e.g., trimmed variances).

4.2. Maximum Likelihood Estimation

MLE selects parameter values that maximize the joint likelihood of the observed sample (Rossi, 2018). Numerical optimization is necessary for this model, and sensible initialization helps facilitate convergence. In this section, we concentrate on the MLE of the unknown parameters from a random sample of the BCZTPG distribution, specifically estimating the parameters λ, $\alpha_1, \alpha_2, \beta_1$, and β_2 by maximizing the joint log-likelihood. The likelihood function for this model can be expressed in terms of λ, $\alpha_1, \alpha_2, \beta_1$, and it can be maximized using the L-BFGS-B algorithm.

The MLEs of the parameters $\lambda, \alpha_1, \alpha_2, \beta_1, \beta_2$ are obtained in a similar from random samples $(N_j, (X_{ji}, Y_{ji})_{i=1}^{N_j})_{j=1}^{m}$, where $N_j \sim \mathrm{ZTP}(\lambda)$, $X_{ji} \sim \mathrm{Gamma}(\alpha_1, \beta_1)$, and $Y_{ji} \sim \mathrm{Gamma}(\alpha_2, \beta_2)$, all of which are mutually independent. The

likelihood function for the complete data is expressed as follows:

$$L(\theta)=\prod_{j=1}^{m}\left[\frac{\lambda^{N_j}e^{-\lambda}}{N_j!(1-e^{-\lambda})}\prod_{i=1}^{N_j}\frac{X_{ji}^{\alpha_1-1}e^{-X_{ji}/\beta_1}}{\beta_1^{\alpha_1}\Gamma(\alpha_1)}\frac{Y_{ji}^{\alpha_2-1}e^{-Y_{ji}/\beta_2}}{\beta_2^{\alpha_2}\Gamma(\alpha_2)}\right],$$

The log-likelihood is given by:

$$\begin{aligned}\log L(\theta)=\sum_{j=1}^{m}\Bigg[&\Big(N_j\log\lambda-\lambda-\log(N_j!)-\log(1-e^{-\lambda})\Big)\\&+\sum_{i=1}^{N_j}\left((\alpha_1-1)\log X_{ji}-\frac{X_{ji}}{\beta_1}-\alpha_1\log\beta_1-\log\Gamma(\alpha_1)\right)\\&+\sum_{i=1}^{N_j}\left((\alpha_2-1)\log Y_{ji}-\frac{Y_{ji}}{\beta_2}-\alpha_2\log\beta_2-\log\Gamma(\alpha_2)\right)\Bigg].\end{aligned}$$

Setting the score functions to zero results in the estimating equations. For (λ):

$$\frac{\partial\log L}{\partial\lambda}=\sum_{j=1}^{m}\frac{N_j}{\lambda}-m-\frac{me^{-\lambda}}{1-e^{-\lambda}}. \tag{6}$$

For (α_1,β_1):

$$\frac{\partial\log L}{\partial\alpha_1}=\sum_{j=1}^{m}\sum_{i=1}^{N_j}\log X_{ji}-\sum_{j=1}^{m}N_j\log\beta_1-\sum_{j=1}^{m}N_j\psi(\alpha_1)=0,$$

and

$$\frac{\partial\log L}{\partial\beta_1}=-\frac{\alpha_1}{\beta_1}\sum_{j=1}^{m}N_j+\frac{1}{\beta_1^2}\sum_{j=1}^{m}\sum_{i=1}^{N_j}X_{ji}=0,$$

which leads to:

$$\hat{\beta}_1=\frac{\sum_{j=1}^{m}\sum_{i=1}^{N_j}X_{ji}}{\hat{\alpha}_1\sum_{j=1}^{m}N_j}.$$

Similar expressions apply to (α_2,β_2):

$$\sum_{j=1}^{m}\sum_{i=1}^{N_j}\log Y_{ji}-\sum_{j=1}^{m}N_j\log\hat{\beta}_2-\sum_{j=1}^{m}N_j\psi(\hat{\alpha}_2)=0,$$

$$\hat{\beta}_2=\frac{\sum_{j=1}^{m}\sum_{i=1}^{N_j}Y_{ji}}{\hat{\alpha}_2\sum_{j=1}^{m}N_j}.$$

Solve for $(\hat{\alpha}_1,\hat{\alpha}_2)$ iteratively, updating $(\hat{\beta}_1,\hat{\beta}_2)$ in closed form at each step. Standard errors are derived from the observed Fisher information evaluated at the MLE.

5. Parameter Estimation of the BCZTPG Distribution under Incomplete Data

The second scenario deals with partially observed data. In scenario A we observe (N_j,S_{1j},S_{2j}), while in Scenario B, we observe only (S_{1j},S_{2j}). When N_j is available (Scenario A), conditioning on N_j eliminates ZTP marginalization and greatly simplifies the computations. In the absence of N_j (Scenario B), the marginal density involves an infinite ZTP mixture and necessitates numerical approximations (e.g., SPA) or alternative strategies.

5.1. Scenario A

In Scenario A, the data consist of triplets $(N_j, S_{1j}, S_{2j})_{j=1}^{m}$, where, for each observation j, both the aggregate sums S_{1j}, S_{2j}, and the corresponding count N_j are fully observed. This setting simplifies parameter estimation because, conditional on N_j, variables S_{1j} and S_{2j} follow independent gamma distributions with shape parameters $N_j\alpha_1$ and $N_j\alpha_2$, and scale parameters β_1 and β_2, respectively. Thus, the likelihood function can be constructed as a product of the ZTP probabilities for N_j and the Gamma densities for S_{1j} and S_{2j}. This structure enables the application of standard estimation techniques, such as the MoM or MLE, using the complete data available for each observation. Having N_j available for each record eliminates the need for marginalization and reduces the computational complexity involved in parameter estimation.

5.1.1. Method of Moments Estimation

In the following, we derive MoM estimators for the BCZTPG model parameters λ, α_1, α_2, β_1, and β_2, refining the construction from (Nascimento et al., 2023) and providing additional clarifications for implementation. Let $N_j \sim \text{ZTP}(\lambda)$ as in (3) and recall the conditional sums

$$S_{1j} \mid N_j \sim \text{Gamma}(N_j\alpha_1, \beta_1), \qquad S_{2j} \mid N_j \sim \text{Gamma}(N_j\alpha_2, \beta_2),$$

with

$$\mathbb{E}[S_{1j} \mid N_j] = N_j\alpha_1\beta_1, \qquad \text{Var}(S_{1j} \mid N_j) = N_j\alpha_1\beta_1^2, \tag{7}$$

and with similar expressions for S_{2j}. Unconditionally, we have:

$$\mathbb{E}[S_{1j}] = \frac{\lambda\alpha_1\beta_1}{1-e^{-\lambda}}, \tag{8}$$

$$\text{Var}(S_{1j}) = \frac{\lambda\alpha_1\beta_1^2}{1-e^{-\lambda}}\left[1+\alpha_1\left(1-\frac{\lambda}{1-e^{-\lambda}}+\lambda\right)\right]. \tag{9}$$

Define:

$$\bar{N} = \frac{1}{m}\sum_{j=1}^{m} N_j, \qquad s_N^2 = \frac{1}{m-1}\sum_{j=1}^{m}(N_j-\bar{N})^2,$$

and

$$\bar{S}_1 = \frac{1}{m}\sum_{j=1}^{m} S_{1j}, \qquad s_{S_1}^2 = \frac{1}{m-1}\sum_{j=1}^{m}(S_{1j}-\bar{S}_1)^2. \tag{10}$$

Solving (5) yields $\hat{\lambda}$. Substituting (8)–(5) directly yields:

$$\hat{\beta}_1 = \frac{\bar{N}s_{S_1}^2 - \bar{S}_1^2(1+\hat{\lambda}-\bar{N})}{\bar{N}\bar{S}_1}, \tag{11}$$

$$\hat{\alpha}_1 = \frac{\bar{S}_1^2}{\bar{N}s_{S_1}^2 - \bar{S}_1^2(1+\hat{\lambda}-\bar{N})}. \tag{12}$$

In a similar manner, we have:

$$\hat{\beta}_2 = \frac{\bar{N}s_{S_2}^2 - \bar{S}_2^2(1+\hat{\lambda}-\bar{N})}{\bar{N}\bar{S}_2}, \qquad \hat{\alpha}_2 = \frac{\bar{S}_2^2}{\bar{N}s_{S_2}^2 - \bar{S}_2^2(1+\hat{\lambda}-\bar{N})}.$$

Equations (5), (8)–(9), (11), and (12) yield the MoM estimators for $(\lambda, \alpha_1, \beta_1, \alpha_2, \beta_2)$ under Scenario A; the positivity conditions are enforced as noted. Next, we present the corresponding MLE formulation for Scenario A in Section 5.1.2.

5.1.2. Maximum Likelihood Estimation

To avoid repetition, we will refer to the description of Scenario A above and highlight only the likelihood components and optimization details necessary for MLE implementation. The objective is to derive the likelihood function for this distribution in terms of $\lambda, \alpha_1, \alpha_2, \beta_1, \beta_2$. This analysis assumes that N_j is observed for each observation in the dataset. To tackle these computational challenges, the L-BFGS-B algorithm is directly applied to the scenario where N_j is observed (Cameron and Trivedi, 2013).

When N_j is observed, the marginal PDF of S_{1j} and S_{2j} does not require summation over the distribution of N_j. Instead, the conditional distribution of S_{1j} and S_{2j} given N_j is utilized, and the likelihood involves the PMF of the ZTP distribution for the observed N_j values. The likelihood of observation j is given by:

$$L_j(\theta; \text{data}_j) = P(N_j; \lambda) \cdot f_{S_{1j},S_{2j}|N_j}(S_{1j}, S_{2j}; \alpha_1, \beta_1, \alpha_2, \beta_2),$$

where $\theta = (\lambda, \alpha_1, \alpha_2, \beta_1, \beta_2)^\top$. The PMF of the ZTP distribution is given by:

$$P(N_j; \lambda) = \frac{\lambda^{N_j} e^{-\lambda}}{N_j!(1-e^{-\lambda})}, \qquad N_j = 1,2,\ldots, \quad \lambda > 0.$$

Given N_j, the conditional distributions for the sums are:

$$S_{1j} \mid N_j \sim \text{Gamma}(N_j\alpha_1, \beta_1), \qquad S_{2j} \mid N_j \sim \text{Gamma}(N_j\alpha_2, \beta_2),$$

with the joint density given by:

$$f_{S_{1j},S_{2j}|N_j}(S_{1j}, S_{2j}) = \frac{S_{1j}^{N_j\alpha_1-1} e^{-S_{1j}/\beta_1}}{\Gamma(N_j\alpha_1)\beta_1^{N_j\alpha_1}} \frac{S_{2j}^{N_j\alpha_2-1} e^{-S_{2j}/\beta_2}}{\Gamma(N_j\alpha_2)\beta_2^{N_j\alpha_2}},$$

where $S_{1j} > 0, S_{2j} > 0$.

The likelihood for all m observations is given by:

$$L(\theta) = \prod_{j=1}^{m} \left[\frac{\lambda^{N_j} e^{-\lambda}}{N_j!(1-e^{-\lambda})} \frac{S_{1j}^{N_j\alpha_1-1} e^{-S_{1j}/\beta_1}}{\Gamma(N_j\alpha_1)\beta_1^{N_j\alpha_1}} \frac{S_{2j}^{N_j\alpha_2-1} e^{-S_{2j}/\beta_2}}{\Gamma(N_j\alpha_2)\beta_2^{N_j\alpha_2}} \right].$$

The log-likelihood function is expressed as follows:

$$\log L(\theta) = \sum_{j=1}^{m} \Big(N_j \log\lambda - \lambda - \log N_j! - \log(1-e^{-\lambda}) + (N_j\alpha_1 - 1)\log S_{1j} - \frac{S_{1j}}{\beta_1} - N_j\alpha_1 \log\beta_1 - \log\Gamma(N_j\alpha_1) + (N_j\alpha_2 - 1)\log S_{2j} - \frac{S_{2j}}{\beta_2} - N_j\alpha_2 \log\beta_2 - \log\Gamma(N_j\alpha_2) \Big)$$

The summary statistics can be defined as

$$N_\cdot = \sum_{j=1}^{m} N_j, \quad S_{1\cdot} = \sum_{j=1}^{m} S_{1j}, \quad S_{2\cdot} = \sum_{j=1}^{m} S_{2j}, \quad T_{1j} = \log S_{1j}, \quad T_{2j} = \log S_{2j}$$

The scoring equations are provided below:

$$\frac{\partial \log L}{\partial \alpha_1} = \sum_{j=1}^{m} \left[N_j \log S_{1j} - N_j \log\beta_1 - N_j \psi(N_j\alpha_1) \right],$$

$$\frac{\partial \log L}{\partial \beta_1} = -\frac{N.\alpha_1}{\beta_1} + \frac{S_{1\cdot}}{\beta_1^2},$$

$$\frac{\partial \log L}{\partial \alpha_2} = \sum_{j=1}^{m} \left[N_j \log S_{2j} - N_j \log \beta_2 - N_j \psi(N_j \alpha_2) \right],$$

$$\frac{\partial \log L}{\partial \beta_2} = -\frac{N.\alpha_2}{\beta_2} + \frac{S_{2\cdot}}{\beta_2^2}.$$

The closed-form solutions for β_1 and β_2 are given by:

$$\hat{\beta}_1 = \frac{S_{1\cdot}}{N.\hat{\alpha}_1}, \qquad \hat{\beta}_2 = \frac{S_{2\cdot}}{N.\hat{\alpha}_2}.$$

These values are substituted into the log-likelihood, and numerical optimization is performed with respect to λ, α_1, α_2, using the observed values N_j, S_{1j}, S_{2j}. The estimate of λ is given by (5).

For gradient-based optimization, the gradients can be approximated using finite differences as follows:

$$\frac{\partial \log L}{\partial \theta_i} \approx \frac{\log L(\theta + \delta e_i) - \log L(\theta - \delta e_i)}{2\delta}, \qquad \delta = 10^{-6}$$

The initial values are selected as follows: for λ, use $\hat{\lambda}_0 = \bar{N}(1 - e^{-\bar{N}})^{-1}$, and for the gamma parameters, use the MoM based on S_{kj}/N_j:

$$\bar{S}_k = \frac{1}{m} \sum_{j=1}^{m} \frac{S_{kj}}{N_j}, \quad \hat{\beta}_{k0} = \frac{\mathrm{Var}(S_{kj}/N_j)}{\bar{S}_k}, \quad \hat{\alpha}_{k0} = \frac{\bar{S}_k}{\hat{\beta}_{k0}}.$$

Standard errors can be computed by numerically evaluating the observed Fisher information as follows:

$$I(\hat{\theta}) = -\left. \frac{\partial^2 \log L}{\partial \theta \, \partial \theta^T} \right|_{\hat{\theta}}, \quad \mathrm{SE}(\hat{\theta}_i) = \sqrt{\left[I^{-1}(\hat{\theta}) \right]_{ii}}$$

It is crucial to ensure numerical stability by using robust implementations of $\log\Gamma(N_j\alpha_k)$ and $\psi(N_j\alpha_k)$, enforcing positivity constraints, and guarding against overflow for large N_j. If desired, the conditional density $f_{S|N}$ can also be approximated using the SPA, which may be beneficial for validation or in situations where evaluating $\Gamma(N_j\alpha_k)$ is unstable. In the current scenario, the exact gamma density is utilized because N_j is known; therefore, the SPA is not necessary.

5.2. Scenario B

In Scenario B, the observed data consist solely of the aggregated pairs $(S_{1j}, S_{2j})_{j=1}^{m}$, while the latent count variable N_j remains unobserved. Each pair can be interpreted as the sum of an unknown number of gamma-distributed contributions. Consequently, the observed totals follow a mixture distribution over all possible event counts, with mixture weights determined by the probability mass function of the ZTP model.

This missing-count setting complicates inference significantly: marginally, (S_{1j}, S_{2j}) is distributed as a mixture over $N_j \in \{1, 2, \dots\}$. Thus, the marginal likelihood for each observation takes the form of an infinite series, where each term combines the probability of N_j under the ZTP model with the joint gamma density of (S_{1j}, S_{2j}) conditional on N_j. Evaluating this series directly is computationally expensive, so practical estimation typically relies on numerical approximations, such as the SPA, or alternative strategies that avoid explicit summation over all possible N_j.

The SPA tackles this difficulty by leveraging the CGF to create a highly accurate local approximation around the saddle-point. This transformation converts an otherwise intractable problem into a manageable numerical optimization task while maintaining accuracy and stability. For further implementation details, including the derivation of the saddle-point equations, numerical solution strategies, and simulation-based illustrations, the reader is referred to (Al-

hejaili and AlGhamedi, 2024a), which complements the current work. Thus, Scenario B necessitates more advanced statistical techniques for reliable parameter estimation compared to the complete data case in Scenario A.

From an identifiability standpoint, Scenario B presents additional challenges because the latent event counts are unobserved, and the model parameters must be inferred solely from aggregated sums. When sample sizes are small or parameter values are large, multiple parameter configurations may yield similar likelihood surfaces, potentially resulting in weak identification and numerical instability during optimization. To address these issues, the estimation procedure utilizes moment-based initialization, positivity constraints, and robust numerical optimization, which collectively enhance convergence behavior and minimize the risk of spurious solutions. Additionally, simulation results indicate that identifiability improves as the sample size increases, aligning with standard asymptotic theory.

5.2.1. Method of Moments Estimation

In Scenario B, we observe aggregated sums S_{1j} and S_{2j}, but the number of events N_j is unobserved. The goal is to estimate the unknown parameters using only the observed data. To streamline the presentation, we adopt the same moment-based principle introduced in Scenario A and directly present the moment equations implied by the latent-count setting. We start by examining the theoretical moments of the distribution, which characterize the expected behavior of the random variables.

For the ZTP distribution, the expected value and variance of N_j are provided by (3). Given the data aggregation, we concentrate on conditional moments. Specifically, for S_{1j}, which is the sum of N_j i.i.d. Gamma random variables $X_{ji} \sim \text{Gamma}(\alpha_1, \beta_1)$, the expected value and variance are as indicated in (7); similar expressions apply to S_{2j}. We then calculate the unconditional moments by examining the expectations of the conditional moments. For S_{1j}, the mean and variance are specified by (8) and (9), respectively, and similar expressions hold for S_{2j}.

The covariance between S_{1j} and S_{2j} can be expressed as:

$$\text{Cov}(S_{1j}, S_{2j}) = \text{Var}(N_j)\,\alpha_1\beta_1\alpha_2\beta_2 = \frac{\lambda}{1-e^{-\lambda}}\left[1 - \frac{\lambda}{1-e^{-\lambda}} + \lambda\right]\alpha_1\beta_1\alpha_2\beta_2.$$

Turning to the sample moments, for S_{1j} and S_{2j}, the sample means and variances are calculated using Equation (10). The sample covariance between S_{1j} and S_{2j} is defined as:

$$\hat{C}_{12} = \frac{1}{m-1}\sum_{j=1}^{m}(S_{1j} - \bar{S}_1)(S_{2j} - \bar{S}_2).$$

We will now proceed with the MoM equations. First, for λ, we consider the following equations:

$$\mathbb{E}[N_j] = \frac{\lambda}{1-e^{-\lambda}}, \qquad \text{Var}(N_j) = \mathbb{E}[N_j]\,(1 - \mathbb{E}[N_j] + \lambda)$$

For the parameters α_1, β_1, α_2, and β_2, we employ the following equations:

$$\bar{S}_1 = \mathbb{E}[N_j]\alpha_1\beta_1, \qquad \bar{S}_2 = \mathbb{E}[N_j]\alpha_2\beta_2,$$

$$s^2_{S_1} = \mathbb{E}[N_j]\alpha_1\beta_1^2 + \text{Var}(N_j)(\alpha_1\beta_1)^2,$$

$$\hat{C}_{12} = \text{Var}(N_j)\alpha_1\beta_1\alpha_2\beta_2.$$

Similarly, the moment equations for S_{2j} apply to the corresponding parameters.

Solving this system results in:

$$\alpha_1\beta_1 = \frac{\bar{S}_1}{\mathbb{E}[N_j]}, \qquad \alpha_2\beta_2 = \frac{\bar{S}_2}{\mathbb{E}[N_j]},$$

Additionally, from the covariance expression, we have:

$$\mathrm{Var}(N_j) = \frac{\hat{C}_{12}}{\alpha_1\beta_1\alpha_2\beta_2} = \frac{\hat{C}_{12}\,\mathbb{E}[N_j]^2}{\bar{S}_1\bar{S}_2}.$$

Substituting into the variance equation for S_{1j} results in:

$$s^2_{S_1} = \mathbb{E}[N_j]\alpha_1\beta_1^2 + \mathrm{Var}(N_j)(\alpha_1\beta_1)^2 = \mathbb{E}[N_j]\alpha_1\beta_1^2 + \frac{\hat{C}_{12}\bar{S}_1}{\bar{S}_2\,\mathbb{E}[N_j]}.$$

Solving for $\alpha_1\beta_1^2$ gives:

$$\alpha_1\beta_1^2 = \frac{s^2_{S_1} - \frac{\hat{C}_{12}\bar{S}_1}{\bar{S}_2\,\mathbb{E}[N_j]}}{\mathbb{E}[N_j]}, \qquad \beta_1^2 = \frac{s^2_{S_1} - \frac{\hat{C}_{12}\bar{S}_1}{\bar{S}_2\,\mathbb{E}[N_j]}}{\bar{S}_1}.$$

Thus, we find that:

$$\hat{\beta}_1 = \sqrt{\frac{s^2_{S_1} - \frac{\hat{C}_{12}\bar{S}_1}{\bar{S}_2\,\mathbb{E}[N_j]}}{\bar{S}_1}}, \qquad \hat{\alpha}_1 = \frac{\bar{S}_1}{\mathbb{E}[N_j]\hat{\beta}_1}.$$

Similarly, the estimators for S_{2j} are derived by replacing the corresponding indices and parameters with $s^2_{S_2}$ instead of $s^2_{S_1}$.

Finally, we utilize the variance of N_j to estimate λ.

$$\mathrm{Var}(N_j) = \mathbb{E}[N_j]\,(1 - \mathbb{E}[N_j] + \lambda), \qquad \frac{\hat{C}_{12}\,\mathbb{E}[N_j]^2}{\bar{S}_1\bar{S}_2} = \mathbb{E}[N_j]\,(1 - \mathbb{E}[N_j] + \lambda).$$

By solving this equation for $\mathbb{E}[N_j] = \frac{\lambda}{1-e^{-\lambda}}$, we can derive the value of λ.

It is crucial to ensure that $s^2_{S_1} > \frac{\hat{C}_{12}\bar{S}_1}{\bar{S}_2\,\mathbb{E}[N_j]}$ (and a similar condition for S_{2j}) to prevent negative or imaginary parameter estimations. For robustness, it is recommended to employ numerical optimization techniques to solve for $\mathbb{E}[N_j]$ and λ, with constraints established to ensure positive parameter estimates.

5.2.2. Maximum Likelihood Estimation

Now, consider the incomplete data setting where only the aggregated pairs $\{(S_{1j}, S_{2j})\}_{j=1}^m$ are observed and the latent event counts N_j are unobserved. The intuition behind the ZTP mixture structure—and the motivation for employing the SPA was introduced at the beginning of Scenario B. In this section, we focus on the likelihood formulation and the ensuing saddle-point-based computation.

We assume that $N_j \sim \mathrm{ZTP}(\lambda)$. Conditional on $N_j = n$, the components of the jth aggregate are formed from independent gamma summands: $X_{ji} \sim \Gamma(\alpha_1, \beta_1)$ and $Y_{ji} \sim \Gamma(\alpha_2, \beta_2)$, with $\{X_{ji}\}$ being independent of $\{Y_{ji}\}$. The complete parameter vector is given by:

$$\theta = (\lambda, \alpha_1, \beta_1, \alpha_2, \beta_2)^\top.$$

.

Since N_j is unobserved, the likelihood incorporates the marginal density:

$$f_S(S_{1j}, S_{2j}; \theta) = \sum_{n=1}^{\infty} \Pr(N_j = n; \lambda)\, f_{S|N=n}(S_{1j}, S_{2j}; \theta),$$

which is typically intractable for exact evaluation. Consequently, we utilize the SPA, which is constructed from the CGF of (S_{1j}, S_{2j}). Closely related conditional and stopped-sum SPA formulations can be found in (Alhejaili and Al-Ghamedi, 2024b) and (Alhejaili and Abd-Elfattah, 2013), respectively.

Let the joint MGF of (X_{ji}, Y_{ji}) be given by:

$$M_{X,Y}(t,s) = (1-\beta_1 t)^{-\alpha_1}(1-\beta_2 s)^{-\alpha_2}, \qquad (t < 1/\beta_1,\ s < 1/\beta_2). \tag{13}$$

For the ZTP distribution, we have:

$$M_N(u) = \frac{e^{\lambda(e^u-1)} - e^{-\lambda}}{1-e^{-\lambda}}.$$

Thus, $M_S(t,s) = M_N\big(\log M_{X,Y}(t,s)\big)$, and the CGF is given by:

$$K(t,s) = \log M_S(t,s) = \log\left(\frac{e^{\lambda\left(M_{X,Y}(t,s)-1\right)} - e^{-\lambda}}{1-e^{-\lambda}}\right). \tag{14}$$

Saddle-point (t_0, s_0) is defined by the first-order conditions $\partial K/\partial t = S_{1j}$ and $\partial K/\partial s = S_{2j}$; explicitly, we have:

$$\begin{aligned}
&\frac{\lambda\alpha_1\beta_1 \exp\{\lambda(1-\beta_1 t)^{-\alpha_1}(1-\beta_2 s)^{-\alpha_2} - \lambda\}}{(1-\beta_1 t)^{\alpha_1}(1-\beta_2 s)^{\alpha_2}(1-\beta_1 t)\left[\exp\{\lambda(1-\beta_1 t)^{-\alpha_1}(1-\beta_2 s)^{-\alpha_2} - \lambda\} - e^{-\lambda}\right]} - S_{1j} = 0,\\
&\frac{\lambda\alpha_2\beta_2 \exp\{\lambda(1-\beta_1 t)^{-\alpha_1}(1-\beta_2 s)^{-\alpha_2} - \lambda\}}{(1-\beta_1 t)^{\alpha_1}(1-\beta_2 s)^{\alpha_2}(1-\beta_2 s)\left[\exp\{\lambda(1-\beta_1 t)^{-\alpha_1}(1-\beta_2 s)^{-\alpha_2} - \lambda\} - e^{-\lambda}\right]} - S_{2j} = 0.
\end{aligned} \tag{15}$$

We numerically solve (15) to obtain (t_0, s_0). Following the approach of (Alhejaili and AlGhamedi, 2024a), one may reduce this to a single fixed-point equation in m_{MGF} and substitute back to obtain (t_0, s_0). After calculating the saddle-point, evaluate the Hessian $K''(t_0, s_0)$ (with entries $\partial^2 K/\partial t^2$, $\partial^2 K/\partial s^2$, $\partial^2 K/\partial t\,\partial s$) for SPA normalization.

For ease of notation, we define:

$$m_{MGF} = (1-\beta_1 t)^{-\alpha_1}(1-\beta_2 s)^{-\alpha_2}, \quad A := 1-\beta_1 t,\ \ B := 1-\beta_2 s, \tag{16}$$

$$r_{1j} := \frac{S_{1j}}{\alpha_1\beta_1}, \qquad r_{2j} := \frac{S_{2j}}{\alpha_2\beta_2}, \qquad k := \alpha_1 + \alpha_2 + 1. \tag{17}$$

To enhance numerical stability in the score equations, we introduce the reweighting factor:

$$\lambda^\star m_{MGF} := \frac{\lambda\, e^{\lambda(m_{MGF}-1)}}{e^{\lambda(m_{MGF}-1)} - e^{-\lambda}} = \frac{\lambda}{1-e^{-\lambda m_{MGF}}},$$

It is important to note that $\lambda^\star(\cdot)$ is a function of m_{MGF} (not an additional parameter). With this notation, (15) can be expressed as:

$$S_{1j} = \lambda^\star(m_{MGF})\,\alpha_1\beta_1 A^{-\alpha_1-1}B^{-\alpha_2} = \lambda^\star(m_{MGF})\,\alpha_1\beta_1\,\frac{m_{MGF}}{A}, \tag{18}$$

$$S_{2j} = \lambda^\star(m_{MGF})\,\alpha_2\beta_2 A^{-\alpha_1}B^{-\alpha_2-1} = \lambda^\star(m_{MGF})\,\alpha_2\beta_2\,\frac{m_{MGF}}{B}. \tag{19}$$

Therefore, we have:

$$A = \frac{\lambda^\star(m_{MGF})\, m_{MGF}}{r_{1j}}, \qquad B = \frac{\lambda^\star(m_{MGF})\, m_{MGF}}{r_{2j}}. \tag{20}$$

Using $m_{MGF} = A^{-\alpha_1} B^{-\alpha_2}$ leads to the consistency relation:

$$\left(\lambda^\star(m_{MGF})\right)^{\alpha_1+\alpha_2} m_{MGF}^{\alpha_1+\alpha_2+1} = r_{1j}^{\alpha_1}\, r_{2j}^{\alpha_2}, \qquad m_{MGF} = \left(\frac{r_{1j}^{\alpha_1} r_{2j}^{\alpha_2}}{\left(\lambda^\star(m_{MGF})\right)^{\alpha_1+\alpha_2}}\right)^{1/k}.$$

Alternatively, a one-dimensional fixed point in m_{MGF} is given by:

$$m_{MGF}^k = \frac{r_{1j}^{\alpha_1} r_{2j}^{\alpha_2}}{\lambda^{\alpha_1+\alpha_2}} \left(1 - e^{-\lambda m_{MGF}}\right)^{\alpha_1+\alpha_2}.$$

Eliminating m_{MGF} from the expressions for A and B results in:

$$A = \left(\lambda^\star(m_{MGF})\right)^{1/k} \frac{r_{2j}^{\alpha_2/k}}{r_{1j}^{(\alpha_2+1)/k}}, \qquad B = \left(\lambda^\star(m_{MGF})\right)^{1/k} \frac{r_{1j}^{\alpha_1/k}}{r_{2j}^{(\alpha_1+1)/k}},$$

Thus, the saddle-point coordinates are given by:

$$t_0 = \frac{1}{\beta_1}\left[1 - \left(\lambda^\star(m_{MGF})\right)^{1/k} \frac{r_{2j}^{\alpha_2/k}}{r_{1j}^{(\alpha_2+1)/k}}\right], \qquad s_0 = \frac{1}{\beta_2}\left[1 - \left(\lambda^\star(m_{MGF})\right)^{1/k} \frac{r_{1j}^{\alpha_1/k}}{r_{2j}^{(\alpha_1+1)/k}}\right].$$

The SPA for the joint density can then be expressed as:

$$\hat{f}_S(S_{1j}, S_{2j}; \theta) \approx \frac{1}{2\pi\sqrt{\det K''(t_0, s_0)}} \exp\left(K(t_0, s_0) - t_0 S_{1j} - s_0 S_{2j}\right).$$

To enhance numerical stability, the CDF can be approximated using a bivariate normal proxy, and the PDF can be recovered through finite differencing as follows (Alhejaili and AlGhamedi, 2024a):

$$\hat{f}_S(S_{1j}, S_{2j}) \approx \frac{1}{4h^2}\left[F_S(S_{1j}+h, S_{2j}+h) - F_S(S_{1j}+h, S_{2j}-h) - F_S(S_{1j}-h, S_{2j}+h) + F_S(S_{1j}-h, S_{2j}-h)\right],$$

where $h = \max(|S_{1j}|, |S_{2j}|, 1) \times 10^{-4}$.
The resulting approximate log-likelihood is given by:

$$\log L(\theta) \approx \sum_{j=1}^{m} \log \hat{f}_S(S_{1j}, S_{2j}; \theta),$$

is maximized under positive (box) constraints using the L-BFGS-B algorithm. In practice, finite-difference gradients are sufficient. Standard errors can be obtained from the inverse of the observed information (via numerical Hessian) or through a parametric bootstrap. For stability, robust solvers are employed for the fixed-point in m_{MGF}, near-singular Hessians are regularized, and validation is conducted by comparing SPA-based densities with Monte Carlo benchmarks (bias/MSE over replicated datasets).

6. Testing in the BCZTPG Model

This section presents likelihood-based procedures for hypothesis testing in the BCZTPG model. Let $\{(S_{1j}, S_{2j})\}_{j=1}^{m}$ denote the observed bivariate sums generated by a shared ZTP count $N_j \sim \text{ZTP}(\lambda)$ and independent Gamma severities $X_{ji} \sim \Gamma(\alpha_1, \beta_1)$ and $Y_{ji} \sim \Gamma(\alpha_2, \beta_2)$. The exact evaluation of the likelihood involves an infinite mixture; therefore, we utilize the SPA to compute the log-likelihood. With CGF $K(t, s)$ and saddle-points (t_{0j}, s_{0j}) satisfying $K_t(t_{0j}, s_{0j}) = S_{1j}$ and $K_s(t_{0j}, s_{0j}) = S_{2j}$ ((14), (15)), the SPA log–likelihood for a parameter vector θ takes the usual form

$$\ell(\theta) \;\approx\; \sum_{j=1}^{m} \Big[K(t_{0j}, s_{0j}) - t_{0j} S_{1j} - s_{0j} S_{2j} - \log(2\pi) - \tfrac{1}{2} \log\{\det K''(t_{0j}, s_{0j})\} \Big],$$

where $K''(t_{0j}, s_{0j})$ is the Hessian with respect to (t, s) (including the off–diagonal terms).

6.1. Likelihood Ratio Test

Let $H_0 : g(\theta) = \mathbf{0}$ impose k smooth restrictions on the full parameter vector $\theta = (\lambda, \alpha_1, \beta_1, \alpha_2, \beta_2)^\top$. We denote by $\hat{\theta}$ the unrestricted SPA–based MLE and by $\hat{\theta}_0$ the restricted SPA–based MLE. The likelihood-ratio statistic is given by:

$$\Lambda = -2\Big\{ \ell\big(\hat{\theta}_0\big) - \ell\big(\hat{\theta}\big) \Big\}$$

is asymptotically χ^2_k under H_0 (Wilks, 1938), provided the usual regularity conditions are satisfied; otherwise, mixture limits may apply (Self and Liang, 1987). This framework encompasses several common hypotheses in the BCZTPG model: symmetry with $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$ ($k = 2$); equality of scales only, $\beta_1 = \beta_2$ ($k = 1$); exponential marginals, $\alpha_1 = \alpha_2 = 1$ ($k = 2$); tests on the count parameter, for instance, $\lambda = \lambda_0$ when testing λ. For a given observed value Λ_{obs}, the nominal p–value is $p = \Pr\{\chi^2_k \geq \Lambda_{\text{obs}}\}$; and rejection at significance level α occurs when $\Lambda_{\text{obs}} > \chi^2_{k,1-\alpha}$.

Since SPA provides an approximation of the exact likelihood, routine validation via a parametric bootstrap under H_0 is recommended for moderate sample sizes: simulated data at $\hat{\theta}_0$, refit the restricted and unrestricted models for each replicate, recompute Λ, and obtain a bootstrap p-value from the empirical reference distribution (Efron and Tibshirani, 1994; Davison and Hinkley, 1997). In large samples, SPA–LRT typically aligns well with the χ^2_k limit, while in smaller samples, the bootstrap corrects the finite-sample bias introduced by truncation and curvature (Daniels, 1954; Lugannani and Rice, 1980; Skovgaard, 2001; Butler, 2007).

For implementation, we replicate the estimation pipeline: First, maximize $\ell(\theta)$ without constraints to obtain $\hat{\theta}$; next, impose $g(\theta) = \mathbf{0}$ and maximize the restricted log–likelihood to obtain $\hat{\theta}_0$; finally, compute Λ and its nominal or bootstrap–calibrated p-value. Numerical stability is enhanced by optimizing over log–parameters ($\log\lambda,\ \log\alpha_1,\ \log\beta_1,\ \log\alpha_2,\ \log\beta_2$) to enforce positivity, utilizing a quasi-Newton method with box constraints and well-scaled initial values. Convergence checks should encompass gradient norms, step sizes, and reproducibility from multiple initializations; reporting the observed information (the inverse Hessian of ℓ at $\hat{\theta}$ and at $\hat{\theta}_0$) provides diagnostics and facilitates standard errors for ancillary summaries.

When the focus is on a subset of parameters, profile likelihood becomes convenient; for instance, to test the equality of scales $\beta_1 = \beta_2$, we handle the nuisance parameters $(\lambda, \alpha_1, \alpha_2)$ by profiling them under both H_0 and H_1 and forming the same Λ. The likelihood–ratio approach is also invariant to smooth reparameterizations; for the symmetry test, one may work with $(\alpha, \delta_\alpha, \beta, \delta_\beta)$ where $\delta_\alpha = \alpha_1 - \alpha_2$ and $\delta_\beta = \beta_1 - \beta_2$, and H_0 corresponds to $(\delta_\alpha, \delta_\beta) = (0, 0)$, which can enhance conditioning near the null. The $(1 - \alpha)$ region for a k–dimensional target is $\{\theta :\ \Lambda(\theta) \leq \chi^2_{k,1-\alpha}\}$, and the intervals of the likelihood ratio for one dimension arise by profiling all nuisance parameters.

Care must be taken in boundary or non-regular cases—for example, very small shape parameters or hypotheses that place parameters on the boundary of the parameter space—where the usual χ^2 limit may not apply and mixture limits or nonstandard asymptotics may arise (Self and Liang, 1987). In such cases, the parametric bootstrap under H_0 serves as the default approach. Finally, we recommend standardized reporting: State the null and k, provide $\hat{\theta}_0$ and $\hat{\theta}$, report Λ, the nominal χ^2_k p–value, and when applicable, the bootstrap p-value, accompanied by condition diagnos-

tics (optimizer status, gradient norms), and likelihood-ratio-based confidence intervals for key contrasts. With SPA and double–precision optimization, extremely small positive values of the statistic may arise (e.g., $\Lambda \approx 10^{-30}$) due to nearly identical restricted/unrestricted fits and floating-point rounding. We report such cases as $\Lambda \approx 0$, which results in $p \approx 1$ (i.e., no rejection of H_0).

6.2. Symmetry Test

We define the BCZTPG model as *marginally symmetric* when the Gamma distributions for X and Y coincide, specifically:

$$H_0 : \ \alpha_1 = \alpha_2, \ \beta_1 = \beta_2 \qquad \text{vs.} \qquad H_1 : \ \alpha_1 \neq \alpha_2 \ \text{ or } \ \beta_1 \neq \beta_2.$$

Since N_j is drawn from a common ZTP(λ), symmetry in the counts is inherently enforced, and the null hypothesis focuses on the gamma parameters. Under H_0 the parameterization reduces from $(\lambda, \alpha_1, \beta_1, \alpha_2, \beta_2)$ to (λ, α, β).
We implement a LRT by fitting (i) the unrestricted model and (ii) the restricted symmetric model, each using SPA-based MLE, and computing the symmetry LRT statistic as follows:

$$\Lambda_{sym} \ = \ -2\big\{\ell_{\text{restricted}} - \ell_{\text{unrestricted}}\big\}.$$

Under H_0 and standard regularity conditions (with SPA serving as an accurate surrogate for the true likelihood), $\Lambda_{sym} \overset{\cdot}{\sim} \chi^2_2$ since two equality constraints are imposed according to Wilks (1938). Large values of the statistic (small p–values) provide evidence *against* symmetry (i.e., they reject H_0); conversely, small values support symmetry.

For convenience, closed–form expressions for the saddle-points under symmetric restrictions can be used to accelerate the evaluation. By writing:

$$r_{1j} = \frac{S_{1j}}{\alpha\beta}, \qquad r_{2j} = \frac{S_{2j}}{\alpha\beta},$$

one can utilize the working quantities:

$$m_{\text{MGF}} = \left(\frac{(r_{1j}r_{2j})^{\alpha}}{\hat{\lambda}^{\,2\alpha}}\right)^{\frac{1}{2\alpha+1}}, \qquad t_{0j} = \frac{1}{\beta}\left[1 - \hat{\lambda}^{\,\frac{1}{2\alpha+1}}\frac{r_{2j}^{\frac{\alpha}{2\alpha+1}}}{r_{1j}^{\frac{\alpha+1}{2\alpha+1}}}\right], \quad s_{0j} = \frac{1}{\beta}\left[1 - \hat{\lambda}^{\,\frac{1}{2\alpha+1}}\frac{r_{1j}^{\frac{\alpha}{2\alpha+1}}}{r_{2j}^{\frac{\alpha+1}{2\alpha+1}}}\right],$$

with the MGF and CGF defined as shown in (13) and (14). In practice, we first optimize $\ell(\lambda, \alpha, \beta)$ under H_0, then optimize $\ell(\lambda, \alpha_1, \beta_1, \alpha_2, \beta_2)$ under H_1, and then report the LRT statistic, its χ^2_2 p–value, and a decision at a predetermined significance level (e.g., $\alpha = 0.05$). Due to the approximate nature of SPA, a parametric bootstrap calibrated at the restricted fit can be utilized to refine the p-value in smaller samples (Daniels, 1954; Lugannani and Rice, 1980; Skovgaard, 2001; Butler, 2007; Efron and Tibshirani, 1994; Davison and Hinkley, 1997).

7. Simulation Study

Following the development of the estimation procedures outlined in Sections 4 and 5, we evaluated their performance through a comprehensive simulation study. To evaluate the proposed procedures for the BCZTPG distribution, we conducted simulations across various data scenarios. In this study, we estimated the BCZTPG parameters using the MoM, the MLE, and the SPA-based MLE (Butler, 2007).

In this study, both the MoM and MLE were applied to evaluate the estimation accuracy of the BCZTPG parameters. For each method, three distinct data scenarios were considered, representing different levels of data completeness: (1) complete data, where $(N_j, (X_{ji}, Y_{ji})_{i=1}^{N_j})_{j=1}^{m}$ are observed; (2) partially observed data with known $(N_j, S_{1j}, S_{2j})_{j=1}^{m}$; and (3) partially observed data where only $(S_{1j}, S_{2j})_{j=1}^{m}$ are observed and N_j is unobserved. For each estimation method and scenario, experiments were conducted across various sample sizes, and performance was compared in terms of bias, variance, and confidence interval coverage (Klugman et al., 2012).

We selected the representative true values $\lambda = 5$, $\alpha_1 = 3$, $\alpha_2 = 4$, $\beta_1 = 2$, and $\beta_2 = 3$. To investigate the impact of sample size on accuracy and efficiency, we explored $m \in \{100, 1000, 10000, \text{and } 100000\}$. For each m, datasets were generated using the BCZTPG data-generating process described in Section 3. To simulate N_j, we utilized the ZTP PMF as described in (2) with sampling implemented via the inverse-CDF method: draw $U \sim \text{Unif}(e^{-\lambda}, 1)$ and set $N_j = F^{-1}_{\text{Poisson}(\lambda)}(U)$ (Johnson et al., 2005; Devroye, 1986).

For each replicate, we first draw $N_j \sim \text{ZTP}(\lambda)$ for $j = 1, \ldots, m$. Then, for $i = 1, \ldots, N_j$, we generate $X_{ji} \sim \Gamma(\alpha_1, \beta_1)$ and $Y_{ji} \sim \Gamma(\alpha_2, \beta_2)$ independently. The aggregates were $S_{1j} = \sum_{i=1}^{N_j} X_{ji}$ and $S_{2j} = \sum_{i=1}^{N_j} Y_{ji}$, yielding a sample of size m consisting of paired observations (S_{1j}, S_{2j}). No filtering for positivity was necessary since the gamma support and zero-truncation ensured valid observations (Johnson et al., 1994; Robert and Casella, 2004).

The parameters were then estimated using MoM, MLE, and MLE based on SPA, maximized with L-BFGS-B (Byrd et al., 1995; Zhu et al., 1997; Butler, 2007). For Scenario B, we employed a stable two-stage approach: first, obtain an MoM estimate of λ from the equation (5), and then, with λ fixed, estimate $(\alpha_1, \beta_1, \alpha_2, \beta_2)$ using an SPA-based MLE (a fully joint SPA–based MLE is also feasible). A bootstrap with 2000 resamples was employed to compute standard errors and 95% confidence intervals (Efron and Tibshirani, 1994; Davison and Hinkley, 1997). Bootstrap samples were created by resampling indices with replacement and re-running the estimation in parallel for reproducibility; runs that did not converge were flagged and excluded. In addition to Wald intervals $\hat{\theta} \pm 1.96$ SE, we also report the percentile intervals $[\hat{\theta}^{*(0.025)}, \hat{\theta}^{*(0.975)}]$.

Performance was summarized using the mean estimate, standard error (SE), bias (mean estimate minus true value), percent error (PE) defined as $\text{PE} = 100 \times |\text{Bias}|/\text{True}$, and 95% confidence intervals. Additionally, the coverage probability is the proportion of instances where the true value lies within the bootstrap 95% interval; simple coverage indicators are also noted to flag under- or over-coverage.

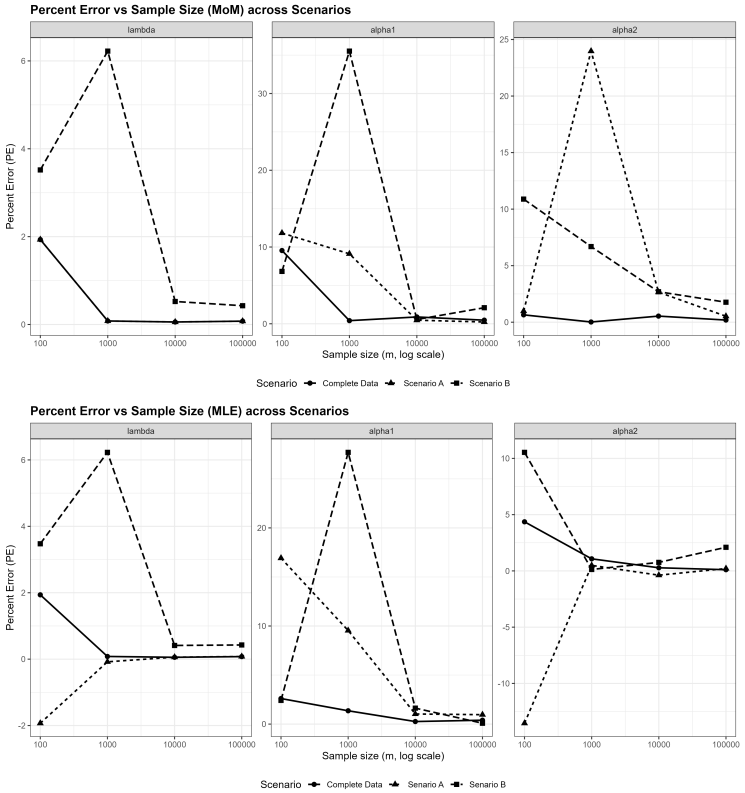

Figure 1: Percent error of the MoM and MLE estimators versus sample size m across the three data scenarios (Complete Data, Scenario A, Scenario B) for λ, α_1, α_2.

To supplement the tabulated results, we plot the percent error against sample size m, clearly distinguishing the three data scenarios (Complete data, Scenario A, and Scenario B). In Figure 1, each curve represents one scenario, allowing for direct comparison of estimator accuracy under varying levels of data availability for the parameters λ, α_1, and α_2. Across all scenarios, percent error diminishes as m increases, providing empirical evidence of consistency. Errors are typically largest in Scenario B, where counts are latent, and smallest under complete data, with Scenario A in between.

Table 1 provides comprehensive numerical summaries—point estimates, bias, standard errors, and confidence intervals—for each scenario and sample size. Overall, increasing m reduces both bias and variance, demonstrating the accuracy and efficiency improvements of the SPA-based estimator. For large samples (e.g., $m = 100{,}000$), empirical coverage approaches the nominal 95% level, consistent with asymptotic theory (van der Vaart, 1998). We estimate parameters and compare efficiency using both the MoM and the SPA-based MLE (Butler, 2007; Lehmann and Casella, 1998).

We also evaluate computational performance across increasing sample sizes and bootstrap configurations. For $m = 100,\ 1{,}000,\ 10{,}000,$ and 100,000, total runtime was approximately 14.66 s, 116.24 s, 2228.96 s, and 52,921.68 s, respectively. The mean time per bootstrap replication increased from 0.26 s to 25.97 s as m grew, indicating approximately linear scaling with respect to both m and the number of bootstrap replications. The likelihood optimization converged in every run (100% convergence), and no bootstrap failures were recorded. In all configurations, the bootstrap stage accounted for the majority of total computation, whereas the core SPA estimation remained fast and numerically stable, particularly when initialized with moment-based estimates.

Together, these results suggest that the proposed framework is both computationally scalable and practically feasible, especially when parallelization is employed. Furthermore, the additional computational cost associated with SPA is justified by the significant improvements in estimation accuracy and inferential reliability.

As an additional diagnostic tool, we conducted a symmetry LRT on an SPA-based fit. The routine reported $\Lambda = 16697.5984$ (df $= 2$) with $p < 2.2 \times 10^{-16}$; therefore, at conventional significance levels (e.g., $\alpha = 0.05$), the null hypothesis of symmetry is decisively rejected. This result provides strong evidence that a symmetric specification is insufficient for the simulated datasets and confirms the presence of asymmetry in the simulated data captured by the fitted model.

Finally, goodness-of-fit diagnostics on selected replicates—specifically, visual comparisons between the bivariate empirical distribution function (BEDF) and the fitted saddle-point CDF $\hat{F}_{\mathbf{S}}(x, y; \hat{\theta}_{\mathrm{MLE}})$—demonstrate close agreement, confirming that the SPA estimator accurately captures the underlying distribution (Genest and Rémillard, 2008). The detailed results for each scenario and estimation method are summarized in Table 1. The table reports the parameter estimates, bias, percent error, standard errors, and 95% confidence intervals for the MoM and SPA-based MLE across $m \in \{100, 1000, 10000, \text{and } 100000\}$.

A synthesis of the simulation results uncovers several important patterns. First, the SPA-based MLE consistently outperforms the MoMs, particularly in scenarios with latent counts, where bias reduction and interval precision are crucial. Second, the advantages of SPA are more pronounced for small and moderate sample sizes, as moment-based estimators often display higher variability under incomplete data structures. As the sample size increases, all estimators demonstrate enhanced accuracy, reflecting their asymptotic consistency; however, the SPA-based MLE retains a distinct efficiency advantage across the scenarios considered. From a practical standpoint, these findings indicate that SPA-based likelihood methods should be favored when event counts are unobserved or when high estimation accuracy is essential in compound modeling frameworks. In summary, the simulation study validates the robustness and practical relevance of the proposed SPA-based estimation framework.

8. Application to Montreal Precipitation and Snowfall Using the BCZTPG Model

The BCZTPG model was applied to a real climatological dataset to demonstrate the practical relevance of the proposed estimation procedures. *Monthly* aggregates were collected for Montreal, Canada, from **1952–2016**, sourced from the NOAA National Centers for Environmental Information (NCEI) Global Historical Climatology Network—Daily (GHCN-Daily; `https://www.ncei.noaa.gov/cdo-web/`). For month j, let S_{1j} represent the monthly total

Table 1: Simulation Results for MoM and MLE Estimators Across Scenarios and Sample Sizes

Scenario	m	Parameter	True	MoM					MLE				
				Estimate	Bias	Perc_Error	Lower_CI	Upper_CI	Estimate	Bias	Perc_Error	Lower_CI	Upper_CI
Complete Data	100	λ	5	4.9033	-0.0967	1.9333	4.4596	5.4114	4.9031	-0.0969	1.9374	4.4626	5.3437
		α_1	3	3.2863	0.2863	9.5437	3.0311	3.7356	3.0782	0.0782	2.6054	2.7131	3.4432
		α_2	4	4.0260	0.0260	0.6496	3.4468	4.4779	3.8254	-0.1746	4.3642	3.3675	4.2833
		β_1	2	1.8646	-0.1354	6.7704	1.6293	2.0548	1.9907	-0.0093	0.4642	1.7343	2.2471
		β_2	3	3.0185	0.0185	0.6167	2.6991	3.4285	3.1768	0.1768	5.8949	2.7705	3.5832
	1000	λ	5	4.9960	-0.0040	0.0806	4.8712	5.1419	4.9959	-0.0041	0.0810	4.8555	5.1364
		α_1	3	3.0122	0.0122	0.4073	2.8868	3.1556	2.9593	-0.0407	1.3582	2.8495	3.0690
		α_2	4	3.9993	-0.0007	0.0180	3.8129	4.2002	4.0429	0.0429	1.0732	3.8910	4.1949
		β_1	2	1.9735	-0.0265	1.3249	1.8735	2.0690	2.0088	0.0088	0.4407	1.9276	2.0900
		β_2	3	3.0113	0.0113	0.3776	2.8483	3.1683	2.9788	-0.0212	0.7061	2.8596	3.0980
	10000	λ	5	5.0029	0.0029	0.0572	4.9576	5.0500	5.0028	0.0028	0.0559	4.9583	5.0472
		α_1	3	3.0266	0.0266	0.8882	2.9832	3.0708	3.0078	0.0078	0.2606	2.9725	3.0431
		α_2	4	3.9784	-0.0216	0.5389	3.9233	4.0307	4.0111	0.0111	0.2776	3.9635	4.0587
		β_1	2	1.9810	-0.0190	0.9512	1.9504	2.0118	1.9934	-0.0066	0.3312	1.9679	2.0188
		β_2	3	3.0177	0.0177	0.5891	2.9754	3.0646	2.9931	-0.0069	0.2292	2.9553	3.0310
	100000	λ	5	5.0038	0.0038	0.0761	4.9894	5.0178	5.0038	0.0038	0.0768	4.9898	5.0179
		α_1	3	3.0141	0.0141	0.4705	3.0004	3.0269	3.0119	0.0119	0.39647	3.0007	3.0231
		α_2	4	4.0076	0.0076	0.1890	3.9882	4.0246	4.0043	0.0043	0.1066	3.9892	4.0193
		β_1	2	1.9911	-0.0089	0.4440	1.9820	2.0004	1.9927	-0.0073	0.3661	1.9846	2.0007
		β_2	3	2.9962	-0.0038	0.1259	2.9828	3.0109	2.9987	-0.0013	0.0446	2.9867	3.0107
Scenario A	100	λ	5	4.9033	-0.0967	1.9333	4.3823	5.2684	4.9033	-0.0967	-1.9300	4.4630	5.3440
		α_1	3	2.6453	-0.3547	11.8235	1.7284	6.5847	3.5075	0.5075	16.9200	2.5470	4.4680
		α_2	4	3.9606	-0.0394	0.9841	2.2629	12.1527	3.4586	-0.5414	-13.5400	2.5120	4.4050
		β_1	2	2.3164	0.3164	15.8213	0.9836	3.5224	1.7470	-0.2530	-12.6500	1.2690	2.2250
		β_2	3	3.0683	0.0683	2.2768	1.1433	5.1741	3.5137	0.5137	17.1200	2.5520	4.4760
	1000	λ	5	4.9960	-0.0040	0.0806	4.8737	5.1265	4.9960	-0.0040	-0.0800	4.8550	5.1360
		α_1	3	3.2734	0.2734	9.1133	2.6399	4.3722	3.2863	0.2863	9.5400	3.0020	3.5710
		α_2	4	3.0411	-0.9589	23.9736	2.4738	3.9247	4.0196	0.0196	0.4900	3.67107	4.3680
		β_1	2	1.8160	-0.1840	9.1980	1.3461	2.2483	1.8089	-0.1911	-9.5500	1.6520	1.9650
		β_2	3	3.9602	0.9602	32.0063	3.0546	4.8661	2.9961	-0.0039	-0.1300	2.7360	3.2560
	10000	λ	5	5.0029	0.0029	0.0572	4.9584	5.0464	5.0030	0.0030	0.0600	4.9590	5.0470
		α_1	3	2.9860	-0.0140	0.4659	2.7867	3.2241	3.0310	0.0310	1.0300	2.9480	3.1140
		α_2	4	4.1063	0.1063	2.6566	3.7678	4.4810	3.9849	-0.0151	-0.3800	3.8760	4.0940
		β_1	2	2.0079	0.0079	0.3963	1.8589	2.1496	1.9781	-0.0219	-1.0900	1.9240	2.0320
		β_2	3	2.9237	-0.0763	2.5420	2.6834	3.1868	3.0128	0.0128	0.4300	2.9300	3.0950
	100000	λ	5	5.0038	0.0038	0.0761	4.9901	5.0179	5.0040	0.0040	0.0800	4.9900	5.0180
		α_1	3	3.0070	0.0070	0.2318	2.9385	3.0781	3.0288	0.0288	0.9600	3.0030	3.0550
		α_2	4	3.9789	-0.0211	0.5276	3.8749	4.0877	4.0085	0.0085	0.2100	3.9740	4.0430
		β_1	2	1.9959	-0.0041	0.2069	1.9496	2.0411	1.9815	-0.0185	-0.9300	1.9640	1.9990
		β_2	3	3.0178	0.0178	0.5935	2.9373	3.0987	2.9955	-0.0045	-0.1500	2.9700	3.0220
Scenario B	100	λ	5	5.1759	0.1759	3.5184	4.0307	8.6563	5.1737	0.1737	3.4747	3.6303	6.6694
		α_1	3	2.7953	-0.2047	6.8241	1.4317	5.1981	3.0725	0.0725	2.4167	1.9178	8.6154
		α_2	4	4.4354	0.4354	10.8855	2.0192	9.3523	3.5786	-0.4214	10.5360	1.9258	15.2400
		β_1	2	2.0804	0.0804	4.0197	1.2417	2.9353	1.9129	-0.0871	4.3543	0.8854	3.0343
		β_2	3	2.6002	-0.3998	13.3266	1.2964	4.4429	3.2572	0.2572	8.5747	0.8037	5.6578
	1000	λ	5	4.6889	-0.3111	6.2229	4.1725	5.3239	4.6886	-0.3114	6.2276	4.1584	5.2984
		α_1	3	4.0657	1.0657	35.5237	3.2644	5.1144	3.8310	0.8310	27.7011	3.0699	4.8599
		α_2	4	3.7325	-0.2675	6.6881	2.9552	4.7270	3.9948	-0.0052	0.1288	3.2050	5.1133
		β_1	2	1.5541	-0.4459	22.2956	1.2915	1.8362	1.6607	-0.3393	16.9644	1.3412	1.9665
		β_2	3	3.4295	0.4295	14.3168	2.7499	4.1688	3.2265	0.2265	7.5485	2.5931	3.8508
	10000	λ	5	5.0262	0.0262	0.5238	4.8523	5.2367	5.0206	0.0206	0.4124	4.8310	5.2190
		α_1	3	2.9832	-0.0168	0.5616	2.7878	3.1992	2.9512	-0.0488	1.6262	2.7606	3.1472
		α_2	4	4.1078	0.1078	2.6944	3.7686	4.4855	4.0303	0.0303	0.7578	3.7162	4.3960
		β_1	2	2.0008	0.0008	0.0420	1.8862	2.1109	2.0357	0.0357	1.7826	1.9226	2.1405
		β_2	3	2.9096	-0.0904	3.0150	2.6987	3.1279	2.9848	-0.0152	0.5078	2.7831	3.1900
	100000	λ	5	4.9786	-0.0214	0.4277	4.9156	5.0419	4.9786	-0.0214	0.4274	4.9174	5.0382
		α_1	3	3.0628	0.0628	2.0922	2.9973	3.1278	3.0026	0.0026	0.0870	2.9433	3.0683
		α_2	4	4.0707	0.0707	1.7664	3.9632	4.1859	4.0839	0.0839	2.0985	3.9782	4.1978
		β_1	2	1.9691	-0.0309	1.5467	1.9354	2.0052	2.0196	0.0196	0.9822	1.9838	2.0545
		β_2	3	2.9642	-0.0358	1.1936	2.8937	3.0318	2.9709	-0.0291	0.9693	2.9050	3.0367

precipitation (mm; the sum of daily mm within month j) and S_{2j} represent the monthly total snowfall (mm; the sum of daily mm within month j). The latent event count N_j (e.g., the number of precipitation or snow events during month j) was unobserved unless stated otherwise. This setting aligns with the observation regimes considered in this study (Scenario A and Scenario B) and enables comparisons of inferences as information about N_j becomes more limited (Wilks, 2011).

In this context, the latent count N_j was estimated based on the number of rainy days in each month. The aggregated rainfall and snowfall were defined as S_{1j} and S_{2j}, respectively. In this context, S_{1j} represents the monthly total precipitation obtained by summing daily precipitation over rainy days only, while S_{2j} represents the monthly total snowfall accumulated on those same rainy days. Thus, (S_{1j}, S_{2j}) represents a monthly rainfall–snowfall pair conditional on $N_j > 0$, and within the BCZTPG framework, these are modeled as compound sums of Gamma random variables with compounding variable $N_j \sim \text{ZTP}(\lambda)$. For Scenario A regimes, N_j may be approximated by the number of event days in month j (e.g., days with precipitation and/or snowfall); for Scenario B, N_j is treated as unobserved. In all cases, months with $S_{1j} > 0$ and $S_{2j} > 0$ were retained to reflect the zero-truncated event mechanism of the model. Months with zeros were excluded to align with the BCZTPG structure, which assumes a ZTP count for event frequency (Menne et al., 2012; Johnson et al., 2005).

In Section 5, the parameters were estimated using MoM and SPA–based MLE, with MoM estimates employed to initialize the SPA optimizer. Uncertainty was quantified using a nonparametric bootstrap with 300 resamples, and we report the standard errors and 95% percentile intervals (Butler, 2007; Daniels, 1954; Lugannani and Rice, 1980; Skovgaard, 2001; Efron and Tibshirani, 1994; Davison and Hinkley, 1997).

Table 2 presents the *monthly* descriptive statistics for the totals (S_1, S_2) used in the estimation. Both marginals exhibit positive skewness, supporting the use of Gamma components for (S_1, S_2).The monthly scatter and kernel density contours demonstrate a strong concentration near the origin with elongation toward the upper right, indicating a positive association and some upper-tail co-variation—features effectively captured by the BCZTPG specification (Johnson et al., 1994; Scott, 2015; Nelsen, 2006). Figure 2 illustrates a dense band along $S_2 \approx 0$ while $S_1 > 0$ (rainy, snow-free months) and a sparse upper-right tail corresponding to winter months with heavy snowfall. The spread of S_2 increases with S_1, indicating heteroscedasticity in the monthly totals. This geometry aligns with BCZTPG's positively associated Gamma marginals compounded by a zero-truncated event mechanism. The empirical contours in Figure 3 display a single dominant mode at small totals, with elongation toward the upper right, indicating positive association and upper-tail co-variation. This geometry is consistent with the *rejection of symmetry* reported below, and the clear asymmetry between the marginals supports permitting distinct gamma parameters for precipitation and snowfall within the BCZTPG specification.

Table 2: Descriptive Statistics for Precipitation and Snowfall, 1952–2016.

Variable	Obs.	Mean	Std. Dev.	Min	Q_1	Median	Q_3	Max	Skewness
N: Rainy Days Per Months	774	13.6047	3.6084	2.0000	11.0000	14.0000	16.0000	25.0000	-0.0300
S_1: Precipitation (mm, monthly total)	774	3.2072	1.4092	0.0200	2.2325	3.0450	4.0600	8.9600	0.6184
S_2: Snowfall (mm, monthly total)	774	7.1845	10.2527	0.0000	0.0000	0.4000	12.6750	52.1000	1.5207

The parameter estimates from both estimation approaches are summarized in Table 3. From a climatological perspective, the estimated Poisson parameter reflects the average frequency of precipitation-related events within a month, suggesting relatively consistent wet periods in Montreal. Differences between the gamma parameters for precipitation and snowfall indicate distinct accumulation dynamics associated with warm and cold season processes. In particular, the greater dispersion observed in snowfall totals corresponds to the episodic and highly variable nature of winter storms, while rainfall generally displays more gradual accumulation patterns. These findings align with the hydroclimatic behavior of cold-region environments, where snowfall is more intermittent yet can produce substantial accumulations over short durations. Additionally, the positive dependence captured by the compound structure suggests shared large-scale meteorological drivers, such as frontal systems, that may concurrently affect rainfall and snowfall totals. Collectively, this interpretability highlights the practical value of the BCZTPG framework in associating statistical inference with physically meaningful climate dynamics.

The confidence intervals correspond to the bootstrap percentiles. The SPA-based MLE method provides more stable

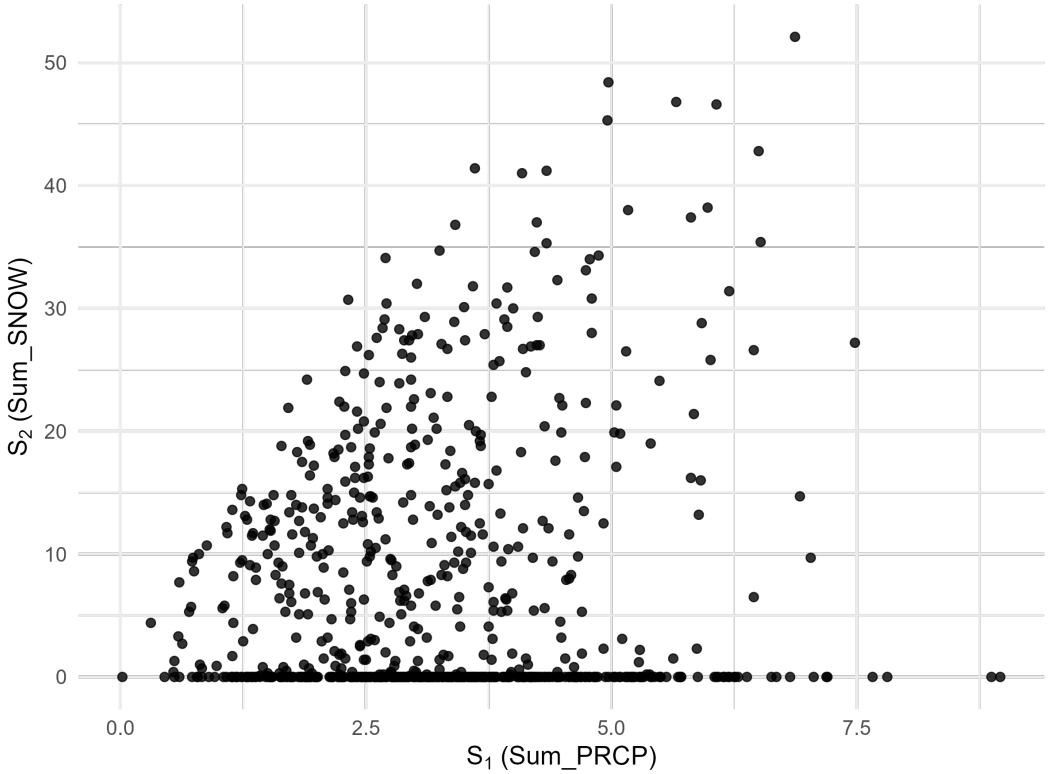

Figure 2: Scatter Plot of *Monthly* Totals S_1 Precipitation and S_2 Snowfall, 1952–2016.

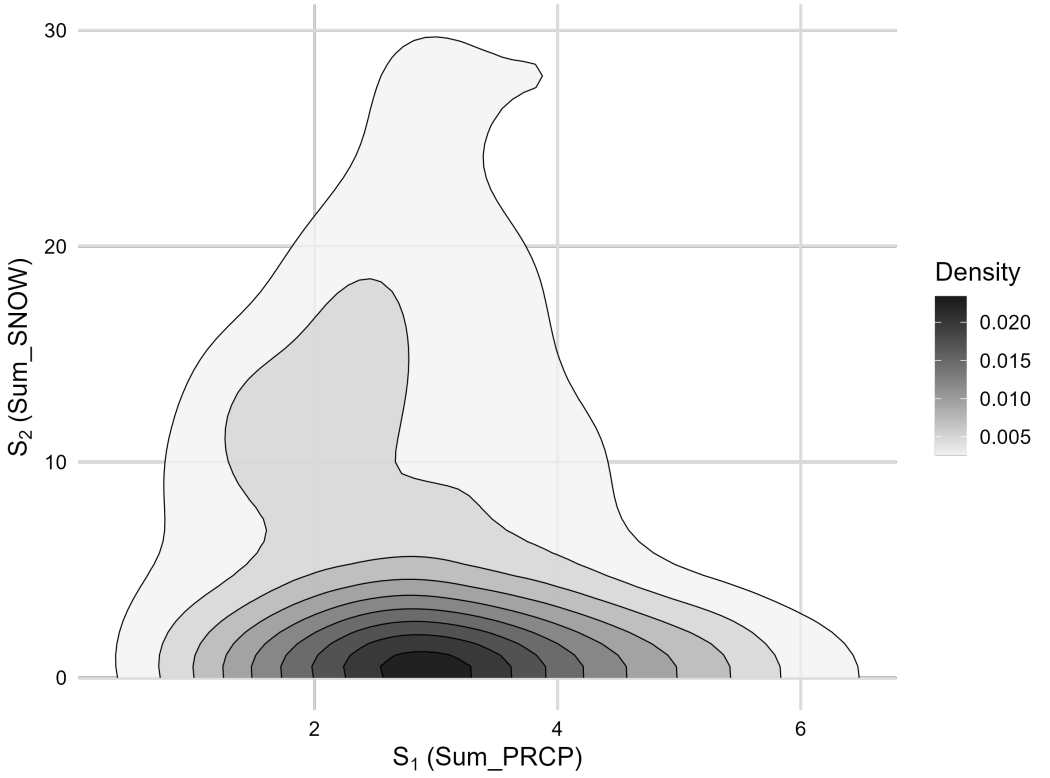

Figure 3: Empirical Kernel-Density Contours for *Monthly* Pairs S_1 and S_2, 1952–2016.

estimates with lower standard errors, confirming the utility of SPA for making inferences (Butler, 2007; Efron and Tibshirani, 1994; Davison and Hinkley, 1997).

Table 3: Estimates of BCZTPG Parameters with 95% Confidence Intervals by Scenario and Estimation Method, Based on Montreal Monthly Data (1952–2016).

Scenario	Method	$\hat{\lambda}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
Complete Data	MoM	13.6046 (13.3398,13.8660)	0.5516 (0.5274,0.5804)	0.1428 (0.1270,0.1584)	0.4273 (0.4039,0.4502)	3.6988 (3.4548,3.9793)
	MLE	13.6046 (13.3448,13.8645)	0.7309 (0.7140,0.7478)	0.0981 (0.0961,0.1001)	0.3225 (0.3122,0.3329)	5.3830 (5.0376,5.7285)
Scenario A	MoM	13.6046 (13.3398,13.8777)	0.5991 (0.5115,0.7184)	0.0374 (0.0332,0.0423)	0.3935 (0.3322,0.4579)	14.1258 (12.8935,15.3606)
	MLE	14.7910 (14.4170,15.1650)	0.4379 (0.3790,0.4960)	0.0910 (0.0800,0.1020)	0.4718 (0.4090,0.5350)	10.1477 (8.8940,11.4020)
Scenario B	MoM	13.7706 (7.7342,31.5653)	0.6030 (0.2025,1.8692)	0.0370 (0.0163,0.0707)	0.3863 (0.2232,0.5512)	14.1095 (13.0783,15.3629)
	MLE	7.9072 (5.9303,10.9445)	1.4460 (0.8477,2.7112)	0.2661 (0.1671,0.3631)	0.1540 (0.1078,0.2258)	11.1703 (9.2064,13.0806)

To evaluate the marginal adequacy of the BCZTPG fit, we compared the empirical distributions of the monthly totals (S_1, S_2) to their fitted marginal CDFs under SPA–based MLE, denoted $\widehat{F}_1(s) = \Pr_{\hat{\theta}}(S_1 \le s)$ and $\widehat{F}_2(s) = \Pr_{\hat{\theta}}(S_2 \le s)$. For a sample size m, we define the probability–integral–transform (PIT) values as follows:

$$U_{1j} = \widehat{F}_1(S_{1j}), \qquad U_{2j} = \widehat{F}_2(S_{2j}), \qquad j = 1, \ldots, m.$$

If the model is correctly specified, (U_{1j}) and (U_{2j}) are i.i.d. Uniform$(0,1)$. We assessed the fit using (i) QQ-plots of the empirical quantiles of S_1 and S_2 against the fitted quantiles $\widehat{F}_k^{-1}\big((r-0.5)/m\big)$, $k \in \{1,2\}$, and (ii) one-sample Kolmogorov-Smirnov (KS) tests on the PITs:

$$D_k = \sup_{u\in[0,1]} \big|\widehat{G}_k(u) - u\big|, \qquad p_k = \Pr\big(D_k^{(m)} \ge D_k\big), \quad k = 1,2,$$

where $\widehat{G}_k$ is the empirical CDF of $(U_{kj})_{j=1}^m$ and $D_k^{(m)}$ is the KS statistic under Uniform$(0,1)$ (we report the standard asymptotic p-values).

Table 4: Kolmogorov–Smirnov Tests of PITs from the Fitted Marginals $\widehat{F}_1$ and $\widehat{F}_2$. Larger p-values Indicate no Evidence of Departure from Uniform$(0,1)$**.**

Variable	m	KS statistic D	p-value
$U_1 = \widehat{F}_1(S_1)$	774	0.0343	0.7665
$U_2 = \widehat{F}_2(S_2)$	774	0.0994	0.0012

The QQ-plots in Figure 4 demonstrate close adherence to the 45° line over most of the support, with slight deviations in the extreme upper tail, which is typical for monthly aggregates. The KS test for S_1 does not reject uniformity at conventional significance levels, while the test for S_2 indicates a modest departure primarily driven by upper-tail behavior (Table 4). Overall, the diagnostics indicate an adequate marginal fit for practical applications. Collectively, the goodness-of-fit diagnostics suggest that the BCZTPG model offers a significantly improved representation of the empirical distribution compared to simpler independent gamma formulations, especially in capturing the observed dependence structure and upper-tail behavior. This underscores the practical advantage of jointly modeling precipitation and snowfall through a compound framework, rather than treating them as unrelated processes.

We also evaluated joint symmetry in the bivariate distribution of precipitation and snowfall. Using the LRT on an

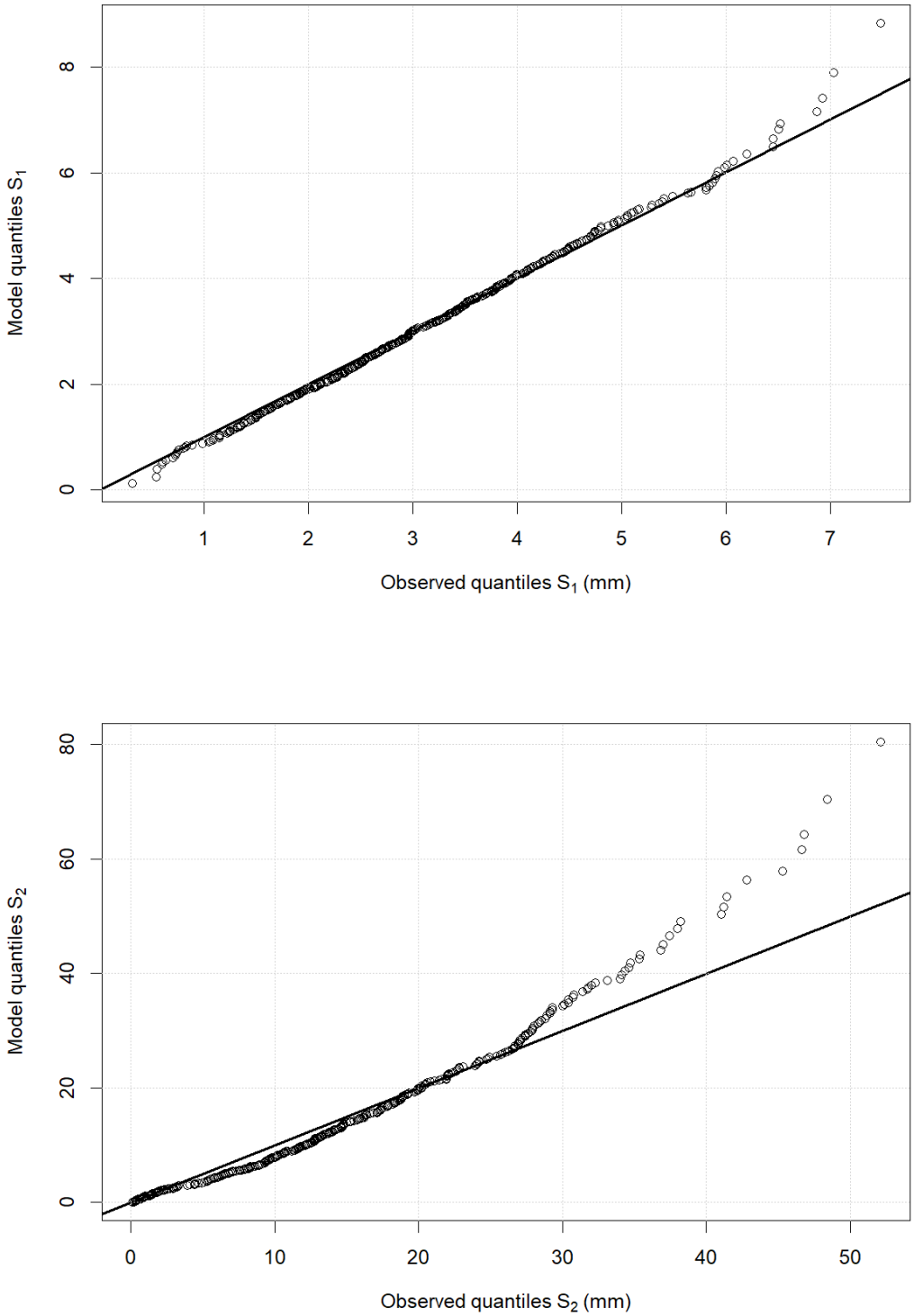

Figure 4: Q–Q Plots of Empirical Quantiles of S_1 (Top) and S_2 (Bottom) Versus the Fitted CZTPG Marginal Quantiles (SPA-based MLE), Monthly Data, 1952–2016.

SPA-based fit, we obtained $\Lambda = 411.7189$ with $p \approx 3.95 \times 10^{-90}$ under the χ^2_2 reference (Wilks 1938); thus, the null hypothesis of symmetry is decisively rejected. We also compared the full BCZTPG specification with a restricted model in which the rainfall and snowfall marginals share common parameters; the likelihood favored the unrestricted specification, indicating that the two processes are not governed by identical marginal shapes and scales. These findings highlight the utility of a flexible bivariate formulation like BCZTPG for climatological aggregates (Genest and Rémillard 2008). Overall, these findings indicate that the BCZTPG model is not only statistically adequate but also capable of capturing essential structural features of real climatological systems.

In conclusion, the BCZTPG model effectively captured both the marginal and joint features of the Montreal monthly rainfall and snowfall records from **1952–2016**, validating its effectiveness for aggregated climatological measurements when event counts were only partially observed (Wilks, 2011).

9. Conclusion

In this study, we developed and validated efficient estimation procedures for the BCZTPG distribution across various data scenarios, including situations where the latent count variable N_j is unobserved. Two complementary approaches were proposed: a classical moment-based estimation method and a high-precision MLE technique utilizing SPA. By deriving a numerically stable and accurate approximation of the SPA-based CDF and applying numerical differentiation, we obtained a tractable likelihood expression suitable for direct maximization. Extensive simulation studies confirmed the statistical advantages of SPA–based MLE over the moment method, particularly regarding bias reduction, lower mean-squared errors, and tighter confidence intervals. The practical applicability of the proposed framework was demonstrated using monthly precipitation and snowfall data from Montreal, where the BCZTPG model effectively captured both marginal skewness and joint dependence.

Despite its strong empirical performance, the SPA-based framework has several notable limitations. First, its numerical behavior may be sensitive with large Poisson rate parameters λ, necessitating careful initialization and stable optimization routines. Second, identifiability can be problematic in Scenario B, where unobserved latent counts may lead to multiple parameter configurations that generate similar distributions. Third, the dependence structure assumed in the BCZTPG model treats gamma components as conditionally independent given the latent count, which may not fully account for residual cross-component dependence arising from unobserved factors. Future research could investigate more flexible dependence formulations, such as copula-based extensions, to improve the model's applicability in contexts exhibiting stronger joint behavior.

Future research directions include incorporating dependence between rainfall and snowfall components through correlated gamma distributions or copula-based structures, thereby enhancing the model's flexibility for applications in insurance and environmental sciences. Additionally, alternative inference paradigms warrant exploration, particularly Bayesian approaches such as MCMC, ABC, and particle filters, which could provide valuable insights into computational-accuracy trade-offs. Robustness checks, including sensitivity analyses of the SPA solver, differentiation step size, and model misspecification, would further improve reliability. Finally, developing a dedicated R package would promote accessibility and reproducibility for practitioners. In conclusion, the BCZTPG model, along with the proposed estimation strategies, establishes a robust framework for modeling positive-valued bivariate sums over random counts, demonstrating utility in climatology and potential applications in actuarial science, reliability engineering, and related fields. We also introduced a formal symmetry test and a LRT for nested model comparison, providing rigorous tools for assessing both symmetry and overall model adequacy.

Acknowledgments

The authors would like to thank the editor and referees for their careful reading and insightful comments, which greatly improved the quality of this paper.

References

1. Alhejaili, A. D. and Abd-Elfattah, E. F. (2013). Saddlepoint approximations for stopped-sum distributions. *Communications in Statistics-Theory and Methods*, 42(20):3735–3743.
2. Alhejaili, A. D. and AlGhamedi, A. A. (2024a). Approximation methods for the bivariate compound truncated poisson gamma distribution. *Pakistan Journal of Statistics and Operation Research*, pages 285–299.
3. Alhejaili, A. D. and AlGhamedi, A. A. (2024b). Conditional saddle-point approximations for bivariate compound distributions. *Journal of King Abdulaziz University: Sciences*, 34(1).
4. Alhejaili, A. D. and AlGhamedi, A. A. (2025). A review of saddle-point approximation: Theory and applications. *Advances and Applications in Statistics*, 92(3):343–390.
5. Asmussen, S. (2000). Saddlepoint approximations for insurance risk. *Scandinavian Actuarial Journal*, 2000(1):69–98.
6. Bowers, N. L. (1997). Actuarial mathematics. *The Society of Actuaries*.
7. Butler, R. W. (2007). *Saddlepoint Approximations with Applications*. Cambridge University Press, Cambridge.
8. Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208.
9. Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*. Number 53. Cambridge university press.
10. Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, pages 631–650.
11. Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.
12. Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer, New York.
13. Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Chapman and Hall/CRC, New York.
14. Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in copulas. *Annals of Statistics*, 36(2):846–870.
15. Grogger, J. T. and Carson, R. T. (1991). Models for truncated counts. *Journal of applied econometrics*, 6(3):225–238.
16. Huzurbazar, V. (1956). Saddle-point approximations in statistics. *The Annals of Mathematical Statistics*, 27(2):253–268.
17. Jensen, J. L. (1991). Saddlepoint approximations to the distribution of the total claim amount in some recent risk models. *Scandinavian Actuarial Journal*, 1991(2):154–168.
18. Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate Discrete Distributions*. Wiley, Hoboken, NJ, 3 edition.
19. Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Volume 1*. Wiley, New York, 2 edition.
20. Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: From Data to Decisions*. Wiley, Hoboken, NJ, 4 edition.
21. Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. CRC Press, Boca Raton, 1st edition. eBook published November 22, 2017.
22. Kolassa, J. (2003a). Saddlepoint distribution function approximations in biostatistical inference. *Statistical Methods in Medical Research*, 12:59 – 71.
23. Kolassa, J. E. (2003b). Saddlepoint approximations for bivariate distributions. *Statistics & Probability Letters*, 65(3):273–279.
24. Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Springer, New York, 2 edition.
25. Lugannani, R. and Rice, S. (1980). Saddle point approximation for the distribution of the sum of independent random variables. *Advances in applied probability*, 12(2):475–490.
26. McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, revised edition.
27. Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G. (2012). An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, 29(7):897–910.
28. Nascimento, A., Rêgo, L. C., and Silva, J. W. (2023). Compound truncated poisson gamma distribution for understanding multimodal sar intensities. *Journal of Applied Statistics*, 50(6):1358–1377.

29. Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer Series in Statistics. Springer, New York, 2 edition.
30. Reid, N. (1988). Saddlepoint methods and statistical inference. *Statistical Science*, pages 213–227.
31. Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York, 2 edition.
32. Rossi, R. J. (2018). *Mathematical statistics: an introduction to likelihood based inference*. John Wiley & Sons.
33. Scott, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, Hoboken, NJ, 2 edition.
34. Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610.
35. Skovgaard, I. M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics*, 28(1):3–32.
36. Sorenson, H. W. (1980). Parameter estimation: principles and problems. *Marcel Dekker, New York*.
37. van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.
38. Wang, S. (1990). Saddlepoint approximations for bivariate distributions. *Journal of applied probability*, 27(3):586–597.
39. Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. Academic Press, Amsterdam, 3 edition.
40. Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9(1):60–62.
41. Willmot, G. E. and Lin, X. S. (2007). On compound distributions with mixed exponential claim amounts. *North American Actuarial Journal*, 11(3):76–92.
42. Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software*, 23(4):550–560.