AIMS Mathematics, 10(12): 30384–30428.
DOI: 10.3934/math.20251334
Received: 21 September 2025
Revised: 15 November 2025
Accepted: 01 December 2025
Published: 25 December 2025

https://www.aimspress.com/journal/Math

Research article

The doubly generalized exponential-geometric frailty distribution

Mohieddine Rahmouni*

Applied College, King Faisal University, Al-Ahsa, Saudi Arabia

* **Correspondence:** Email: mrahmouni@kfu.edu.sa.

Abstract: We proposed the doubly generalized exponential-geometric frailty (DGEGF) distribution, a hierarchical lifetime model for settings with a decreasing hazard, a random number of failure-prone components, and shared latent heterogeneity. The construction combined a geometric k-out-of-n failure rule, in which the system failed at the kth component failure rather than at the first, with gamma frailty acting on exponential component lifetimes. This hierarchy implied that every member of the family has a strictly decreasing failure rate, so the model was intended for burn-in or early-failure reliability data and for heterogeneous survival cohorts where risk decayed over time. We derived closed-form expressions for the marginal density, distribution, and survival functions and showed that the model was identifiable for both fixed and unknown k. A Monte Carlo study over several parameter regimes indicated that the baseline rate and geometric parameter were accurately estimated in moderate samples, whereas the frailty parameter can be highly variable in small samples, in line with known numerical-identifiability issues in multi-parameter lifetime models. In four benchmark applications, we compared DGEGF with exponential, exponential-geometric, and shared-frailty alternatives. The results showed that, under decreasing hazards, DGEGF offered a transparent way to encode redundancy and unobserved heterogeneity while remaining competitive in fit. We also indicated how the same hierarchical construction can be coupled with Weibull or log-logistic baselines to accommodate non-monotone hazards when needed.

Keywords: DGEGF distribution; geometric mixing; frailty; order statistics; decreasing failure rate; reliability
Mathematics Subject Classification: 60E05, 62E15, 62F10

1. Introduction

Time-to-event data arises in many scientific fields, including reliability engineering [1–3], biomedical research [4–7], economics [8, 9], and actuarial science [10, 11]. Survival and lifetime models provide the basis for quantifying risks, comparing designs or treatments, and forecasting the

timing of events such as mechanical failure, disease recurrence, or default [12, 13]. Classical parametric families, most prominently the exponential and Weibull distributions, remain widely used because of their interpretability and analytical tractability. However, real systems often exhibit features such as latent heterogeneity, competing risks, random system sizes, and threshold-based failure rules that are only partially captured by these baselines [14]. A large literature has therefore developed on generalized and compounded lifetime models, including numerous reviews of Weibull-type extensions and their reliability applications [15–17], as well as survey articles on frailty-based survival models and their role in explaining heterogeneity in time-to-event data [14, 18].

One productive strand of this work compounds a continuous lifetime distribution with a discrete count model to represent a random number of latent "failure opportunities". A foundational example is the exponential–geometric (EG) distribution proposed by Adamidis and Loukas [19], which models the time to first failure in a system with a geometrically distributed number of exponentially distributed components. The EG model accommodates overdispersion and decreasing hazard rates, and has inspired a broad family of geometric and power-series mixtures in reliability and survival analysis [17, 20, 21]. Parallel developments generalize the baseline (e.g., Weibull-geometric or exponential-Poisson models) to capture monotone, bathtub-shaped, and unimodal hazard profiles while retaining a random-count interpretation [17, 20, 22].

A second strand uses order statistics to represent failure times in systems with redundancy or threshold-based mechanisms. In k-out-of-n configurations, load-sharing networks, and multicomponent devices, system failure occurs only after the kth component has failed. This naturally leads to modeling the time to the kth failure using order statistics, with applications in both engineering and biomedical contexts [23–27]. A third and complementary strand introduces frailty models, which multiply the hazard by an unobserved random effect (often gamma or lognormal) to capture latent variability arising from manufacturing, environmental, or genetic factors [14, 18]. Frailty mixing can substantially alter the population hazard shape, for example, by inducing decreasing failure rates (DFR) through early selection of high-risk units.

While frailty models, order-statistic constructions, and random opportunity counts are each well developed, they are typically treated in isolation; in particular, models that simultaneously allow (i) randomness in system size, (ii) threshold-based failure, and (iii) shared frailty within a single analytically tractable hierarchy remain relatively rare in the literature, especially when paired with a clear reliability interpretation and systematic study of identifiability.

In this paper we propose the doubly generalized exponential-geometric frailty (DGEGF) distribution, a survival model that integrates these three ingredients. In its basic form, the model defines failure as the kth order statistic among a random number of conditionally exponential lifetimes, where (a) the number of components follows a k-truncated geometric distribution and (b) each component's hazard rate is scaled by a gamma-distributed frailty variable. This hierarchical construction yields a marginal distribution that (i) generalizes the EG model of Adamidis and Loukas [19], (ii) reduces to a gamma-frailty k-out-of-n model when the component count is fixed, and (iii) allows, in principle, alternative frailty specifications at the cost of losing some closed-form tractability.

Because the DGEGF model is obtained by exponential mixing over a gamma frailty, its population hazard rate is monotonically decreasing: heterogeneity induces an early selection effect and a declining aggregate hazard over time [14, 18]. The proposed distribution is therefore intentionally specialized. It

targets burn-in or heterogeneity-dominated regimes, where risk is highest shortly after deployment and then decays, rather than classic wear-out or bathtub-shaped patterns for which Weibull-type models are known to be more appropriate [15,16]. In this sense, DGEGF is not intended as a universal replacement for standard models, but as a structured alternative tailored to a specific and practically important class of failure mechanisms.

Why DGEGF?

The DGEGF distribution is motivated by applications where three features co-occur but are rarely modeled jointly: randomness in the number of latent failure opportunities, thresholded failure, and unobserved shared heterogeneity.

(1) Random opportunity counts: Many systems exhibit unit-specific opportunity counts (for example, micro-defects in electronics, potential crack-initiation sites in materials, or on-demand submodules in standby configurations). A geometric-type model for the latent count N provides a parsimonious way to represent this variability and extends the intuition of the EG family [19–21]. The k-truncated specification used in DGEGF ensures that the system contains at least k potential failure modes.
(2) Thresholded failure via order statistics: In redundant architectures and multi-step biological processes, failure or event occurrence requires multiple underlying "hits". Modeling the time to the kth failure via order statistics is natural in k-out-of-n reliability theory and allows k to play the role of a tolerance parameter that governs robustness to early breakdowns [23–27].
(3) Shared frailty and decreasing hazards: Gamma frailty introduces a multiplicative random effect on the component hazard, capturing unmeasured cluster-level risk factors such as manufacturing batches, patient-level biology, or environmental conditions. Exponential-gamma mixing yields a marginal distribution with a strictly decreasing population hazard, reflecting early removal of high-risk units and convergence toward a lower-risk subpopulation [14, 18].

Unlike EG-type models, which incorporate random N but omit thresholding and frailty [19, 20], or standard frailty models, which allow heterogeneity but assume a fixed cluster size [14, 18], the DGEGF distribution embeds all three mechanisms within a single coherent hierarchy. This leads to interpretable parameters: k encodes the failure rule (system fails at the kth hit), the geometric parameter η controls dispersion in the effective number of active failure opportunities, and the frailty shape parameter λ regulates early-life risk and tail thickness.

The model is most appropriate when (i) the hazard is believed to be decreasing over time, (ii) failure can plausibly be viewed as a k-hit phenomenon in a randomly sized system, and (iii) there is evidence of shared, unmeasured heterogeneity. In such settings, DGEGF can jointly represent heavy early tails, tolerance-to-failure through k, and residual frailty within an analytically tractable framework. By contrast, in strictly increasing or bathtub-shaped hazard regimes, classical Weibull or related extended lifetime models remain the preferred tools [15–17].

The primary objective of this paper is to formally introduce the DGEGF distribution, derive its main theoretical properties, and assess its practical performance in regimes where decreasing hazards and latent heterogeneity are scientifically plausible. Section 2 presents the model construction and marginal distribution. Section 3 develops key analytical properties, including the survival and hazard functions, moments, quantiles, and special/limiting cases. Section 4 describes maximum likelihood

estimation via an EM algorithm and discusses computational details. Section 5 reports a simulation study that examines finite-sample behavior and numerical identifiability. Section 6 analyzes four real datasets from reliability and survival analysis, highlighting regimes where DGEGF is competitive with or superior to classical models and regimes where it is not. Finally, Section 7 summarizes the findings, candidly discusses limitations, and outlines directions for future work, including non-exponential baselines to accommodate non-monotone hazards.

2. Model specification

The DGEGF* distribution is a hierarchical lifetime model that jointly accounts for (i) unobserved heterogeneity across units, (ii) randomness in the number of latent failure opportunities, and (iii) thresholded failure based on a k-out-of-n rule. In reliability terms, the system fails not necessarily at the first component failure, but when the kth component fails, with $k \geq 1$ fixed. The components are defined as follows:

2.1. Hierarchical construction

The DGEGF distribution is defined through four stochastic layers:

(1) Shared frailty. Let Z denote an unobserved multiplicative risk factor shared by all components in the same system (or by all event processes in the same subject). We assume

$$Z \sim \text{Gamma}(\lambda, \lambda),$$

with shape and rate $\lambda > 0$. The corresponding probability density function (PDF) is

$$g(z) = \frac{\lambda^\lambda}{\Gamma(\lambda)} z^{\lambda-1} e^{-\lambda z}, \qquad z > 0.$$

Under this parameterization, $\mathbb{E}[Z] = 1$ and $\text{Var}(Z) = 1/\lambda$. Small λ therefore corresponds to high frailty variance (strong unobserved heterogeneity), which in turn induces a strongly decreasing population hazard.

(2) Random number of opportunities. Let N denote the (latent) number of available failure opportunities/components. We assume N follows a k-truncated geometric distribution [28]:

$$\Pr(N = n) = (1 - \eta)\, \eta^{n-k}, \qquad n = k, k+1, \ldots, \quad 0 < \eta < 1,$$

which enforces $N \geq k$ and provides a parsimonious way to model random system size. Larger η puts more mass on larger n.

*The term "doubly generalized" here refers (i) to generalizing the failure rule from first to kth order statistic, i.e., k-out-of-n failure, and (ii) to introducing a shared gamma-frailty to capture unobserved heterogeneity and induce decreasing failure rates. This usage is unrelated to "doubly generalized" regression or GLM terminology.

(3) Component lifetimes. Conditional on $Z = z$ and $N = n$, we assume

$$X_1, \ldots, X_n \,\big|\, (Z = z, N = n) \overset{\text{iid}}{\sim} \text{Exp}(\theta z), \qquad \theta > 0,$$

so that

$$F_X(y \mid Z = z) = 1 - e^{-z\theta y}, \qquad f_X(y \mid Z = z) = \theta z\, e^{-z\theta y}, \qquad y > 0.$$

Here, θ is a baseline rate parameter. The frailty Z multiplies that rate, so higher Z corresponds to globally higher risk for all components in the same unit.

(4) System failure time. The observed lifetime is

$$Y = X_{(k)},$$

the kth order statistic of $\{X_1, \ldots, X_n\}$. In reliability terms, this is a k-out-of-n failure rule: the system is declared failed when k components have failed. Larger k represents more redundancy or tolerance to individual failures.

Together, the four parameters $(\theta, \eta, \lambda, k)$ admit direct interpretation: θ is the baseline stress/load; η controls dispersion in N (random system size or number of latent failure opportunities); λ controls frailty heterogeneity; and k encodes the failure threshold. Marginalizing the exponential components over the shared gamma-frailty yields a monotone survival function and, hence, a decreasing population hazard [14, 18].

2.2. Conditional and marginal density

We next derive the unconditional PDF and CDF (cumulative distribution function) of Y by integrating over Z and N.

Step 1: kth order statistic given $(Z = z, N = n)$.

For fixed $z > 0$ and $n \geq k$, the PDF of $Y = X_{(k)}$ (the kth order statistic of n i.i.d. exponentials with rate θz) is the standard order-statistic formula:

$$f_{Y|Z,N}(y \mid z, n) = k\binom{n}{k}[F_X(y \mid z)]^{k-1}[1 - F_X(y \mid z)]^{n-k} f_X(y \mid z), \qquad y > 0. \tag{2.1}$$

Substituting $F_X(y \mid z) = 1 - e^{-z\theta y}$ and $f_X(y \mid z) = \theta z e^{-z\theta y}$ gives

$$f_{Y|Z,N}(y \mid z, n) = k\binom{n}{k}\theta z \left(1 - e^{-z\theta y}\right)^{k-1} e^{-z(n-k+1)\theta y}.$$

We now expand the power term via the binomial theorem:

$$\left(1 - e^{-z\theta y}\right)^{k-1} = \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j e^{-z\theta y\, j}.$$

This yields

$$f_{Y|Z,N}(y \mid z, n) = k\binom{n}{k}\theta z \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j \exp\Big\{ - z\theta y[(n - k + 1) + j]\Big\}. \tag{2.2}$$

For convenience define

$$\tau_{n,j} \equiv n - k + j + 1, \qquad j = 0, \ldots, k-1,$$

so

$$f_{Y|Z,N}(y \mid z, n) = k\binom{n}{k}\theta z \sum_{j=0}^{k-1} \binom{k-1}{j}(-1)^j \exp\{-z\theta y\, \tau_{n,j}\}.$$

Step 2: Integrating out the frailty Z.

To obtain $f_{Y|N}(y \mid n)$ we integrate the Eq (2.2) with respect to the gamma-frailty $Z \sim \text{Gamma}(\lambda, \lambda)$:

$$f_{Y|N}(y \mid n) = \int_0^\infty f_{Y|Z,N}(y \mid z, n)\, g(z)\, dz, \qquad g(z) = \frac{\lambda^\lambda}{\Gamma(\lambda)} z^{\lambda-1} e^{-\lambda z}.$$

Plugging in the expression above, each summand in j contributes an integral of the form

$$\int_0^\infty z \cdot z^{\lambda-1} \exp\Big\{-z\Big[\lambda + \theta y\, \tau_{n,j}\Big]\Big\}\, dz = \int_0^\infty z^\lambda e^{-az} dz, \quad a = \lambda + \theta y\, \tau_{n,j}.$$

The gamma integral identity

$$\int_0^\infty z^\lambda e^{-az} dz = \frac{\Gamma(\lambda+1)}{a^{\lambda+1}} = \frac{\lambda\,\Gamma(\lambda)}{a^{\lambda+1}}$$

gives

$$f_{Y|N}(y \mid n) = \theta\, \lambda^{\lambda+1}\, k\binom{n}{k} \sum_{j=0}^{k-1} \binom{k-1}{j}(-1)^j \Big[\lambda + \theta y\, \tau_{n,j}\Big]^{-(\lambda+1)}.$$

Explicitly, since $\tau_{n,j} = n - k + j + 1$,

$$f_{Y|N}(y \mid n) = \theta\, \lambda^{\lambda+1}\, k\binom{n}{k} \sum_{j=0}^{k-1} \binom{k-1}{j}(-1)^j \Big[\,(n-k+j+1)\,\theta y + \lambda\Big]^{-(\lambda+1)}, \qquad y > 0. \tag{2.3}$$

Step 3: Mixing over the random system size N.

Finally, the unconditional PDF of Y is obtained by averaging the Eq (2.3) over N:

$$f_Y(y) = \sum_{n=k}^\infty \Pr(N = n)\, f_{Y|N}(y \mid n), \qquad \Pr(N = n) = (1-\eta)\,\eta^{n-k}.$$

Hence

$$f_Y(y) = \theta\, \lambda^{\lambda+1}\, k \sum_{n=k}^\infty \binom{n}{k}(1-\eta)\,\eta^{n-k} \sum_{j=0}^{k-1} \binom{k-1}{j}(-1)^j \Big[\,(n-k+j+1)\,\theta y + \lambda\Big]^{-(\lambda+1)}, \quad y > 0. \tag{2.4}$$

Equation (2.4) is the marginal PDF of the DGEGF distribution.

2.3. Distribution function

Integrating the Eq (2.3) with respect to y (or, equivalently, starting from the binomial expansion of the conditional CDF of the kth order statistic and then averaging over Z and N) yields a convenient closed form for the cumulative distribution function $F_Y(y) = \Pr(Y \le y)$:

$$F_Y(y) = k \sum_{n=k}^{\infty} (1-\eta)\,\eta^{n-k} \binom{n}{k} \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j \frac{1}{n-k+j+1} \times \left[1 - \left(\frac{\lambda}{\lambda + (n-k+j+1)\,\theta y}\right)^{\lambda}\right], \qquad y > 0. \tag{2.5}$$

An algebraically equivalent expression, obtained by expanding the conditional binomial term

$$\Pr(Y \le y \mid Z = z, N = n) = \sum_{r=k}^{n} \binom{n}{r} [F_X(y \mid z)]^r [1 - F_X(y \mid z)]^{n-r}$$

and then integrating term-by-term in z, is

$$F_Y(y) = \sum_{n=k}^{\infty} (1-\eta)\,\eta^{n-k} \sum_{r=k}^{n} \binom{n}{r} \sum_{m=0}^{r} \binom{r}{m} (-1)^m \left(\frac{\lambda}{\lambda + (m+n-r)\,\theta y}\right)^{\lambda}, \quad y > 0. \tag{2.6}$$

Equations (2.5) and (2.6) are mathematically identical; (2.5) keeps the (n, j) structure naturally tied to the kth order statistic, and is typically more convenient numerically.

Extensions to non-monotone hazards. The present DGEGF construction uses conditionally exponential lifetimes and gamma mixing; this combination yields a decreasing population hazard, so the model is tailored to early-failure / burn-in or heterogeneity-dominated regimes. If increasing or bathtub-shaped hazards are expected, the same hierarchy can be coupled with a richer baseline-for example a Weibull, gamma, or log-logistic component distribution-while retaining the k-truncated geometric distribution for N and the shared gamma-frailty. Such a "Weibull-DGEGF" or "log-logistic-DGEGF" extension would preserve the interpretability of $(\theta, \eta, \lambda, k)$ (baseline stress, opportunity dispersion, frailty, and tolerance-to-failure) but would allow non-monotone hazards, at the cost of heavier likelihood expressions; see [17, 20, 22] for related geometric-mixing constructions.

Figure 1 compares the proposed DGEGF density against standard lifetime models (exponential, EG, Weibull) for fixed $\theta = 1.5$ and $\eta = 0.5$. Two structural patterns emerge. First, for fixed frailty shape λ, increasing the redundancy level k shifts mass to the right and lowers the peak, since systems that fail only after the kth component breaks tend to survive longer. Second, across panels (increasing λ), the tail becomes lighter and the peak becomes sharper: large λ corresponds to weak frailty (low unobserved heterogeneity), so the model approaches exponential behavior, whereas small λ produces heavier tails and higher early risk. This separation of roles (k governing the threshold mechanism and λ governing tail weight and early-risk intensity) is not jointly available in simpler models. The EG effectively assumes $k = 1$ and no frailty, while the Weibull can shape the hazard but does not encode a k-out-of-n failure rule with common frailty. The DGEGF model can do both, under a strictly decreasing hazard.

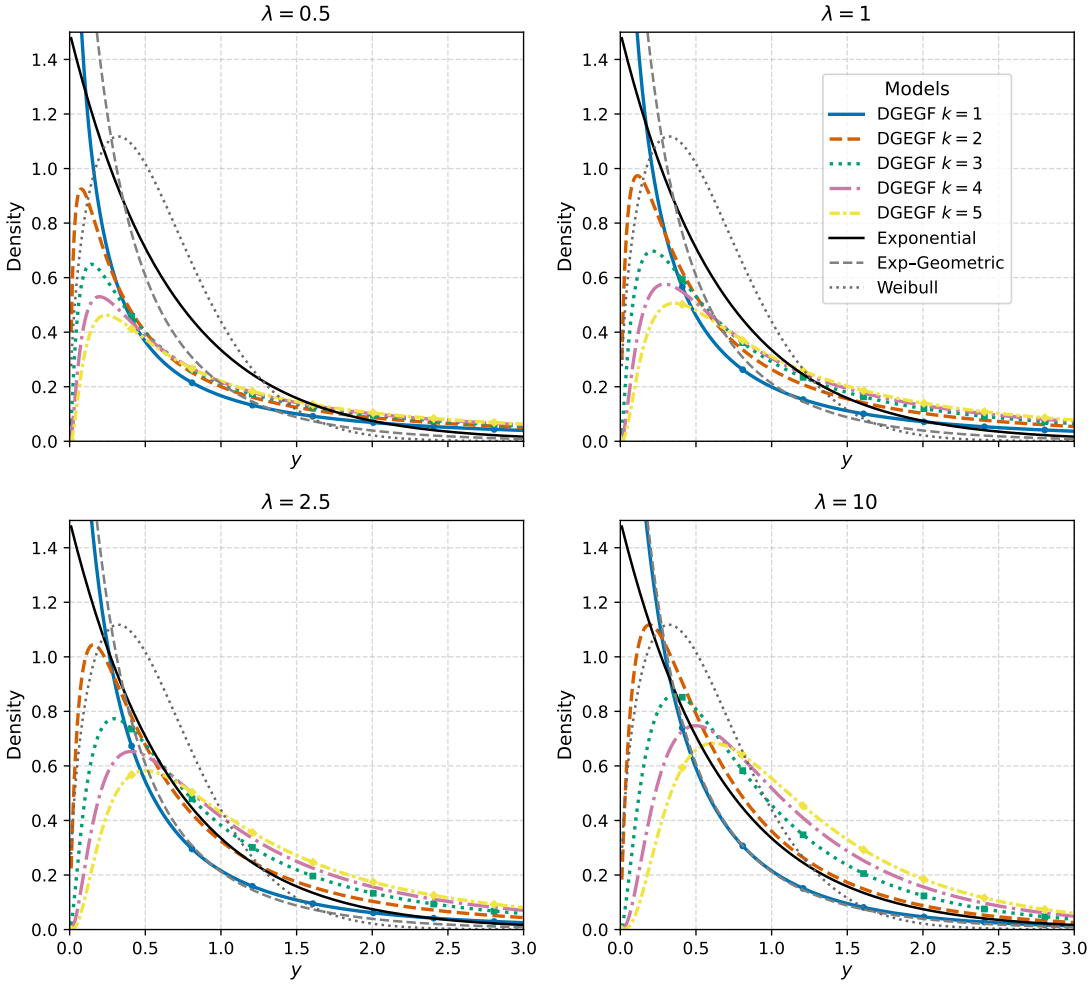


Figure 1. Marginal densities of the proposed DGEGF model (colored curves) for $k \in \{1, 2, 3, 4, 5\}$, shown at four frailty shapes $\lambda \in \{0.5, 1, 2.5, 10\}$ with $\theta = 1.5$ and $\eta = 0.5$. Gray/black curves give exponential, EG, and Weibull benchmarks.

3. Properties

This section studies analytical and numerical identifiability properties of the DGEGF model. We first establish formal identifiability of the parameter vector $(\theta, \eta, \lambda, k)$, and then examine finite–sample behavior using targeted likelihood diagnostics. The latter highlight regimes in which the frailty parameter λ is numerically fragile, even though it is identifiable at the population level.

3.1. Analytical identifiability

A basic requirement for meaningful parametric inference is identifiability: distinct parameter values must correspond to distinct probability distributions. For the DGEGF model, analytical identifiability ensures that population-level inference on $(\theta, \eta, \lambda, k)$ is well posed.

Proposition 3.1. *(Identifiability for fixed k) Let* $Y \sim \text{DGEGF}(\theta, \eta, \lambda, k)$ *with k fixed. If*

$$\text{DGEGF}(\theta_1, \eta_1, \lambda_1, k) \equiv \text{DGEGF}(\theta_2, \eta_2, \lambda_2, k),$$

then $(\theta_1, \eta_1, \lambda_1) = (\theta_2, \eta_2, \lambda_2)$.

Proof. The pdf decomposes into components that depend on θ (exponential rate), on η (geometric weights over N), and on λ (frailty term). Equality in distribution therefore forces equality of the corresponding transforms and series coefficients. Full derivations are given in Appendix A.1. □

Corollary 3.1. *(Identifiability with unknown k) If*

$$\mathrm{DGEGF}(\theta_1, \eta_1, \lambda_1, k_1) \equiv \mathrm{DGEGF}(\theta_2, \eta_2, \lambda_2, k_2),$$

then $k_1 = k_2$ *and* $(\theta_1, \eta_1, \lambda_1) = (\theta_2, \eta_2, \lambda_2)$.

Proof. The behavior of $F_Y(y)$ as $y \downarrow 0$ shows that k determines the leading power of y. Since this cannot be mimicked by varying (θ, η, λ), we must have $k_1 = k_2$, after which Proposition 3.1 applies. □

3.2. Numerical identifiability and aliasing

Analytical identifiability does not, by itself, guarantee stable estimation in finite samples. As emphasized by Hutson and Vexler [29] for multi–parameter lifetime models, different parameter combinations can generate densities and hazards that are nearly indistinguishable at realistic sample sizes. This phenomenon, often called numerical aliasing, manifests as flat regions or ridges in the likelihood surface, with corresponding instability or multimodality of the maximum likelihood estimator.

The DGEGF model is particularly susceptible to aliasing in λ, because several components influence the early-time behavior of the hazard. The baseline rate θ and frailty shape λ both act as scale parameters after mixing, while the geometric parameter η controls the effective number of competing risks. For small or moderate samples, these effects can partially compensate one another: increasing λ while suitably adjusting θ and η can leave the induced marginal density almost unchanged. In such regimes, the data contains little information about λ, and likelihood-based estimators of frailty heterogeneity can behave erratically even though $(\theta, \eta, \lambda, k)$ are identifiable at the population level.

3.3. Profile-likelihood diagnostics in a weak-frailty regime

To make these numerical issues concrete, we carried out focused likelihood diagnostics in a weak-frailty, first-hit scenario. Specifically, we considered parameter setting D:

$$(\theta, \eta, \lambda, k) = (1.0,\ 0.2,\ 0.5,\ 1),$$

generated independent samples of sizes $n = 80$ and $n = 300$, and, for each dataset, evaluated profile log-likelihoods in turn for λ, θ, and η. For a generic parameter $\varphi \in \{\theta, \eta, \lambda\}$, the profile log-likelihood

$$\ell_p(\varphi) = \max_{\phi} \ell(\varphi, \phi)$$

was obtained by maximizing the full log-likelihood over the remaining components ϕ. The resulting curves were centered at their respective maxima for ease of comparison.

Figure 2 displays the profile log-likelihood in λ. For $n = 80$, the curve is very flat over a wide interval of λ values: a broad range of frailty shapes produce nearly indistinguishable fits. In contrast,

the $n = 300$ profile is much more sharply peaked and penalizes extreme values of λ strongly. This behavior illustrates numerical aliasing of the frailty parameter in small samples: even though λ is formally identifiable, the likelihood surface contains an extended ridge along which λ can move with little change in fit quality. Only when n is sufficiently large does the ridge collapse into a well-defined maximum.

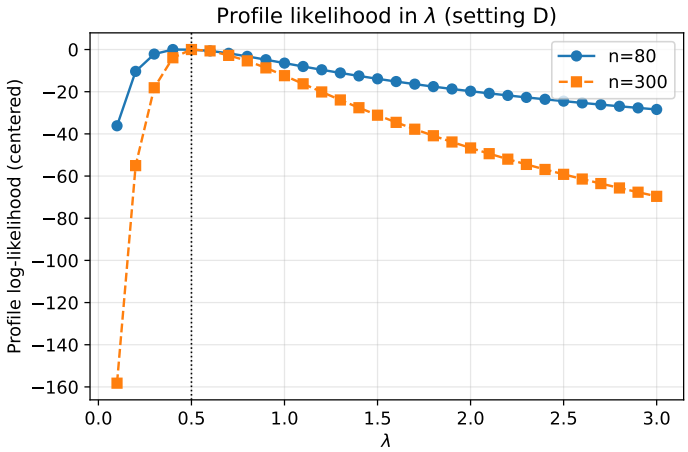


Figure 2. Profile log-likelihood in λ under setting D $(\theta, \eta, \lambda, k) = (1.0, 0.2, 0.5, 1)$ for sample sizes $n = 80$ and $n = 300$. Curves are centered at their maxima; the vertical line marks the generating value $\lambda = 0.5$.

Figures 3 and 4 show analogous profiles for θ and η. In both cases, the $n = 80$ profile already exhibits clear curvature, and the $n = 300$ curve is distinctly more concentrated. The profile in θ is well behaved for both sample sizes, with a relatively narrow plateau around the optimum and rapid decay away from it. The profile in η is somewhat flatter for $n = 80$, reflecting limited information about the geometric tail in very small samples, but nevertheless has a single, discernible mode and becomes sharply peaked at $n = 300$.

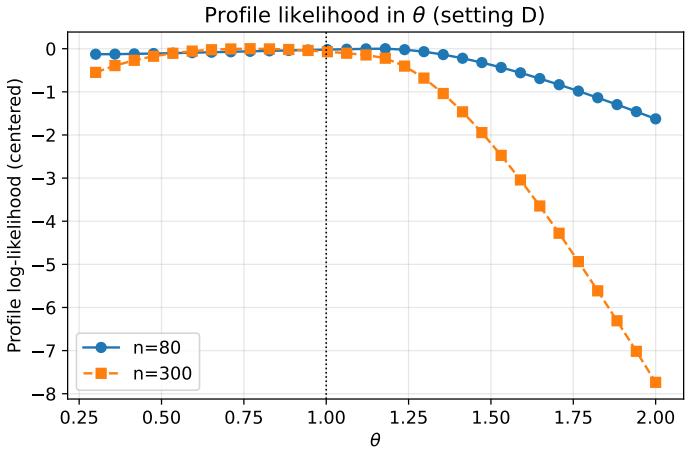


Figure 3. Profile log-likelihood in θ under setting D for $n = 80$ and $n = 300$. Curves are centered at their maxima; the vertical line marks the generating value $\theta = 1.0$.

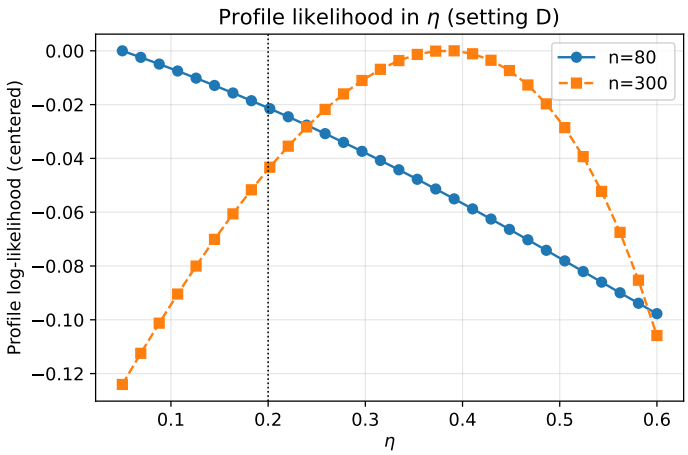


Figure 4. Profile log-likelihood in η under setting D for $n = 80$ and $n = 300$. Curves are centered at their maxima; the vertical line marks the generating value $\eta = 0.2$.

Taken together, these diagnostics support three conclusions:

(i) In weak-frailty regimes, the frailty shape λ can be numerically under-identified in small samples, with a wide, nearly flat profile log-likelihood.

(ii) The baseline rate θ and geometric parameter η are numerically much more stable: their profiles show clear curvature even for $n = 80$, and become sharply peaked as n increases.

(iii) Profile-likelihood plots are a practical tool for detecting aliasing in applied work: shallow or multimodal profiles signal that certain parameters, particularly λ, may require larger samples or regularization for reliable estimation.

These numerical investigations bridge the gap between formal identifiability and practical estimation, showing explicitly how aliasing can arise for λ and how it disappears as the amount of information in the data increases.

3.4. Moment generating function

The moment generating function (MGF) of Y, $M_Y(t) = \mathbb{E}[e^{tY}]$, can be expressed by substituting the marginal PDF $f_Y(y)$:

$$M_Y(t) = \theta\lambda^{\lambda+1}k(1-\eta)\sum_{n=k}^{\infty}\binom{n}{k}\eta^{n-k}\sum_{j=0}^{k-1}\binom{k-1}{j}(-1)^j\int_0^{\infty}\frac{e^{ty}}{[(n-k+j+1)\theta y+\lambda]^{\lambda+1}}\,\mathrm{d}y, \quad t\in\mathbb{R}.$$

By change of variables and recognizing the upper incomplete gamma function, the inner integral can be written in closed form. For $t < 0$ (i.e., for $s = -t > 0$), this yields the Laplace transform

$$\mathcal{L}_Y(s) = M_Y(-s) = \theta\lambda^{\lambda+1}k(1-\eta)\sum_{n=k}^{\infty}\binom{n}{k}\eta^{n-k}\sum_{j=0}^{k-1}\binom{k-1}{j}(-1)^j\frac{s^{\lambda}}{a_{n,j}^{\lambda+1}}\exp\left(\frac{s\lambda}{a_{n,j}}\right)\Gamma\left(-\lambda,\frac{s\lambda}{a_{n,j}}\right),$$

where $a_{n,j} = (n-k+j+1)\theta$ and $\Gamma(\cdot,\cdot)$ is the upper incomplete gamma function. This series converges for $s > 0$ and, therefore, $M_Y(t)$ exists for $t \le 0$. A full derivation, including all substitutions, is provided in Appendix B.

Differentiating $M_Y(t)$ at $t = 0^-$ gives the raw moments, so, in particular, finite moments can be recovered both from the Laplace transform and from direct integration (Section 3.5).

3.5. Moments

Let $Y = X_{(k)}$ denote the kth order statistic in the DGEGF hierarchy. The rth raw moment $\mathbb{E}[Y^r]$ is obtained by conditioning on $N = n$ and on the shared frailty $Z \sim \text{Gamma}(\lambda, \lambda)$, and then marginalizing:

$$\mathbb{E}[Y^r] = k\theta^{-r}\lambda^r\Gamma(r+1)\frac{\Gamma(\lambda - r)}{\Gamma(\lambda)}\sum_{n=k}^{\infty}(1-\eta)\eta^{n-k}\binom{n}{k}\sum_{j=0}^{k-1}\binom{k-1}{j}\frac{(-1)^j}{(n-k+j+1)^{r+1}}.$$

The gamma ratio term

$$\int_0^\infty z^{-r}\frac{\lambda^\lambda z^{\lambda-1}e^{-\lambda z}}{\Gamma(\lambda)}\,\mathrm{d}z = \lambda^r\frac{\Gamma(\lambda-r)}{\Gamma(\lambda)}$$

implies that $\mathbb{E}[Y^r]$ is finite, if and only if, $r < \lambda$. Thus, λ governs tail heaviness: small λ implies heavier right tails and fewer finite moments.

Two important special cases are:

The mean ($r = 1$), valid for $\lambda > 1$:

$$\mathbb{E}[Y] = \frac{k\lambda}{\theta(\lambda-1)}\sum_{n=k}^{\infty}(1-\eta)\eta^{n-k}\binom{n}{k}\sum_{j=0}^{k-1}\binom{k-1}{j}\frac{(-1)^j}{(n-k+j+1)^2}.$$

The second moment ($r = 2$), valid for $\lambda > 2$:

$$\mathbb{E}[Y^2] = \frac{k\lambda^2}{\theta^2(\lambda-1)(\lambda-2)}\sum_{n=k}^{\infty}(1-\eta)\eta^{n-k}\binom{n}{k}\sum_{j=0}^{k-1}\binom{k-1}{j}\frac{(-1)^j}{(n-k+j+1)^3}.$$

When $k = 1$ (first-failure time), this simplifies further because the inner sum collapses and $\binom{n}{1} = n$. Averaging over N then gives a closed-form mean in terms of $-\log(1-\eta)$. The full step-by-step derivation, including the order-statistic integral and the gamma integration over Z, is given in Appendix B.

3.6. Quantile function and median

Let $Q(p)$ denote the pth quantile of Y, so $F_Y(Q(p)) = p$ for $p \in (0, 1)$. From the DGEGF construction, the marginal CDF can be written as

$$F_Y(y) = k\sum_{n=k}^{\infty}(1-\eta)\eta^{n-k}\binom{n}{k}\sum_{j=0}^{k-1}\binom{k-1}{j}\frac{(-1)^j}{n-k+j+1}\left[1-\left(\frac{\lambda}{(n-k+j+1)\theta y+\lambda}\right)^\lambda\right].$$

Hence, $Q(p)$ is defined implicitly by $F_Y(Q(p)) = p$. In general there is no closed-form inversion because the righthand side is an infinite geometric sum with nested binomial terms. The median m_Y is obtained by setting $p = 0.5$.

For $k = 1$, the expression simplifies to a single geometric sum over n, which can be solved numerically by a one-dimensional root-finding routine (e.g., bisection or Newton-Raphson). For $k > 1$, the same numerical strategy applies. Practical computation of $Q(p)$ and m_Y therefore proceeds by evaluating $F_Y(y)$ on a grid and solving $F_Y(y) - p = 0$. Implementation details and the algebraic form specialized to $k = 1$ (first-failure case) are provided in Appendix B.

3.7. Hazard Function

The hazard function (or failure rate) of the DGEGF distribution is defined as

$$h_Y(y) = \frac{f_Y(y)}{S_Y(y)}, \quad y > 0,$$

where $f_Y(y)$ is the marginal PDF and $S_Y(y) = 1 - F_Y(y)$ is the survival function, given by

$$S_Y(y) = 1 - k \sum_{n=k}^{\infty} (1-\eta)\eta^{n-k} \binom{n}{k} \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{(-1)^j}{n-k+j+1} \left[1 - \left(\frac{\lambda}{(n-k+j+1)\theta y + \lambda} \right)^{\lambda} \right].$$

Thus, the hazard function can be written as

$$h_Y(y) = \frac{\theta \lambda^{\lambda+1} k \sum_{n=k}^{\infty} \binom{n}{k} (1-\eta)\eta^{n-k} \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j (n-k+j+1) \left[(n-k+j+1)\theta y + \lambda \right]^{-(\lambda+1)}}{1 - k \sum_{n=k}^{\infty} (1-\eta)\eta^{n-k} \binom{n}{k} \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{(-1)^j}{n-k+j+1} \left[1 - \left(\frac{\lambda}{(n-k+j+1)\theta y + \lambda} \right)^{\lambda} \right]}.$$

Special case: $k = 1$**.** When $k = 1$, we obtain

$$f_Y(y) = \theta \lambda^{\lambda+1} (1-\eta) \sum_{m=0}^{\infty} \eta^m (m+1) \left[(m+1)\theta y + \lambda \right]^{-(\lambda+1)},$$

$$S_Y(y) = (1-\eta) \sum_{m=0}^{\infty} \eta^m \left(\frac{\lambda}{(m+1)\theta y + \lambda} \right)^{\lambda},$$

and therefore

$$h_Y(y) = \frac{\theta \lambda^{\lambda+1} (1-\eta) \sum_{m=0}^{\infty} \eta^m (m+1) \left[(m+1)\theta y + \lambda \right]^{-(\lambda+1)}}{(1-\eta) \sum_{m=0}^{\infty} \eta^m \left(\frac{\lambda}{(m+1)\theta y + \lambda} \right)^{\lambda}}.$$

Qualitative shape and parameter effects. The DGEGF hazard $h_Y(y)$ is monotonically decreasing in y: it begins at a finite, strictly positive value at $y \to 0^+$, and then decays toward zero as y increases. Intuitively, two structural features drive this decrease. First, the multiplicative gamma-frailty induces a "selection" effect: highly fragile units (or high-risk subjects) tend to fail early, so the surviving subpopulation at larger y is, on average, more robust, which lowers the effective hazard over time [14, 18]. Second, system failure is defined through the kth order statistic rather than the first failure, meaning the system is not declared failed until multiple underlying components have already failed; this naturally suppresses late-time hazard for larger k.

The influence of each parameter on the shape and level of $h_Y(y)$ can be summarized as follows.

– Smaller λ implies higher frailty variance, i.e., stronger unobserved heterogeneity. In this regime, the initial hazard can be relatively high (reflecting a vulnerable subpopulation at early times), but it also drops sharply because the most fragile units fail first. Larger λ corresponds to weaker frailty and produces a more homogeneous population; the hazard curve is then flatter near the origin and closer to that of a standard exponential model.
– Increasing k lowers the hazard uniformly across y. Requiring the kth failure before declaring system failure mimics redundancy in k-out-of-n systems: higher k postpones catastrophic failure and therefore reduces the instantaneous failure risk at any given time.
– Larger η increases the expected number of latent failure opportunities N. When many potential failure modes are present at baseline, the initial hazard can be elevated, since there are more components that can fail early. Smaller η has the opposite effect.

Taken together, these effects imply that the DGEGF hazard captures "infant mortality" or burn-in behavior; elevated early risk followed by rapid stabilization as weak subcomponents fail or high-risk individuals fail early. This pattern is common in reliability settings (e.g., electronic components with early manufacturing flaws) and in biomedical or epidemiological applications with heterogeneous patient risk shortly after baseline.

Asymptotic behavior. As $y \to 0^+$, the hazard approaches a finite positive value, reflecting the initial instantaneous failure rate under shared frailty. As $y \to \infty$, the survival and density satisfy

$$S_Y(y) \sim C_S\, y^{-\lambda}, \qquad f_Y(y) \sim C_f\, y^{-(\lambda+1)},$$

for constants $C_S, C_f > 0$ depending on $(\theta, \eta, \lambda, k)$. Hence,

$$h_Y(y) \sim \frac{\lambda\, \theta^{\lambda}}{y}, \quad y \to \infty,$$

so $h_Y(y)$ decays to 0 at rate $1/y$. This slow, algebraic decay is consistent with heavy-tailed survival when λ is small, and reflects the persistence of long-lived, low-risk units at late times.

Numerical illustration. Figure 5 displays the DGEGF hazard function $h_Y(y)$ for frailty shapes $\lambda \in \{0.5, 1, 2.5, 10\}$ (one value per panel), with common $\theta = 1.5$, $\eta = 0.5$, and redundancy levels $k \in \{1, 2, 3, 4, 5\}$ shown as distinct curves. In every panel the hazard is strictly decreasing, which is exactly what the exponential baseline combined with gamma-frailty implies. Smaller λ (i.e., stronger unobserved heterogeneity) produces a high initial hazard followed by a sharp early drop; larger λ flattens the decay and makes the curve more exponential-looking. Within each panel, increasing k uniformly lowers the hazard at all times, because the system must tolerate more component failures before failing at the system level. Across panels, the dominant effect is thus the frailty shape λ (early-risk intensity), while within panels the spread of curves is driven by k (tolerance-to-failure).

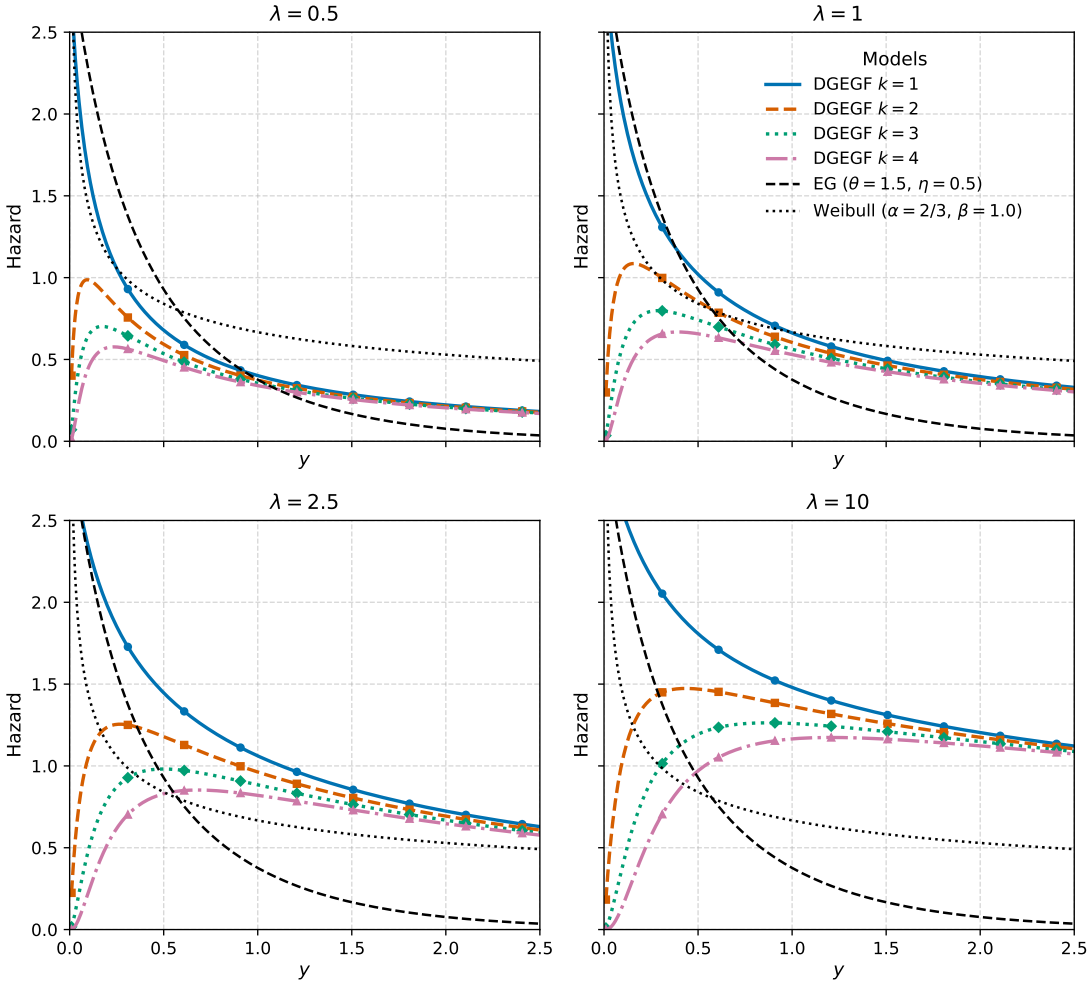


Figure 5. Hazard functions $h_Y(y)$ under the DGEGF model for $\theta = 1.5$ and $\eta = 0.5$. Each panel fixes the frailty shape $\lambda \in \{0.5, 1, 2.5, 10\}$ (top-left to bottom-right). Within every panel, colored curves correspond to redundancy levels $k \in \{1, 2, 3, 4\}$, using distinct color/linestyle/marker. Larger k means the system fails only after more component failures and, therefore, produces a uniformly *lower* hazard. Smaller λ (e.g., $\lambda = 0.5$) implies stronger unobserved heterogeneity and yields a high initial hazard followed by a rapid decline; as λ increases, the hazard becomes smoother and closer to exponential. For visual reference, the same θ and η are used in all panels so that differences can be attributed to (λ, k).

Figure 6 presents the same information from the complementary angle: each panel fixes the redundancy level $k \in \{1, 2, 3, 4\}$, and the curves correspond to $\lambda \in \{0.5, 1, 2.5, 10\}$. Holding k fixed isolates the role of the frailty: small λ yields elevated early hazard and heavier effective tail, whereas large λ produces a smoother, lower curve. Moving from the $k = 1$ panel to higher k panels, the entire hazard surface is pushed downward, reflecting delayed system failure under a "fail at the kth hit" rule. Independent y-axes are used to make these magnitude differences visible.

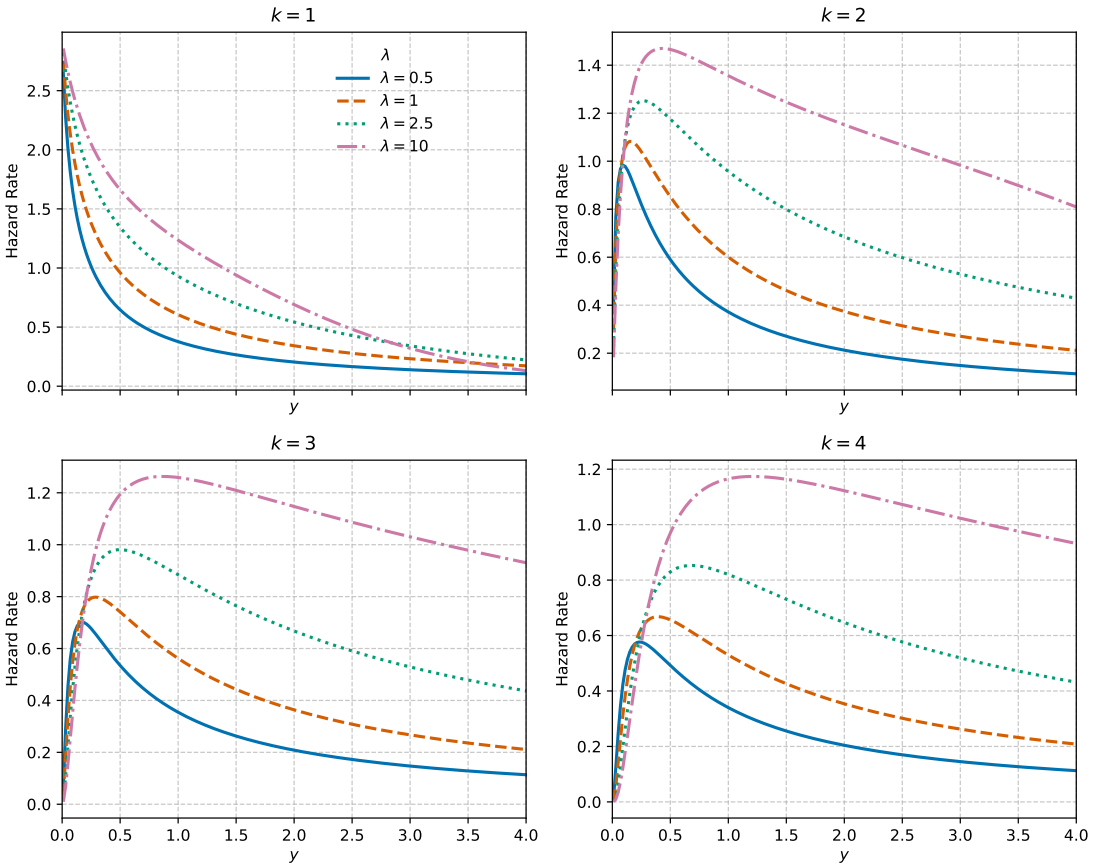


Figure 6. Hazard functions $h_Y(y)$ for $k \in \{1, 2, 3, 4\}$ (panels), with $\theta = 1.5$, $\eta = 0.5$, and $\lambda \in \{0.5, 1, 2.5, 10\}$. Larger k suppresses the hazard at all times by requiring multiple component failures before system-level failure. Smaller λ increases frailty heterogeneity, resulting in high initial hazard and heavier tails.

Practical implications and limitations. The decreasing hazard implied by the DGEGF makes it well suited for (i) systems with early-life or burn-in failures that stabilize over time, such as electronic or mechanical components subject to latent manufacturing defects, (ii) biomedical or epidemiological settings in which risk is initially high but falls as the most vulnerable patients fail early, and (iii) redundant k-out-of-n reliability structures in which system-level failure requires multiple component failures.

By construction, however, the DGEGF cannot reproduce increasing hazards or bathtub-shaped hazards (e.g., classical wear-out or aging effects). In such cases, alternative baselines (e.g., Weibull-based geometric-frailty generalizations) or non-monotone hazard models are more appropriate.

3.8. Limiting and special cases

The DGEGF model provides a flexible framework for the k-th order statistic $Y = X_{(k)}$ under shared gamma-frailty $Z \sim \Gamma(\lambda, \lambda)$ and random system size $N \sim \text{Geometric}_k(1 - \eta)$, where $P(N = n) = (1 - \eta)\eta^{n-k}$ for $n = k, k + 1, \ldots$. The interplay between frailty, redundancy, and failure order allows

the model to generalize several classical lifetime distributions. Below, we examine key special and limiting cases that highlight its versatility (Figure 7).

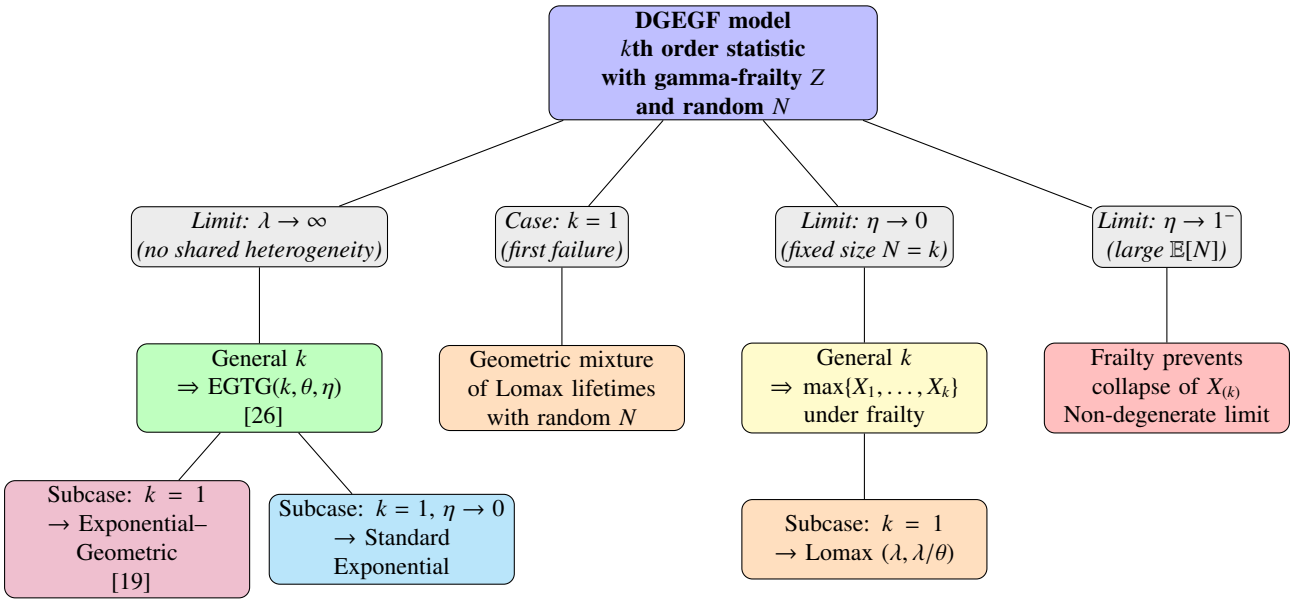


Figure 7. Limiting and special cases of the DGEGF model. The proposed model recovers the EG distribution [19], the Lomax (Pareto type II) distribution, and the standard exponential distribution as special or limiting cases, and admits k-out-of-n system interpretations.

Case 1: No shared heterogeneity ($\lambda \to \infty$). As $\lambda \to \infty$, the gamma-frailty $Z \sim \Gamma(\lambda, \lambda)$ degenerates at 1 (since $\mathbb{E}[Z] = 1$ and $\mathrm{Var}(Z) \to 0$), so conditional lifetimes become i.i.d. Exp(θ) with no unobserved heterogeneity. In this limit, the marginal distribution of $Y = X_{(k)}$ converges to the exponential–generalized truncated geometric (EGTG) model of [26]:

$$f_Y(y) = \theta k(1-\eta)\frac{e^{-\theta y}(1-e^{-\theta y})^{k-1}}{(1-\eta e^{-\theta y})^{k+1}}, \quad y > 0,$$

with CDF

$$F_Y(y) = \left(\frac{1-e^{-\theta y}}{1-\eta e^{-\theta y}}\right)^k, \quad y > 0.$$

This corresponds to the k-th order statistic among $N \sim \mathrm{Geometric}_k(1-\eta)$ i.i.d. Exp(θ) lifetimes, i.e., a k-out-of-n failure model without frailty.

Subcase: EG ($k = 1$, $\lambda \to \infty$). When $k = 1$ and $\lambda \to \infty$, $Y = X_{(1)}$ is the time to first failure in a system with a geometrically distributed number of exponential components, which is exactly the EG distribution of [19]:

$$f_Y(y) = \frac{\theta(1-\eta)e^{-\theta y}}{1-\eta e^{-\theta y}}, \quad y > 0,$$

with strictly decreasing hazard rate. This classical model describes early failure (time to first defect) in systems whose size is itself random.

Subcase: Standard exponential ($k = 1$, $\lambda \to \infty$, $\eta \to 0$). If we further let $\eta \to 0$, then $P(N = 1) \to 1$, so $N = 1$ a.s. and $Y = X_{(1)} = X_1$. In this joint limit ($k = 1$, $\lambda \to \infty$, $\eta \to 0$), the DGEGF model reduces to the standard exponential distribution

$$f_Y(y) = \theta e^{-\theta y}, \quad y > 0,$$

showing that the exponential distribution is nested as an extreme special case.

Case 2: Fixed system size ($\eta \to 0$). As $\eta \to 0$, we have $P(N = k) = 1 - \eta \to 1$, so the system size becomes deterministic: $N = k$. Then

$$Y = X_{(k)} = \max\{X_1, \ldots, X_k\},$$

the maximum of k conditionally exponential lifetimes under shared gamma-frailty. The resulting marginal PDF is

$$f_Y(y) = \theta\lambda^{\lambda+1}k \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{(-1)^j}{[(j+1)\theta y + \lambda]^{\lambda+1}}, \quad y > 0.$$

For $k = 1$, this simplifies to the Lomax (Pareto type II) distribution,

$$F_Y(y) = 1 - \left(1 + \frac{\theta}{\lambda} y\right)^{-\lambda}, \quad y > 0,$$

which is the well-known marginal survival model obtained by mixing an exponential hazard with gamma-frailty [14, 18].

Case 3: First failure ($k = 1$). When $k = 1$, $Y = X_{(1)}$ is the time to first failure among $N \sim \text{Geometric}_1(1 - \eta)$ components, with $P(N = n) = (1 - \eta)\eta^{n-1}$. The marginal PDF is

$$f_Y(y) = \theta\lambda^{\lambda+1}(1 - \eta) \sum_{n=1}^{\infty} \frac{n\eta^{n-1}}{(n\theta y + \lambda)^{\lambda+1}}, \quad y > 0.$$

Conditionally on $N = n$, $Y \mid N = n$ follows a Lomax distribution with scale $\lambda/(n\theta)$ and shape λ. Thus, Y is a geometric mixture of Lomax components. This case is useful for modeling early-life failures in settings where both the number of active failure modes and the underlying risk environment are uncertain.

Case 4: Highly redundant systems ($\eta \to 1^-$). As $\eta \to 1^-$, the expected system size

$$\mathbb{E}[N] = k + \frac{\eta}{1 - \eta} \to \infty.$$

For fixed frailty $Z = z$, the k-th order statistic of a very large number of exponential components tends to zero in probability: with so many potential failure modes, some fail almost immediately. However,

after marginalizing over $Z \sim \Gamma(\lambda, \lambda)$, extremely small-$Z$ realizations (i.e., "robust" systems with low effective hazard) counterbalance that degeneracy. The unconditional density remains non-degenerate:

$$f_Y(y) = \theta\lambda^{\lambda+1}k \sum_{n=k}^{\infty} \binom{n}{k}(1-\eta)\eta^{n-k} \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{(-1)^j}{[(n-k+j+1)\theta y + \lambda]^{\lambda+1}}, \quad y > 0,$$

illustrating how frailty regularizes the limit even when the system size becomes very large.

The DGEGF model therefore unifies several classical survival distributions through interpretable limits. It extends the exponential–geometric and Lomax families while accommodating (i) random system size, (ii) thresholded failure at the kth order statistic, and (iii) shared gamma-frailty. This joint structure allows it to represent heterogeneous k-out-of-n systems with random effective size, something that is not captured by standard exponential, EG, or single-cluster frailty models [14, 18, 19, 26].

4. Parameter estimation

Estimating the parameters of the DGEGF model presents nontrivial computational and statistical challenges due to its hierarchical structure and the presence of infinite series in the marginal likelihood. In this section, we derive the maximum likelihood estimators (MLEs), present the expectation-maximization (EM) algorithm as a natural estimation framework, and introduce a practical, robust implementation strategy that ensures stable inference in finite samples.

4.1. MLE

Let $\boldsymbol{\varphi} = (\theta, \eta, \lambda)$ denote the vector of continuous parameters, with $k \in \mathbb{N}^+$ treated as known or estimated separately (e.g., via profile likelihood or information criteria). Given an i.i.d. sample $\{y_i\}_{i=1}^m$, the likelihood function is:

$$L(\boldsymbol{\varphi}) = \prod_{i=1}^{m} f_Y(y_i; \theta, \eta, \lambda, k),$$

where $f_Y(y_i)$ is the marginal density derived in Section 2:

$$f_Y(y_i) = \theta\lambda^{\lambda+1}k \sum_{n=k}^{\infty}(1-\eta)\eta^{n-k} \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{(-1)^j}{\left[(n-k+j+1)\theta y_i + \lambda\right]^{\lambda+1}}.$$

The log-likelihood function is:

$$\ell(\boldsymbol{\varphi}) = \sum_{i=1}^{m} \log f_Y(y_i) = \sum_{i=1}^{m} \left[\log\theta + (\lambda+1)\log\lambda + \log k + \log(1-\eta) - k\log\eta + \log S_i(\boldsymbol{\varphi})\right],$$

where

$$S_i(\boldsymbol{\varphi}) = \sum_{n=k}^{\infty} \eta^n \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{(-1)^j}{\left[(n-k+j+1)\theta y_i + \lambda\right]^{\lambda+1}}.$$

The score functions are as follows.

For θ:

$$\frac{\partial\ell}{\partial\theta} = \sum_{i=1}^{m} \left[\frac{1}{\theta} + \frac{1}{S_i(\boldsymbol{\varphi})}\frac{\partial S_i(\boldsymbol{\varphi})}{\partial\theta}\right],$$

with

$$\frac{\partial S_i(\boldsymbol{\varphi})}{\partial \theta} = -(\lambda+1)\sum_{n=k}^{\infty}\eta^n\sum_{j=0}^{k-1}\binom{k-1}{j}(-1)^j\frac{(n-k+j+1)y_i}{[(n-k+j+1)\theta y_i+\lambda]^{\lambda+2}}.$$

For η:

$$\frac{\partial \ell}{\partial \eta} = \sum_{i=1}^{m}\left[-\frac{1}{1-\eta}-\frac{k}{\eta}+\frac{1}{S_i(\boldsymbol{\varphi})}\frac{\partial S_i(\boldsymbol{\varphi})}{\partial \eta}\right],$$

with

$$\frac{\partial S_i(\boldsymbol{\varphi})}{\partial \eta} = \sum_{n=k}^{\infty}n\eta^{n-1}\sum_{j=0}^{k-1}\binom{k-1}{j}\frac{(-1)^j}{[(n-k+j+1)\theta y_i+\lambda]^{\lambda+1}}.$$

For λ:

$$\frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^{m}\left[\frac{\lambda+1}{\lambda}+\log\lambda+\frac{1}{S_i(\boldsymbol{\varphi})}\frac{\partial S_i(\boldsymbol{\varphi})}{\partial \lambda}\right],$$

with

$$\frac{\partial S_i(\boldsymbol{\varphi})}{\partial \lambda} = \sum_{n=k}^{\infty}\eta^n\sum_{j=0}^{k-1}\binom{k-1}{j}(-1)^j\left[\frac{-(\lambda+1)\log((n-k+j+1)\theta y_i+\lambda)-1}{[(n-k+j+1)\theta y_i+\lambda]^{\lambda+1}}\right].$$

Due to the presence of infinite series and nonlinear, parameter-dependent denominators, direct numerical maximization of $\ell(\boldsymbol{\varphi})$ can be challenging: standard optimization routines (e.g., Broyden-Fletcher-Goldfarb-Shanno (BFGS); Nelder-Mead) are sensitive to starting values and may converge slowly. This initially motivated consideration of the EM algorithm [30], which, in principle, could exploit the latent-variable structure of the DGEGF model. However, in practice we found the EM formulation to be less stable: convergence was slower, estimates were more variable, and boundary divergence was common. Consequently, we adopted a direct likelihood-based optimization strategy, which proved to be more robust and reproducible across simulation settings.

4.2. EM algorithm

The EM algorithm treats the frailty Z_i and the random system size N_i as latent variables. The observed data are $\{y_i\}_{i=1}^m$; the complete data are $\{(y_i, z_i, n_i)\}_{i=1}^m$. The complete-data likelihood for the i-th observation is:

$$\begin{aligned}L_c(\boldsymbol{\varphi}; y_i, z_i, n_i) &= P(N_i = n_i \mid \eta, k)\, f_{X_{(k)}|Z}(y_i \mid n_i, \theta, z_i)\, g(z_i \mid \lambda)\\ &= (1-\eta)\eta^{n_i-k}\cdot k\binom{n_i}{k}\theta z_i(1-e^{-z_i\theta y_i})^{k-1}e^{-z_i(n_i-k+1)\theta y_i}\cdot\frac{\lambda^\lambda}{\Gamma(\lambda)}z_i^{\lambda-1}e^{-\lambda z_i}.\end{aligned}$$

The complete-data log-likelihood (up to additive constants) is:

$$\begin{aligned}\log L_c(\boldsymbol{\varphi}) = \sum_{i=1}^{m}\Big[&\log(1-\eta)+(n_i-k)\log\eta+\log\theta+\log z_i-z_i(n_i-k+1)\theta y_i\\ &+(k-1)\log(1-e^{-z_i\theta y_i})+\lambda\log\lambda+(\lambda-1)\log z_i-\lambda z_i\Big].\end{aligned}$$

E-step. Given current estimates $\boldsymbol{\varphi}^{(r)} = (\theta^{(r)}, \eta^{(r)}, \lambda^{(r)})$, compute the expected complete-data log-likelihood:

$$Q(\boldsymbol{\varphi} \mid \boldsymbol{\varphi}^{(r)}) = \mathbb{E}_{Z,N|Y,\boldsymbol{\varphi}^{(r)}}[\log L_c(\boldsymbol{\varphi}) \mid \{y_i\}, \boldsymbol{\varphi}^{(r)}].$$

This requires The posterior probability $\tau_{i,n}^{(r)} = P(N_i = n \mid y_i, \boldsymbol{\varphi}^{(r)})$ and the conditional expectations:

$$u_{i,n}^{(r)} = \mathbb{E}[Z_i \mid y_i, N_i = n, \boldsymbol{\varphi}^{(r)}], \quad v_{i,n}^{(r)} = \mathbb{E}[\log Z_i \mid y_i, N_i = n, \boldsymbol{\varphi}^{(r)}],$$

$$w_{i,n}^{(r)}(\theta) = \mathbb{E}[\log(1 - e^{-Z_i \theta y_i}) \mid y_i, N_i = n, \boldsymbol{\varphi}^{(r)}].$$

These expectations involve integrals over the gamma-frailty distribution and are evaluated numerically using Gauss-Laguerre quadrature.

M-step. Maximize $Q(\boldsymbol{\varphi} \mid \boldsymbol{\varphi}^{(r)})$ to obtain $\boldsymbol{\varphi}^{(r+1)}$.

For η, the update is explicit:

$$\eta^{(r+1)} = \frac{\sum_{i=1}^{m} \mathbb{E}[N_i - k \mid y_i, \boldsymbol{\varphi}^{(r)}]}{m + \sum_{i=1}^{m} \mathbb{E}[N_i - k \mid y_i, \boldsymbol{\varphi}^{(r)}]}.$$

For θ and λ, numerical optimization is required. The derivative with respect to λ involves the digamma function $\psi(\lambda)$:

$$\frac{\partial Q}{\partial \lambda} = m(1 + \log \lambda - \psi(\lambda)) + \sum_{i=1}^{m} \sum_{n=k}^{\infty} \tau_{i,n}^{(r)} \left(v_{i,n}^{(r)} - u_{i,n}^{(r)} \right).$$

These equations are solved using Newton-Raphson or quasi-Newton methods.

4.3. Practical implementation and robust estimation

The EM algorithm, presented in the last subsection, proved unstable in our experiments; convergence was slow, and estimates for λ often diverged or displayed excessive variance; a phenomenon well documented in frailty model estimation [14, 18]. To ensure reliable inference, we therefore adopted a hybrid direct-likelihood approach [31, 32], which was used in all simulations and empirical analyzes reported in this paper.

Our estimation procedure is as follows:

(1) The infinite sum over n was truncated at $N_{\max} = 120$, ensuring that the omitted tail probability $P(N > N_{\max})$ was negligible ($< 10^{-6}$ for all $\eta \leq 0.99$).
(2) The integral over frailty z was approximated using Gauss-Laguerre quadrature with 8 nodes, which provides high accuracy for gamma-weighted integrals.
(3) Optimization was carried out via the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS-B) algorithm (which accommodates box constraints) with 10 random initializations drawn from plausible ranges:

$$\theta^{(0)} \sim \text{Uniform}(0.1, 5.0), \quad \eta^{(0)} \sim \text{Uniform}(0.05, 0.95), \quad \lambda^{(0)} \sim \text{Uniform}(0.5, 10.0).$$

(4) Among all converged runs, the solution with the highest log-likelihood was retained.

This strategy sacrifices some theoretical purity for substantial gains in stability and reproducibility. It consistently achieved full convergence in our simulation study and produced interpretable estimates in all settings.

Remark. Although alternative estimators (for example, Bayesian inference with priors on (θ, η, λ), CDF-based minimum-distance / least-squares fitting, or percentile-matching approaches) can in principle be constructed for the DGEGF model, these methods require additional tuning choices and do not yet have established large-sample guarantees for this hierarchical specification. In contrast, the direct maximum likelihood estimator is asymptotically well behaved under standard regularity arguments [33, 34], and can be computed robustly using the truncated-likelihood strategy described above. Moreover, our experiments showed that EM updates for the frailty parameter λ can become unstable in moderate samples, which is consistent with known numerical difficulties in shared frailty models [14, 18]. For these reasons, all inference reported in this paper is based on direct maximum likelihood. A systematic comparison with Bayesian and minimum-distance estimators, especially for stabilizing λ in very small samples, is an important direction for future work.

5. Simulation study

This section investigates the finite-sample behavior of the maximum likelihood estimators of the DGEGF model and links their numerical performance to the identifiability and aliasing issues discussed in Section 3. We focus on three questions:

- How the bias and root mean squared error (RMSE) of $(\hat{\theta}, \hat{\eta}, \hat{\lambda})$ evolve as the sample size n increases;
- How often the likelihood optimization converges to a well-behaved interior maximum;
- How the frailty parameter λ behaves in regimes where the likelihood is nearly flat in the frailty direction, as indicated by the diagnostics in Section 3.3.

The design is constructed to stress both formal identifiability and practical numerical stability. We vary the redundancy level k, the geometric opportunity parameter η, and the frailty parameter λ across four settings and examine sample sizes ranging from very small to large. The main text reports results for the representative sample sizes $n \in \{20, 80, 200, 500\}$, while the full grid and additional diagnostics are documented in Appendix C.

5.1. Experimental design

To describe how estimator behavior changes with the amount of information, we consider the grid of sample sizes

$$n \in \{20, 40, 60, 80, 120, 160, 200, 300, 500\}.$$

This grid allows us to follow the evolution of bias, variance, and numerical stability as n increases, and to detect possible nonlinear or plateau-like regions in the RMSE of the estimators. In particular, it enables us to identify sample-size ranges in which the frailty parameter λ is weakly identified and the ranges in which all components (θ, η, λ) become stably estimable.

We generated data from four qualitatively different DGEGF parameter configurations,

$$(\theta, \eta, \lambda, k) \in \begin{cases} \text{Setting A:} & (1.2, 0.4, 1.5, 2), \\ \text{Setting B:} & (0.8, 0.6, 0.8, 1), \\ \text{Setting C:} & (2.0, 0.3, 3.0, 3), \\ \text{Setting D:} & (1.0, 0.2, 0.5, 1), \end{cases}$$

chosen to combine moderate and strong frailty, first-hit and multi-hit failure rules, and lighter and heavier tails.

For each setting and each sample size n in the grid, we generated R_n independent datasets from the hierarchical DGEGF construction in Section 2. The replication numbers decrease with n (e.g., $R_{20} = 24$, $R_{80} = 14$, $R_{200} = 8$, $R_{500} = 4$) to balance Monte Carlo error against computational cost. For every dataset we computed the maximum likelihood estimator $(\hat{\theta}, \hat{\eta}, \hat{\lambda})$ by direct maximisation of the marginal likelihood as described in Section 4.

For each combination of setting and n we summarized performance by:

- empirical bias and RMSE of $\hat{\theta}$, $\hat{\eta}$ and $\hat{\lambda}$;
- empirical convergence rate of the optimizer (proportion of runs converging to an interior maximum);
- the empirical distribution of the estimators across the sample-size grid;
- RMSE trajectories RMSE(n) across the entire grid, highlighting any non-monotone or plateau-like behavior.

To keep the presentation concise, we report complete numerical summaries in the main text for the four representative sample sizes

$$n \in \{20, 80, 200, 500\},$$

which illustrate very small, moderate, and large samples. The full results for $n \in \{20, 40, 60, 80, 120, 160, 200, 300, 500\}$, including RMSE curves, boxplots of the sampling distributions, and additional stability diagnostics, are provided in Appendix C.

5.2. Numerical results

Table 1 reports the average bias, RMSE, and convergence rates for $\hat{\theta}$, $\hat{\eta}$, and $\hat{\lambda}$ at $n = 20, 80, 200, 500$ across the four parameter settings A–D. These four sample sizes summarize the transition from very small to large samples.

At the smallest size $n = 20$, all three estimators exhibit appreciable variability, and the frailty parameter λ is particularly unstable in Settings A–C. For example, the bias of $\hat{\lambda}$ is around 5.2 in Setting A, 2.4 in Setting B, and 11.8 in Setting C, with corresponding RMSE values between roughly 7 and 18. By contrast, Setting D already yields a very accurate frailty estimate at $n = 20$ (bias $\approx$ 0.09, RMSE $\approx$ 0.26), so that numerical aliasing is concentrated in the designs with stronger shared heterogeneity or higher redundancy. Even in this small-sample regime, however, the estimators $\hat{\theta}$ and $\hat{\eta}$ display only moderate bias (typically below 0.4 in absolute value) and reasonably bounded RMSE, indicating that the baseline and geometric components already carry non-negligible information.

As the sample size increases to $n = 80$ and $n = 200$, the baseline rate θ and geometric parameter η behave well in all settings. At $n = 80$, the biases for $\hat{\theta}$ and $\hat{\eta}$ are modest and centered near the truth, with RMSEs between about 0.3 and 0.6. By $n = 200$, both parameters show a clear reduction in RMSE and smaller biases across all designs. This is consistent with the identifiability results in Section 3.1.

Table 1. Bias, RMSE, and convergence rates for $\hat{\theta}$, $\hat{\eta}$, and $\hat{\lambda}$ at $n = 20, 80, 200, 500$ across settings A–D. "Conv." denotes the number of successful convergences out of the total number of simulated datasets R_n.

Setting	n	Conv.	Bias(θ)	Bias(η)	Bias(λ)	RMSE(θ)	RMSE(η)	RMSE(λ)
A	20	24/24	0.07	-0.04	5.15	0.75	0.34	10.85
A	80	14/14	0.07	-0.03	0.12	0.50	0.30	0.51
A	200	8/8	0.18	-0.07	0.06	0.48	0.25	0.40
A	500	4/4	0.14	-0.12	0.11	0.29	0.25	0.21
B	20	24/24	0.34	-0.19	2.36	0.99	0.41	6.97
B	80	14/14	0.33	-0.25	0.02	0.63	0.41	0.12
B	200	8/8	0.32	-0.20	0.08	0.72	0.40	0.24
B	500	4/4	0.20	-0.20	0.02	0.41	0.33	0.06
C	20	24/24	-0.18	0.16	11.84	0.75	0.30	17.51
C	80	14/14	-0.01	0.01	1.47	0.55	0.26	3.31
C	200	8/8	0.23	-0.14	0.14	0.55	0.29	1.10
C	500	4/4	-0.30	0.14	0.97	0.52	0.23	1.36
D	20	24/24	0.29	0.02	0.09	1.05	0.28	0.26
D	80	14/14	-0.14	0.07	0.03	0.28	0.24	0.07
D	200	8/8	-0.25	0.18	0.05	0.35	0.26	0.07
D	500	4/4	0.09	-0.13	0.00	0.16	0.17	0.02

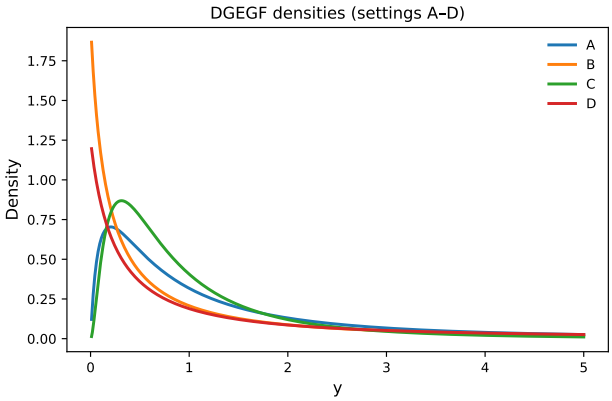
The frailty parameter λ improves more slowly. In Settings A and B, the RMSE of $\hat{\lambda}$ drops rapidly as n grows, and from $n = 80$ upward the estimator is essentially unbiased with RMSE well below 1. Setting D is favorable for λ at all sample sizes, with very small bias and RMSE even at $n = 20$. Setting C, which combines strong frailty with $k = 3$, remains challenging: although the RMSE of $\hat{\lambda}$ decreases markedly between $n = 20$ and $n = 200$ (from about 17.5 to about 1.1), a noticeable positive bias and relatively large RMSE (about 1.36) persist even at $n = 500$. This behaviour is consistent with the strong heterogeneity and higher-order redundancy in Setting C, which create flatter likelihood regions in the frailty direction.

At the largest sample size $n = 500$, the estimators of θ and η are tightly concentrated around their true values in all settings, with small biases and RMSEs. The frailty estimator $\hat{\lambda}$ also stabilizes substantially compared with smaller n: Settings A, B and D show biases close to zero and RMSE between about 0.02 and 0.21, while Setting C displays a moderate positive bias (about 0.97) and RMSE around 1.36.

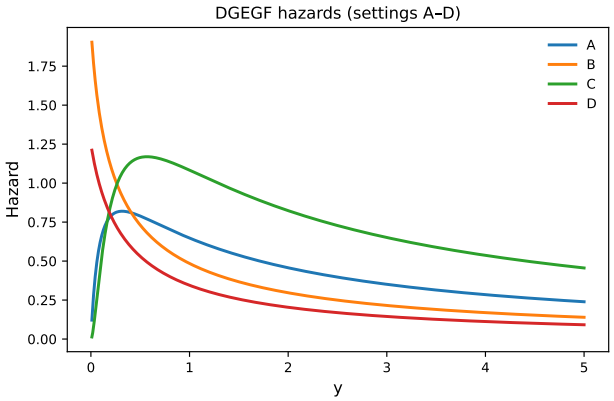
Over all Monte Carlo replications at $n \in \{20, 80, 200, 500\}$, the L-BFGS-B optimizer converged to a feasible interior solution in every run ("Conv." equals R_n in all rows of Table 1). Even in cases with flat likelihood regions in the frailty direction, the algorithm successfully located an interior maximum, although the resulting $\hat{\lambda}$ may be numerically unstable when the likelihood ridge is wide.

Figure 8 displays the marginal densities and hazard functions of the DGEGF model for the four parameter settings. Panel 8a illustrates that the density can change shape across settings, reflecting different combinations of rate, geometric, frailty, and redundancy parameters. Panel 8b confirms that the corresponding hazards are monotonically decreasing in every case, as implied by the exponential baseline combined with gamma mixing. Parameter variation shifts location and spread without altering

the strictly DFR nature of the model.

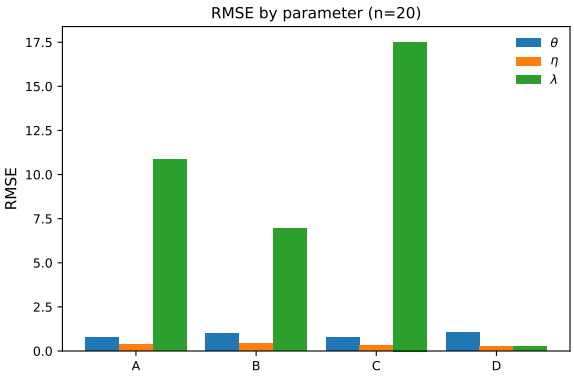


(a) Marginal densities under parameter settings A–D.

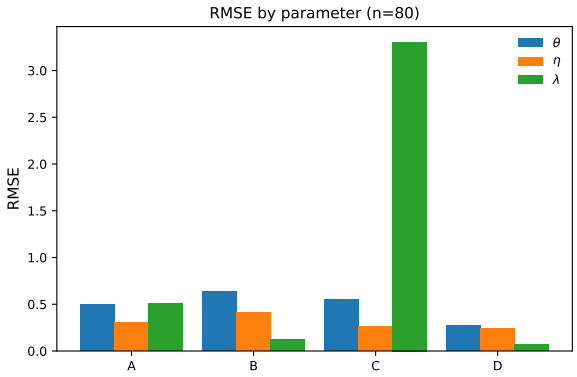


(b) Hazard rate functions under parameter settings A–D.

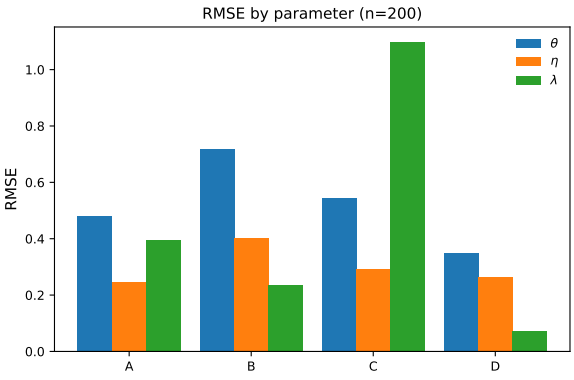
Figure 8. Densities and hazard functions of the DGEGF model across settings A–D.



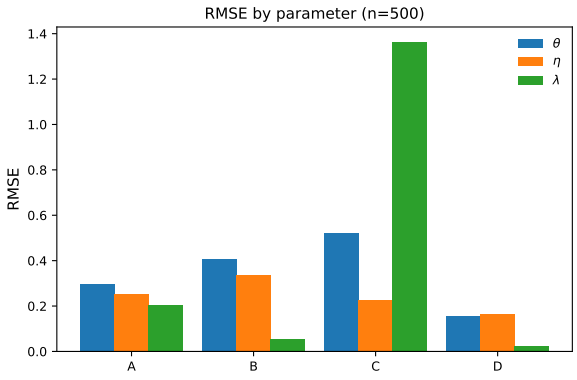
(a) $n = 20$



(b) $n = 80$



(c) $n = 200$



(d) $n = 500$

Figure 9. RMSE barplots comparing parameter estimation accuracy across settings A–D at four representative sample sizes ($n = 20, 80, 200, 500$).

To illustrate the dependence of estimation accuracy on sample size more directly, Figure 9 presents RMSE barplots for the three parameters (θ, η, λ) at the four representative values $n = 20, 80, 200, 500$. For $n = 20$, RMSEs are large for all parameters and particularly pronounced for $\hat{\lambda}$ in Settings A–C, reflecting the instability of frailty estimation under moderate or strong shared heterogeneity. At $n = 80$ and $n = 200$, the RMSE of $\hat{\theta}$ and $\hat{\eta}$ declines steadily across all settings, while the RMSE of $\hat{\lambda}$ decreases but remains comparatively large in Setting C. At $n = 500$, the RMSEs of all three parameters are small in Settings A, B and D, and although $\hat{\lambda}$ is still the least stable parameter in Setting C, its variability is much reduced relative to smaller n.

Additional graphical summaries, including RMSE trajectories over the full grid of n, boxplots of the sampling distributions, and histogram panels, are reported in Appendix C. These confirm that $(\hat{\theta}, \hat{\eta})$ stabilize quickly as n increases, whereas $\hat{\lambda}$ is numerically more fragile in the strong-frailty, multi-hit design (Setting C) and converges more slowly.

5.3. Discussion

The simulation study leads to several conclusions that complement the identifiability and aliasing analysis in Section 3.

First, the baseline rate θ and geometric parameter η are reliably estimable in moderate and large samples. Across all parameter settings, their biases and RMSE decrease steadily as n grows, and the likelihood surface is well behaved in these directions. For $n \geq 80$, the sampling distributions of $\hat{\theta}$ and $\hat{\eta}$ are already reasonably concentrated, and by $n = 200$ and $n = 500$ they are tightly clustered around the true values in all four designs.

Second, the frailty parameter λ is numerically the most delicate component. In small samples, especially under moderate or strong frailty (Settings A–C), the likelihood is relatively flat in the λ direction, which manifests as large variability and occasional extreme values of $\hat{\lambda}$. This behavior matches the near-aliasing phenomenon discussed in Section 3.3: multi-start MLE runs and profile likelihoods show that many different values of λ can yield almost identical fits when n is small. As the sample size increases along the grid, the ridge in the likelihood surface sharpens, the multi-start clouds contract, and the profile likelihoods become more curved, indicating improved numerical identifiability of λ, though in Setting C a moderate bias persists even at $n = 500$.

Third, the grid of sample sizes provides a genuinely gradual view of estimator evolution. The RMSE trajectories reported in Appendix C show that $(\hat{\theta}, \hat{\eta})$ improve monotonically with n, while $\hat{\lambda}$ often exhibits a plateau region (most visibly in Setting C between $n \approx 40$ and $n \approx 160$) before eventually stabilizing at larger sample sizes.

Finally, numerical experiments indicate that EM-type schemes (Section 4.3) are prone to slow or unstable convergence in the presence of near-aliasing, whereas direct maximization of the marginal likelihood with multiple starting values is more robust. All results in this section are therefore based on the direct likelihood approach.

In summary, the simulation study suggests that in applications with moderate or large sample sizes the DGEGF model can be fitted reliably by maximum likelihood, with θ and η estimated accurately and λ estimated with acceptable precision except in the most challenging strong-frailty, multi-hit regimes. In small samples or in designs where the frailty component is weakly identified, it is advisable to combine likelihood-based estimation with the numerical diagnostics of Section 3.2 and, if necessary, to regularize λ to avoid excursions along flat likelihood ridges.

6. Applications

We now study the empirical behavior of the proposed DGEGF model on four benchmark datasets from reliability engineering, hydrology, and biomedical survival analysis, including real datasets alongside simulation studies is standard in recent reliability work, where new lifetime models are expected to demonstrate practical utility on observed data; see, e.g., ElSaeed and Almetwally [35] for recent examples. For each dataset we fit the DGEGF model for redundancy levels $k \in \{1, 2, 3, 4, 5\}$ and estimate (θ, η, λ) by maximum likelihood as described in Section 4. For a fixed k, numerical maximization is carried out under box constraints ($\theta > 0$, $0 < \eta < 1$, $\lambda > 0$) using L–BFGS–B. The geometric tail in the latent opportunity count N is truncated adaptively so that the omitted mass is below a fixed tolerance. We then select a "best" DGEGF specification by the Bayesian information criterion

$$\mathrm{BIC} = p \log n - 2 \log \hat{L}, \qquad p = 3,$$

i.e., we choose the k attaining the smallest Bayesian information criterion (BIC). It should be noted that our goal is not to construct a model that uniformly dominates classical parametric families in terms of information criteria AIC/BIC across all possible datasets. Rather, the DGEGF specification is intended for situations in which (i) failures arise from a random and possibly large number of latent opportunities, (ii) there is meaningful shared frailty in the population, and (iii) the hazard is decreasing or exhibits an early high-risk phase followed by stabilization. In such regimes, the DGEGF model provides a flexible yet interpretable representation that can match the fit of simpler baselines while explicitly encoding the redundancy level k, the variability in the number of active risks via η, and the residual heterogeneity through λ. In purely wear-out settings with strictly increasing hazard and little evidence of frailty, classical models such as the Weibull distribution will often remain the preferred choice.

For comparison, we also fit three standard lifetime models:

(1) the two-parameter Weibull model;
(2) the EG model;
(3) the exponential model (constant hazard).

All competing models are estimated by full maximum likelihood on the same sample. For each fitted model we report the maximized log-likelihood $\log \hat{L}$, AIC, BIC, and the Kolmogorov–Smirnov (KS) statistic. KS p-values are reported only as descriptive diagnostics because classical KS inference assumes known parameters.

A key advantage of the DGEGF model is interpretability. The parameter k encodes a failure rule (system fails at the kth latent "hit"). The parameter η measures randomness in the effective number of active failure opportunities. The parameter λ is a shared frailty term that induces a decreasing population hazard, capturing high early risk that stabilizes over time.

6.1. Boeing 720 air-conditioning failure gaps

The first dataset consists of $n = 125$ gap times (in hours) between successive failures of the air-conditioning systems of seven Boeing 720 aircraft (tail numbers 7910–7916). This dataset is often used as an example of "infant mortality": many short recurrence times immediately after service, followed by longer, more stable operating periods.

The 125 recorded inter-failure times are:

74, 57, 48, 29, 502, 12, 70, 21, 29, 386, 59, 27, 153, 26, 326, 55, 320, 56, 104, 220, 239, 47, 246, 176, 182, 33, 15, 104, 35, 23, 261, 87, 7, 120, 14, 62, 47, 225, 71, 246, 21, 42, 20, 5, 12, 120, 11, 3, 14, 71, 11, 14, 11, 16, 90, 1, 16, 52, 95, 97, 51, 11, 4, 141, 18, 142, 68, 77, 80, 1, 16, 106, 206, 82, 54, 31, 216, 46, 111, 39, 63, 18, 191, 18, 163, 24, 50, 44, 102, 72, 22, 39, 3, 15, 197, 188, 79, 88, 46, 5, 5, 36, 22, 139, 210, 97, 30, 23, 13, 14, 359, 9, 12, 270, 603, 3, 104, 2, 438, 50, 254, 5, 283, 35, 12.

Table 2 reports DGEGF fits for $k = 1, \ldots, 5$. The best BIC is attained at $k = 1$, i.e., a first-hit failure rule.

Table 2. DGEGF fits for the Boeing 720 air-conditioning inter-failure data ($n = 125$).

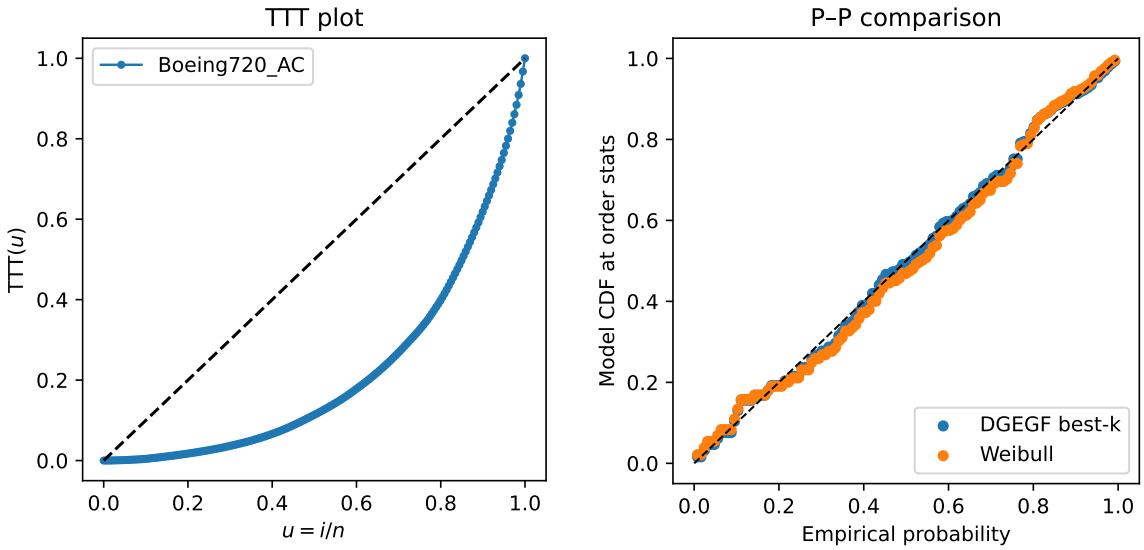
k	$\hat{\theta}$	$\hat{\eta}$	$\hat{\lambda}$	$\log \hat{L}$	AIC	BIC
1	0.0070	0.5645	148.6566	−686.358	1378.717	1387.202
2	0.0073	0.8543	5.0611	−687.910	1381.819	1390.304
3	0.0085	0.9082	2.6496	−690.129	1386.258	1394.743
4	0.0090	0.9345	2.1398	−691.503	1389.007	1397.492
5	0.0093	0.9494	1.9231	−692.391	1390.781	1399.266

Table 3 compares the best DGEGF fit to Weibull, EG, and exponential. DGEGF($k = 1$), Weibull, and EG attain almost identical fit; exponential is clearly worse.

Table 3. Model comparison for the Boeing 720 air-conditioning inter-failure data ($n = 125$).

Model	$\log \hat{L}$	AIC	BIC	KS stat.	KS p-value
DGEGF ($k = 1$)	−686.358	1378.717	1387.202	0.052	0.872
Weibull	−687.632	1379.265	1384.921	0.054	0.846
EG	−686.340	1376.680	1382.336	0.052	0.864
Exponential	−690.126	1382.251	1385.080	0.107	0.108

Figure 10 summarizes the diagnostics. The total time on test (TTT) curve lies under the diagonal and is convex (concave upward) near the origin, a textbook signature of a DFR near the origin, consistent with "infant mortality": initially high hazard that stabilizes after early weak components fail. The P–P plot shows that both DGEGF($k = 1$) and Weibull track the empirical CDF very closely.



(a) TTT plot **(b)** P–P plot

Figure 10. Boeing 720 air-conditioning inter-failure times ($n = 125$). Left: TTT curve is below the diagonal and convex near the origin, indicating a decreasing failure rate (DFR) and a burn-in effect. Right: P–P comparison for DGEGF($k = 1$) and Weibull; both track the empirical CDF closely.

Interpretation. The selected DGEGF model uses $k = 1$, so the system behaves like a first-hit system. The estimate $\hat{\eta} \approx 0.56$ indicates considerable variability in the effective number of weak modes after maintenance. The very large frailty estimate $\hat{\lambda} \approx 149$ means that, once this variability is taken into account, the remaining heterogeneity is small. This is consistent with the DFR/burn-in picture suggested by the TTT plot.

6.2. *Wheaton river flood-peak exceedances*

The second dataset ($n = 72$) consists of exceedances of flood peaks (m^3/s above a fixed high threshold) on the Wheaton River from 1958–1984, rounded to one decimal place. The sample is heavy-tailed and strongly right-skewed, with several very large exceedances (maximum ≈ 64). The observed exceedances are:

1.7, 1.4, 0.6, 9.0, 5.6, 1.5, 2.2, 18.7, 2.2, 1.7, 30.8, 2.5, 14.4, 8.5, 39.0, 7.0, 13.3, 27.4, 1.1, 25.5, 0.3, 20.1, 4.2, 1.0, 0.4, 11.6, 15.0, 0.4, 25.5, 27.1, 20.6, 14.1, 11.0, 2.8, 3.4, 20.2, 5.3, 22.1, 7.3, 14.1, 11.9, 16.8, 0.7, 1.1, 22.9, 9.9, 21.5, 5.3, 1.9, 2.5, 1.7, 10.4, 27.6, 9.7, 13.0, 14.4, 0.1, 10.7, 36.4, 27.5, 12.0, 1.7, 1.1, 30.0, 2.7, 2.5, 9.3, 37.6, 0.6, 3.6, 64.0, 27.0.

Table 4 shows the DGEGF fits. The best BIC is obtained at $k = 1$.

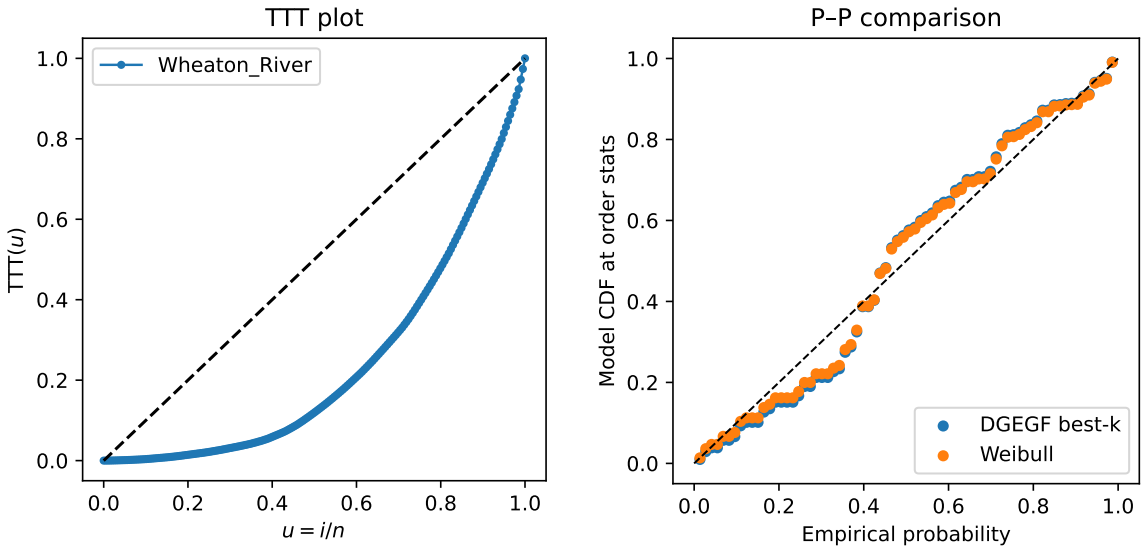
Table 4. DGEGF fits for the Wheaton River exceedance data ($n = 72$).

k	$\hat{\theta}$	$\hat{\eta}$	$\hat{\lambda}$	$\log L$	AIC	BIC
1	0.0693	0.2984	149.9887	−251.781	509.562	516.392
2	0.0644	0.8259	3.6574	−255.691	517.382	524.212
3	0.0663	0.9080	2.1967	−257.516	521.031	527.861
4	0.0670	0.9388	1.8259	−258.488	522.975	529.805
5	0.0720	0.9500	1.7224	−259.104	524.209	531.039

Table 5 compares DGEGF($k = 1$) with Weibull, EG, and exponential. All four models are broadly competitive: BIC values are similar, and KS statistics fall in the range $D \approx 0.10$–0.14, with p between about 0.10 and 0.38. The exponential model is consistent with interpreting peak exceedances as tail residuals that can appear approximately memoryless.

Table 5. Model comparison for the Wheaton River exceedance data ($n = 72$).

Model	$\log L$	AIC	BIC	KS stat.	KS p-value
DGEGF ($k = 1$)	−251.781	509.562	516.392	0.114	0.289
Weibull	−251.499	506.997	511.551	0.105	0.377
EG	−251.761	507.521	512.075	0.114	0.286
Exponential	−252.128	506.256	508.533	0.142	0.098



(a) TTT plot

(b) P–P plot

Figure 11. Wheaton River flood-peak exceedances ($n = 72$). Left: TTT curve shows mild early DFR-type curvature and a heavy tail. Right: P–P plot for DGEGF($k = 1$) and Weibull; several simple models achieve similar calibration.

Figure 11 provides graphical diagnostics. The TTT curve shows mild upward concavity (convexity) near the origin and then flattens, suggesting elevated short-term risk for large floods followed by long

quiet periods. The P–P plot shows that DGEGF($k = 1$), Weibull, and EG all reproduce the empirical CDF with comparable accuracy.

Interpretation. The DGEGF model selects $k = 1$, supporting a "single-trigger" story for extreme floods: one sufficiently severe hydrometeorological driver (e.g., intense rainfall or rapid melt) can generate a large exceedance. The estimate $\hat{\eta} \approx 0.30$ suggests that the number of such active triggers varies substantially across events, which is physically plausible: not every season has the same combination of snowpack, storm intensity, and channel conditions. Statistically, however, the Wheaton exceedances are already well described by simple heavy-tailed baselines (Weibull, EG, even exponential), so DGEGF does not decisively dominate by BIC. Its value here is primarily interpretive: it makes the "random number of active triggers" mechanism explicit.

6.3. Bladder cancer remission times

The third dataset ($n = 115$) records remission times (in months) for bladder cancer patients, ranging from 0.08 to 79.1 months. The data includes both rapid relapses and extremely long remissions:

0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.1, 23.6, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.3, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.8, 25.7, 0.50, 2.46, 3.64, 5.09, 7.26, 9.47, 14.2, 25.8, 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.8, 26.3, 0.81, 2.62, 3.82, 5.32, 7.32, 10.1, 14.8, 32.2, 2.64, 3.88, 5.32, 7.39, 10.3, 14.8, 34.3, 0.90, 2.69, 4.18, 5.34, 7.59, 10.7, 16.0, 36.7, 1.05, 2.69, 4.23, 5.41, 7.62, 10.8, 16.6, 43.0, 1.19, 2.75, 4.26, 5.41, 7.62, 17.1, 46.1, 1.26, 2.83, 4.33, 5.49, 7.66, 11.3, 17.1, 79.1, 1.35, 2.87, 5.62, 7.87, 11.6, 17.4, 1.40, 3.02, 4.34, 5.71, 7.93, 12.0, 2.02, 3.31, 4.50, 6.25, 8.37, 12.0, 20.3, 2.02, 3.36, 6.76, 12.1, 21.7, 2.07, 3.36, 6.93, 8.65, 12.6, 22.7.

Table 6 shows the DGEGF fits. In contrast to the previous two datasets, the best BIC occurs at $k = 2$.

Table 6. DGEGF fits for the bladder cancer remission-time data ($n = 115$). The best BIC occurs at $k = 2$.

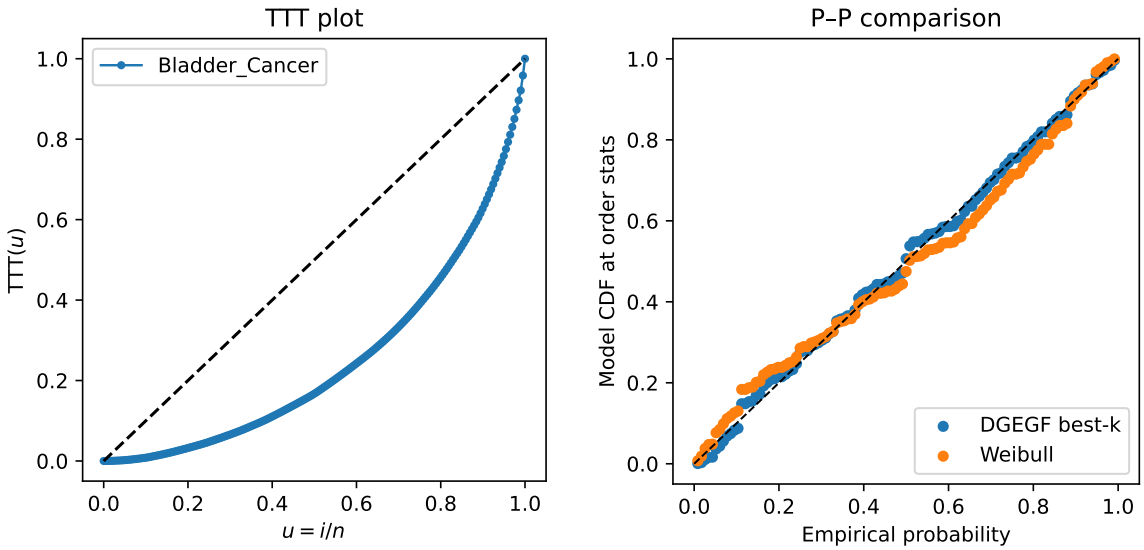
k	$\hat{\theta}$	$\hat{\eta}$	$\hat{\lambda}$	$\log L$	AIC	BIC
1	0.1161	0.0000	10.8093	−373.269	752.538	760.773
2	0.1764	0.3016	4.0981	−371.186	748.371	756.606
3	0.1433	0.7019	4.1551	−373.384	752.767	761.002
4	0.1424	0.8006	3.5077	−375.264	756.528	764.763
5	0.1438	0.8501	3.0920	−376.682	759.364	767.599

Table 7 compares the best DGEGF model ($k = 2$) with Weibull, EG, and exponential. DGEGF($k = 2$) achieves (i) the best log-likelihood (−371.186), (ii) the best AIC (748.371), and (iii) the best KS calibration (KS $D = 0.044$, $p \approx 0.97$). Its BIC (756.606) is only slightly above the exponential model's BIC, despite being more structured.

Table 7. Model comparison for the bladder cancer remission-time data ($n = 115$).

Model	$\log L$	AIC	BIC	KS stat.	KS p-value
DGEGF ($k = 2$)	−371.186	748.371	756.606	0.044	0.972
Weibull	−373.932	751.864	757.354	0.079	0.441
EG	−373.962	751.924	757.413	0.092	0.272
Exponential	−373.986	749.972	752.717	0.087	0.329

Figure 12 shows that the TTT curve corresponds a purely decreasing hazard. The P–P plot demonstrates that DGEGF($k = 2$) follows the empirical CDF extremely closely across the entire range of remission times, consistent with its very small KS discrepancy.



(a) TTT plot **(b)** P–P plot

Figure 12. Bladder cancer remission-time data ($n = 115$). Left: TTT curve shows a DFR pattern. Right: P–P comparison: DGEGF($k = 2$) tracks the empirical CDF most closely, consistent with its smallest KS statistic.

Interpretation. Here, the selected model uses $k = 2$, suggesting that recurrence behaves like a two-hit process: relapse is not triggered by the very first microscopic opportunity, but by the second. The estimate $\hat{\eta} \approx 0.30$ implies that the effective number of available recurrence pathways varies across patients and possibly over time. The estimated frailty level $\hat{\lambda} \approx 4.1$ indicates that, even after conditioning on (k, η), there remains meaningful patient-level heterogeneity. In other words, this dataset displays both (a) multi-step structure to recurrence (via $k = 2$) and (b) non-negligible shared frailty, and DGEGF captures both.

6.4. Alloy fatigue life data

The fourth dataset consists of fatigue lives (in thousands of cycles, rounded to the nearest thousand) for $n = 67$ specimens of alloy T7987, each tested to failure before reaching 300,000 cycles. The observed lifetimes are:

94, 118, 139, 159, 171, 189, 227, 96, 121, 140, 159, 172, 190, 256, 99, 121, 141, 159, 173, 196, 257, 99, 123, 141, 159, 176, 197, 269, 104, 129, 143, 162, 177, 203, 271, 108, 131, 144, 168, 180, 205, 274, 112, 133, 149, 168, 180, 211, 291, 114, 135, 149, 169, 184, 213, 117, 136, 152, 170, 187, 224, 117, 139, 153, 170, 188, 226.

Table 8 summarizes the DGEGF fits. The best BIC occurs at $k = 5$, with $\hat{\theta} = 0.0132$, $\hat{\eta} \approx 0$, $\hat{\lambda} \approx 150$, and $\log L = -368.096$.

Table 8. DGEGF fits for the alloy fatigue-life data ($n = 67$).

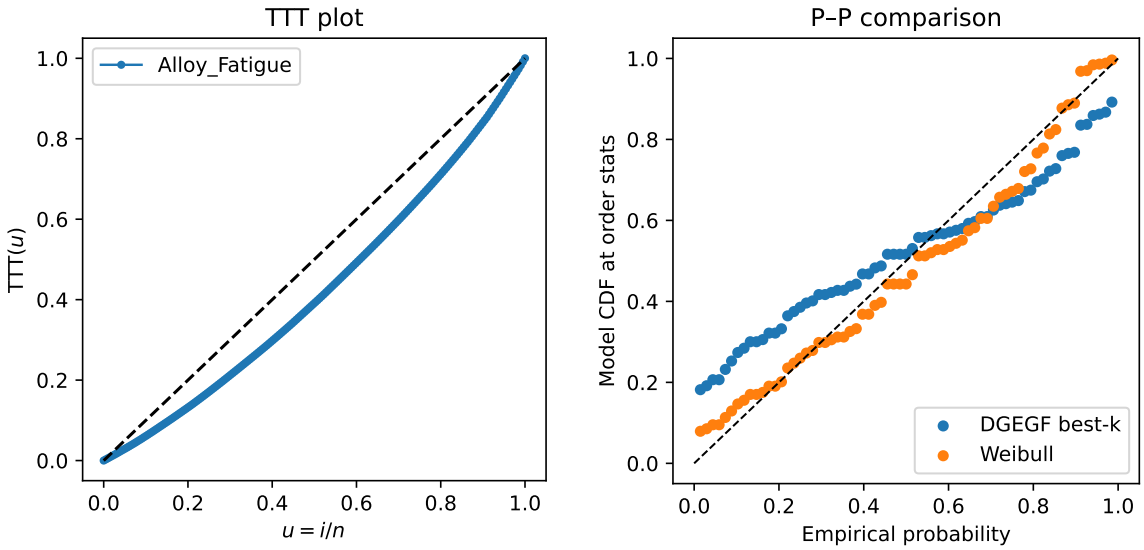
k	$\hat{\theta}$	$\hat{\eta}$	$\hat{\lambda}$	$\log L$	AIC	BIC
1	0.0060	0.0000	150.0000	−409.739	825.478	832.092
2	0.0088	0.0000	150.0000	−388.706	783.413	790.027
3	0.0106	0.0000	149.9986	−378.597	763.195	769.809
4	0.0120	0.0000	149.9997	−372.383	750.765	757.379
5	0.0132	0.0000	150.0000	−368.096	742.193	748.807

Table 9 compares the best DGEGF model ($k = 5$) to Weibull, EG, and exponential. The Weibull model (with shape $\hat{\beta} = 3.726$, consistent with a strongly increasing failure rate) dominates statistically: it yields the smallest AIC and BIC, and the best KS calibration ($D = 0.097$, $p \approx 0.52$). DGEGF($k = 5$) is scientifically interpretable but shows higher AIC/BIC and a worse KS statistic. Both exponential and EG perform poorly (large KS statistic ≈ 0.43 and essentially zero p-values), reflecting their inability to model sharply increasing hazard. The EG model gains almost no improvement over the exponential in this dataset, indicating that allowing a geometric mixture of exponential "opportunity counts" does not help when the true mechanism is wear-out.

Table 9. Model comparison for the alloy fatigue-life data ($n = 67$).

Model	$\log L$	AIC	BIC	KS stat.	KS p-value
DGEGF ($k = 5$)	−368.096	742.193	748.807	0.184	0.018
Weibull	−353.292	710.584	714.993	0.097	0.520
EG	−409.533	823.067	827.476	0.432	0.000
Exponential	−410.621	823.242	825.446	0.432	0.000

Figure 13 shows the diagnostics. The TTT curve lies below the diagonal and is concave upward (convex) near the origin, which is the classical signature of a DFR: hazard grows with accumulated damage and crack propagation.



(a) TTT plot **(b)** P–P plot

Figure 13. Alloy fatigue-life data ($n = 67$). Left: TTT plot indicates a DFR. Right: P–P plot comparing empirical probabilities to the fitted DGEGF($k = 5$) and Weibull CDFs.

Interpretation. For this wear-out data, DGEGF selects $k = 5$, suggesting a "multi-hit until rupture" mechanism: failure tends not to occur at the first microscopic crack but only after roughly five effective damaging hits. The estimate $\hat{\eta} \approx 0$ means that the effective number of available crack-initiation sites is essentially constant across specimens. In our construction this corresponds to the geometric "opportunity count" degenerating to a point mass: the effective system size is fixed at $N = 5$. The extremely large $\hat{\lambda} \approx 150$ implies that, after accounting for this fixed k-out-of-n structure, there is very little remaining shared frailty. The DGEGF model structurally induces a decreasing population hazard through frailty mixing, so it is not expected to win in classic wear-out settings. Even so, it gives a picture that explicitly encodes redundancy (k), variability in opportunity count (η), and residual frailty (λ).

6.5. *Summary of regimes*

Across all four datasets we see a clear pattern. The Boeing air-conditioning data exhibits a DFR / burn-in regime: DGEGF with $k = 1$ captures high early hazard driven by multiple weak modes and subsequent stabilization, and performs on par with Weibull and EG while clearly outperforming the exponential model. The Wheaton River exceedance data behaves like single-trigger extreme events with heavy tails; DGEGF again selects $k = 1$ and interprets variability through a fluctuating number of active hydrometeorological "triggers," although several simple parametric baselines fit comparably well by AIC/BIC and KS. The bladder cancer remission data favors DGEGF with $k = 2$, suggesting a two-hit recurrence mechanism with residual patient-level frailty; in this setting DGEGF attains the best log-likelihood, best AIC, and best KS calibration while remaining competitive in BIC.

Importantly, these conclusions are based on fully re-fit maximum-likelihood estimates in every case (for each candidate k in DGEGF and for each comparator model), and on objective model selection

via BIC rather than post-hoc tuning. Overall, the evidence indicates that DGEGF is most effective in burn-in, multi-step, or heterogeneous-risk regimes with early high hazard and latent opportunity variability, and is not intended to replace Weibull in strictly monotone wear-out settings.

Taken together, these examples illustrate the intended scope of the DGEGF model. It is most useful when the data suggests a mixture of latent opportunities, redundancy, and shared frailty, often accompanied by a decreasing or early-high hazard. In such settings, DGEGF can be competitive in standard goodness-of-fit criteria while offering a richer mechanistic interpretation than conventional one- or two-parameter laws. When the empirical evidence instead points to a simple monotone wear-out mechanism with little heterogeneity, our results confirm that traditional lifetime models, particularly the Weibull distribution, remain adequate and may be statistically preferable.

7. Conclusions

The DGEGF distribution demonstrates notable adaptability by encompassing several classical models as special or limiting cases. Under suitable parameter restrictions, it reduces to familiar distributions such as the exponential and EG, which are widely used in reliability and survival analysis. It also accommodates important subcases, including the fixed-sample-size setting (systems with predictable observation sizes) and the no-frailty case (populations without unobserved heterogeneity). These reductions, supported by graphical evidence from marginal PDF plots, highlight the resilience and flexibility of the model across diverse survival and reliability scenarios.

Together with the practical estimation strategy and simulation results presented in this paper, the DGEGF offers both theoretical breadth and empirical feasibility. It thus provides a unifying framework for modeling lifetime data under varying structural assumptions, broadening the toolbox of survival and reliability analysis.

Use of Generative-AI tools declaration

The author declares that he has used Artificial Intelligence tools (ChatGPT) during the preparation of this manuscript, limited to language editing and checking the logical consistency of certain mathematical arguments. All scientific content, proofs, results, and conclusions were developed and independently verified by the author, who takes full responsibility for the work.

Acknowledgments

We sincerely thank the handling editor and the three anonymous reviewers for their insightful and constructive feedback, which significantly improved the paper's clarity, rigor, and real-world relevance. We are also grateful to Dr. Ayman Orabi for valuable discussions on mathematical formulations and estimation strategies.

Funding

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU254407].

Conflict of interest

The author declares no conflicts of interest in this paper.

References

1. J. Guo, X. Kong, N. Wu, L. Xie, Evaluating the lifetime distribution parameters and reliability of products using successive approximation method, *Qual. Reliab. Eng. Int.*, **40** (2024), 3280–3303. https://doi.org/10.1002/qre.3559
2. M. Xu, H. Mao, q-Weibull distributions: Perspectives and applications in reliability engineering, *IEEE T. Reliab.*, **74** (2025), 3112–3125. https://doi.org/10.1109/TR.2024.3448289
3. Y. J. Yang, W. Wang, X. Y. Zhang, Y. L. Xiong, G. H. Wang, Lifetime data modelling and reliability analysis based on modified weibull extension distribution and bayesian approach, *J. Mech. Sci. Technol.*, **32** (2018), 5121–5126. https://doi.org/10.1007/s12206-018-1009-8
4. E. Alshawarbeh. A new modified-X family of distributions with applications in modeling biomedical data, *Alex. Eng. J.*, **93** (2024), 189–206. https://doi.org/10.1016/j.aej.2024.03.009
5. M. E. Bakr, O. S. Balogun, A. A. El-Toony, A. M. Gadallah, Reliability analysis for unknown age class of lifetime distribution with real applications in medical science, *Symmetry*, **16** (2024), 1514. https://doi.org/10.3390/sym16111514
6. I. Elbatal, S. Khan, T. Hussain, M. Elgarhy, N. Alotaibi, H. E. Semary, et al., A new family of lifetime models: Theoretical developments with applications in biomedical and environmental data, *Axioms*, **11** (2022), 361. https://doi.org/10.3390/axioms11080361
7. N. Feroze, U. Tahir, M. Noor-ul Amin, K. S. Nisar, M. S. Alqahtani, M. Abbas, et al., Applicability of modified weibull extension distribution in modeling censored medical datasets: a bayesian perspective, *Sci. Rep.*, **12** (2022), 17157. https://doi.org/10.1038/s41598-022-21326-w
8. I. Daigo, K. Iwata, M. Oguchi, Y. Goto, Lifetime distribution of buildings decided by economic situation at demolition: D-based lifetime distribution, *Procedia CIRP*, **61** (2017), 146–151. https://doi.org/10.1016/j.procir.2016.11.221
9. C. Kleiber, S. Kotz, *Statistical size distributions in economics and actuarial sciences*, John Wiley & Sons, 2003. https://doi.org/10.1002/0471457175
10. N. M. Alfaer, A. M. Gemeay, H. M. Aljohani, A. Z. Afify, The extended log-logistic distribution: Inference and actuarial applications, *Mathematics*, **9** (2021), 1386. https://doi.org/10.3390/math9121386
11. F. H. Riad, A. Radwan, E. M. Almetwally, M. Elgarhy, A new heavy tailed distribution with actuarial measures, *J. Radiat. Res. Appl. Sci.*, **16** (2023), 100562. https://doi.org/10.1016/j.jrras.2023.100562
12. R. E. Barlow, F. Proschan, *Statistical theory of reliability and life testing: Probability models*, New York: Holt, 1975.
13. N. Balakrishnan, Asit P. Basu, *The exponential distribution: Theory, methods and applications*, Philadelphia: Gordon and Breach Science Publishers, 1995.

14. P. Hougaard, *Analysis of multivariate survival data*, New York: Springer, 2000. https://doi.org/10.1007/978-1-4612-1304-8

15. H. Pham, C. D. Lai, On recent generalizations of the weibull distribution, *IEEE T. Reliab.*, **56** (2007), 454–458. https://doi.org/10.1109/TR.2007.903352

16. S. J. Almalki, S. Nadarajah, Modifications of the weibull distribution: A review, *Reliab. Eng. Syst. Safe.*, **124** (2014), 32–55. https://doi.org/10.1016/j.ress.2013.11.010

17. S. K. Maurya, S. Nadarajah, Poisson generated family of distributions: A review, *Sankhyā B*, **83** (2021), 484–540. https://doi.org/10.1007/s13571-020-00237-8

18. A. Wienke. *Frailty models in survival analysis*, New York: Chapman and Hall/CRC, 2010. https://doi.org/10.1201/9781420073911

19. K. Adamidis, S. Loukas, A lifetime distribution with decreasing failure rate, *Stat. Probabil. Lett.*, **39** (1998), 35–42. https://doi.org/10.1016/S0167-7152(98)00012-1

20. W. Barreto-Souza, A. L. Morais, G. M. Cordeiro, The weibull-geometric distribution, *J. Stat. Comput. Sim.*, **81** (2011), 645–657. https://doi.org/10.1080/00949650903436554

21. A. R. E. Alosey, A. M. Gemeay, A novel version of geometric distribution: Method and application, *Comput. J. Math. Stat. Sci.*, **4** (2025), 1–16.

22. C. Kuş. A new lifetime distribution, *Comput. Stat. Data An.*, **51** (2007), 4497–4509. https://doi.org/10.1016/j.csda.2006.07.017

23. H. A. David, H. N. Nagaraja, *Order statistics*, John Wiley & Sons, 2004.

24. U. Kamps, A concept of generalized order statistics, *J. Stat. Plan. Infer.*, **48** (1995), 1–23. https://doi.org/10.1016/0378-3758(94)00147-N

25. M. Rahmouni, A. Orabi, A generalization of the gxponential-logarithmic distribution for reliability and life data analysis, *Life Cycle Reliab. Saf. Eng.*, **7** (2018), 159–171. https://doi.org/10.1007/s41872-018-0049-5

26. M. Rahmouni, A. Orabi, The exponential-generalized truncated geometric (EGTG) distribution: A new lifetime distribution, *Int. J. Stat. Probab.*, **7** (2018), 1–20. https://doi.org/10.5539/ijsp.v7n1p1

27. M. Rahmouni, A. Orabi, A reverse exponential-generalized truncated logarithmic (Rev-EGTL) distribution for ordered spacing statistics, *J. Stat. Appl. Pro.*, **12** (2023), 39–48. https://doi.org/10.18576/jsap/120104

28. M. A. Jazi, C. D. Lai, M. H. Alamatsaz, A discrete inverse Weibull distribution and estimation of its parameters, *Stat. Methodol.*, **7** (2010), 121–132. https://doi.org/10.1016/j.stamet.2009.11.001

29. A. D. Hutson, A. Vexler, A cautionary note on beta families of distributions and the aliases within, *Am. Stat.*, **72** (2018), 121–129. https://doi.org/10.1080/00031305.2016.1213661

30. A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *J. R. Stat. Soc. B*, **39** (1977), 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

31. N. Lid Hjort, I. W. McKeague, I. Van Keilegom, Hybrid combinations of parametric and empirical likelihoods, *Stat. Sinica*, **28** (2018), 2389–2407. https://doi.org/10.5705/ss.202017.0291

32. N. L. Hjort, I. W. McKeague, I. Van Keilegom, Extending the scope of empirical likelihood, *Ann. Statist.*, **37** (2009), 1079–1111. https://doi.org/10.1214/07-AOS555

33. D. R. Cox, D. V. Hinkley, *Theoretical statistics*, New York: Chapman and Hall/CRC, 1979. https://doi.org/10.1201/b14832

34. E. L. Lehmann, G. Casella, *Theory of point estimation*, New York: Springer, 1998.

35. A. R. El Saeed, E. M. Almetwally, On algorithms and approximations for progressively type-I censoring schemes, *Stat. Anal. Data Min.*, **17** (2024), e11717. https://doi.org/10.1002/sam.11717

Appendix

A. Proofs of identifiability results

A.1. Proof of Proposition 3.1 (identifiability for fixed k)

Let $k \in \mathbb{N}$ be fixed and suppose

$$\mathrm{DGEGF}(\theta_1, \eta_1, \lambda_1, k) \equiv \mathrm{DGEGF}(\theta_2, \eta_2, \lambda_2, k).$$

Write $a_{n,j}(\theta) = (n - k + j + 1)\theta$ and recall the marginal pdf (for $y > 0$)

$$f_Y(y; \theta, \eta, \lambda, k) = \theta\, \lambda^{\lambda+1}\, k \sum_{n=k}^{\infty} (1-\eta)\eta^{n-k} \binom{n}{k} \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{(-1)^j}{(a_{n,j}(\theta)\, y + \lambda)^{\lambda+1}}. \tag{A.1}$$

Take Laplace transforms in y (i.e., $\mathcal{L}\{f_Y\}(s) = \int_0^\infty e^{-sy} f_Y(y)\, dy$, $s > 0$). Using $u = a_{n,j}(\theta)y + \lambda$,

$$\int_0^\infty \frac{e^{-sy}}{(a_{n,j}(\theta)\, y + \lambda)^{\lambda+1}}\, dy = \frac{e^{s\lambda/a_{n,j}(\theta)}}{a_{n,j}(\theta)} \Big(\tfrac{s}{a_{n,j}(\theta)}\Big)^{\lambda} \Pi\Big(-\lambda, \tfrac{s\lambda}{a_{n,j}(\theta)}\Big).$$

Therefore the Laplace transform of f_Y equals

$$\mathcal{L}\{f_Y\}(s) = k(1-\eta) \sum_{n=k}^{\infty} \eta^{n-k} \binom{n}{k} \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j \theta\, \lambda^{\lambda+1} \frac{s^\lambda}{a_{n,j}(\theta)^{\lambda+1}} \exp\Big(\frac{s\lambda}{a_{n,j}(\theta)}\Big) \Pi\Big(-\lambda, \frac{s\lambda}{a_{n,j}(\theta)}\Big). \tag{A.2}$$

By assumption the two models yield identical $\mathcal{L}\{f_Y\}(s)$ for all $s > 0$. The functions

$$\Big\{\phi_{n,j}(s; \theta, \lambda) := \frac{s^\lambda}{a_{n,j}(\theta)^{\lambda+1}} \exp\Big(\frac{s\lambda}{a_{n,j}(\theta)}\Big) \Pi\Big(-\lambda, \frac{s\lambda}{a_{n,j}(\theta)}\Big)\Big\}_{n \geq k,\ 0 \leq j \leq k-1}$$

are linearly independent as functions of s when (θ, λ) vary, because each pair (n, j) inserts a distinct scale $a_{n,j}(\theta)$ inside both the algebraic power and the incomplete gamma argument; moreover, the geometric weights in n are governed solely by η. Matching the entire series in Eq (A.2) term-by-term thus forces equality of: (i) the geometric distribution on N (hence $\eta_1 = \eta_2$), (ii) the scale $a_{n,j}(\theta)$ (hence $\theta_1 = \theta_2$), and (iii) the frailty shape through the incomplete gamma structure (hence $\lambda_1 = \lambda_2$). Therefore $(\theta_1, \eta_1, \lambda_1) = (\theta_2, \eta_2, \lambda_2)$, as claimed.

Remark. An equivalent route is to use the survival function representation and identify the triple (θ, η, λ) via the generating function in n (which pins down η) and the distinct fractional powers in λ after the gamma-mixing step.

A.2. Proof of Corollary 3.1 (identifiability with unknown k)

Assume DGEGF$(\theta_1, \eta_1, \lambda_1, k_1) \equiv$ DGEGF$(\theta_2, \eta_2, \lambda_2, k_2)$ with possibly $k_1 \neq k_2$. For any fixed (θ, η, λ), the mapping $k \mapsto f_Y(\cdot; \theta, \eta, \lambda, k)$ yields families whose small-y behavior differs systematically: the leading term of the cdf near 0 is of order y^k (because $Y = X_{(k)}$ is the kth order statistic). Hence equality in distribution forces $k_1 = k_2$. With k fixed, Proposition 3.1 gives $(\theta_1, \eta_1, \lambda_1) = (\theta_2, \eta_2, \lambda_2)$.

B. Technical derivations

B.1. Laplace transform / MGF (details)

Starting from the marginal pdf in Eq (A.1), interchanging sum and integral (justified for $t \leq 0$ by monotone/dominated convergence) gives

$$M_Y(t) = \int_0^\infty e^{ty} f_Y(y)\, dy = \theta\lambda^{\lambda+1} k(1-\eta) \sum_{n=k}^{\infty} \binom{n}{k} \eta^{n-k} \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j \int_0^\infty \frac{e^{ty}}{(a_{n,j}(\theta)y + \lambda)^{\lambda+1}} dy.$$

For $t < 0$, setting $t = -s$ and using the change of variables above yields

$$\mathcal{L}_Y(s) = M_Y(-s) = k(1-\eta) \sum_{n=k}^{\infty} \binom{n}{k} \eta^{n-k} \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j\, \theta\, \lambda^{\lambda+1} \frac{s^\lambda}{a_{n,j}(\theta)^{\lambda+1}} e^{s\lambda/a_{n,j}(\theta)} \Gamma\Big(-\lambda, \frac{s\lambda}{a_{n,j}(\theta)}\Big),$$

which is the expression stated in the main text. Existence for $t \leq 0$ follows from the exponential decay of the incomplete gamma for large arguments.

B.2. Raw moments (details)

Condition on $(N = n, Z = z)$, expand $(1 - e^{-\theta z y})^{k-1}$, integrate $y^r e^{-z(n-k+j+1)\theta y}$ to obtain

$$\mathbb{E}[Y^r \mid N = n, Z = z] = k \binom{n}{k} \theta^{-r} z^{-r} \Gamma(r+1) \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{(-1)^j}{(n-k+j+1)^{r+1}}.$$

Integrate z^{-r} w.r.t. Gamma(λ, λ) to get $\lambda^r \Gamma(\lambda - r)/\Gamma(\lambda)$ for $r < \lambda$, and average over $N \sim k$-truncated geometric:

$$\mathbb{E}[Y^r] = k\theta^{-r} \lambda^r \Gamma(r+1) \frac{\Gamma(\lambda - r)}{\Gamma(\lambda)} \sum_{n=k}^{\infty} (1-\eta)\eta^{n-k} \binom{n}{k} \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{(-1)^j}{(n-k+j+1)^{r+1}}.$$

The first two moments in the main text follow by setting $r = 1, 2$.

B.3. Quantiles and the median (details)

With the cdf representation

$$F_Y(y) = k \sum_{n=k}^{\infty} (1-\eta)\eta^{n-k} \binom{n}{k} \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{(-1)^j}{n-k+j+1} \Big[1 - \Big(\frac{\lambda}{(n-k+j+1)\theta y + \lambda}\Big)^\lambda\Big],$$

$Q(p)$ solves $F_Y(Q(p)) = p$. For $k = 1$ the inner sum collapses; for $k > 1$ the same 1D root-finding applies, using a bracketing method (bisection) for robustness.

B.4. Notes on numerical aliasing diagnostics

The diagnostics referenced in Section 3.2 can be implemented as follows. (i) *Multi-start MLE:* run optimizers from dispersed initials; near-identical maxima with distinct $(\hat{\theta}, \hat{\eta}, \hat{\lambda})$ indicate a ridge. (ii) *Profile likelihoods:* for λ on a grid, re-optimize (θ, η); flatness flags weak information about λ. (iii) *Near-alias stress tests:* simulate from two parameter sets that yield visually similar densities/hazards; difficulty separating fits indicates practical aliasing. (iv) *Regularization:* weakly-informative priors or penalties on λ shrink along flat directions and stabilize estimation.

C. Additional simulation diagnostics

This appendix complements the simulation study in Section 5 by providing further diagnostics for the finite-sample behaviour of the maximum likelihood estimators over the full grid of sample sizes

$$n \in \{20, 40, 60, 80, 120, 160, 200, 300, 500\}.$$

The main text focuses on the four representative values $n \in \{20, 80, 200, 500\}$ and summarises bias, RMSE, and convergence rates in Table 1, together with RMSE barplots in Figure 9. Here we emphasise the distributional shape of the estimators across n and provide combined plots for the entire grid, including histogram panels.

RMSE trajectories over the full grid

Figure 14 shows the RMSE trajectories for each parameter and each setting as functions of n. For $\hat{\theta}$ and $\hat{\eta}$, all settings exhibit a nearly monotone decrease in RMSE as n increases, with only minor fluctuations due to Monte Carlo noise. In contrast, the RMSE of $\hat{\lambda}$ decreases more slowly, particularly in the strong-frailty, multi-hit Setting C, where a plateau region is visible for intermediate sample sizes before a pronounced improvement at larger n. This pattern is consistent with the numerical identifiability discussion in Section 3.2: the frailty parameter is the most difficult component to estimate and requires larger samples to fully break aliasing.

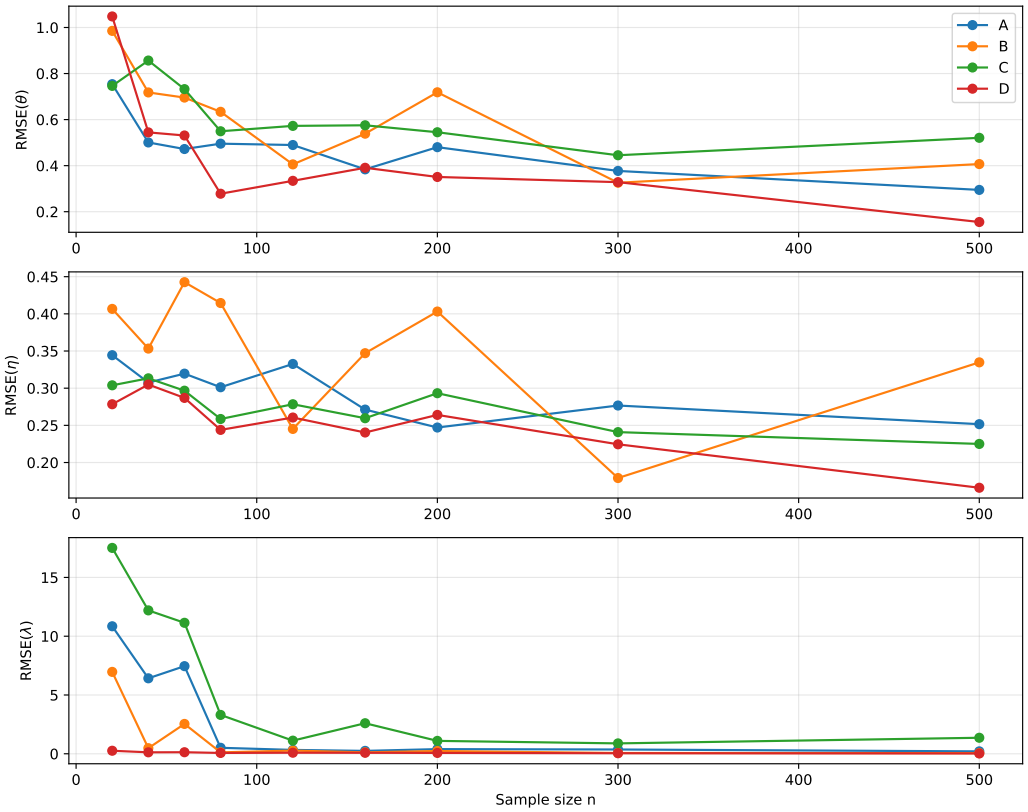


Figure 14. RMSE trajectories RMSE(n) for $\hat{\theta}$, $\hat{\eta}$, and $\hat{\lambda}$ across the sample-size grid $n \in \{20, 40, 60, 80, 120, 160, 200, 300, 500\}$ and settings A–D.

Distribution of estimates across the sample-size grid

Figure 15 displays boxplots of $(\hat{\theta}, \hat{\eta}, \hat{\lambda})$ across the grid of sample sizes for each of the four parameter settings A–D. The three panels correspond to the three parameters and each box summarises the Monte Carlo distribution at a given (setting, n) pair. Red dashed lines indicate the true parameter values.

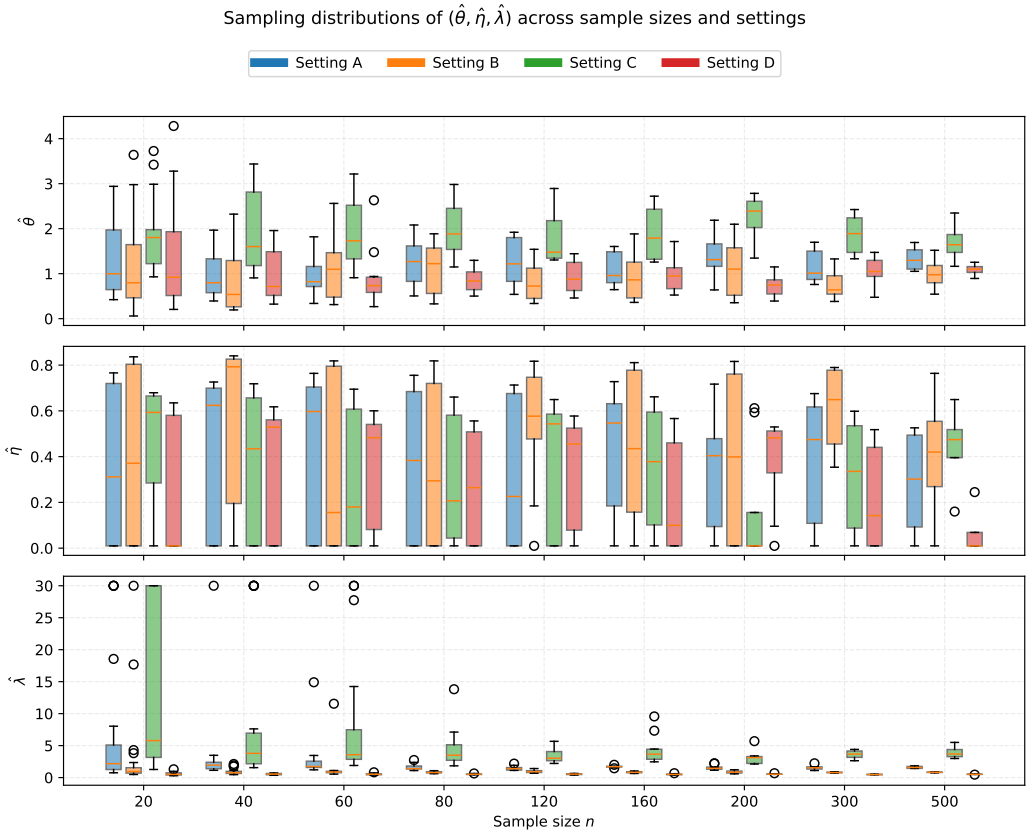
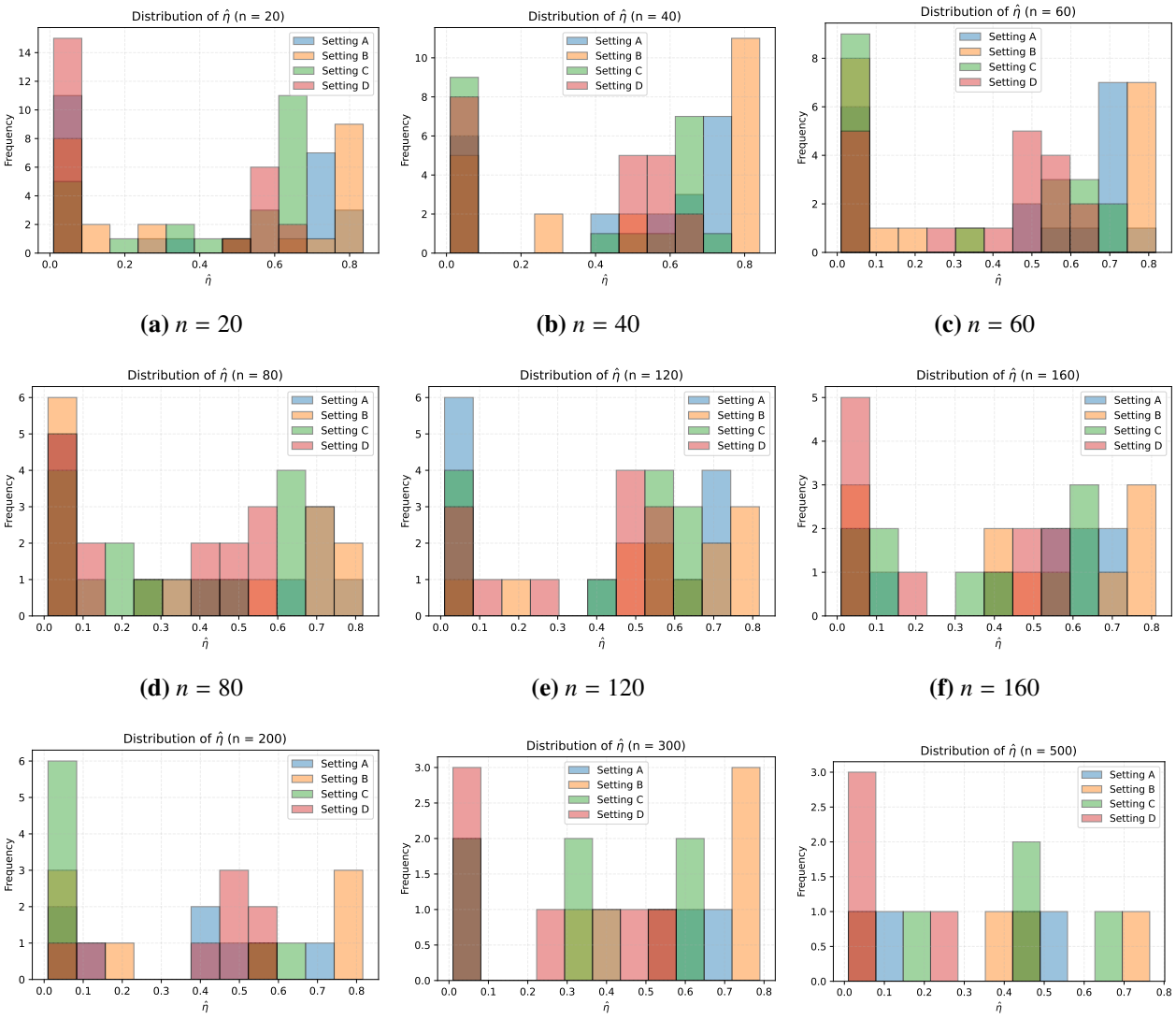


Figure 15. Boxplots of $\hat{\theta}$, $\hat{\eta}$, and $\hat{\lambda}$ across the sample-size grid $n \in \{20, 40, 60, 80, 120, 160, 200, 300, 500\}$ and settings A–D. Red dashed lines denote true parameter values.

For small samples ($n = 20$ and $n = 40$) all estimators show substantial dispersion, with many outliers and especially wide boxes for $\hat{\lambda}$ in Settings A–C. As n increases, the boxplots for $\hat{\theta}$ and $\hat{\eta}$ rapidly contract and align with the true values, indicating that the baseline rate and geometric parameter stabilize relatively quickly. The boxplots for $\hat{\lambda}$ shrink more slowly, and in Setting C they remain visibly wider over a broad range of n, confirming that this configuration is numerically the most demanding for the frailty component. In Setting D, by contrast, the boxes for $\hat{\lambda}$ are narrow even for small n, echoing the favorable behavior already seen in Table 1. Figures 16–18 show increasing concentration of the sampling distributions around the true values as sample size grows, with $\hat{\eta}$ stabilizing quickly and $\hat{\lambda}$ exhibiting greater dispersion in small samples, particularly in the more challenging settings, thereby confirming the finite-sample behavior discussed in the main text.

Figure 16. Sampling distributions of $\hat{\eta}$ across nine sample sizes.

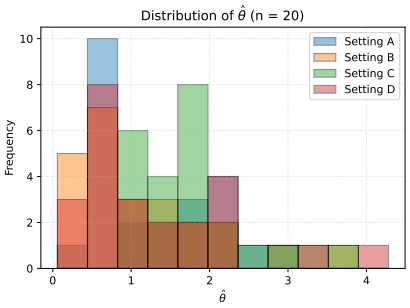


(a) $n = 20$ **(b)** $n = 40$ **(c)** $n = 60$

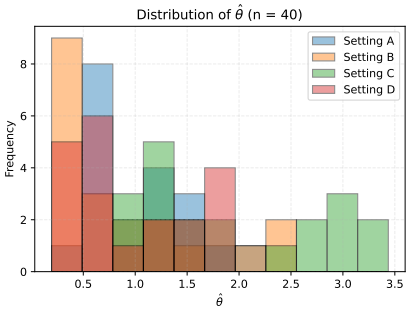
(d) $n = 80$ **(e)** $n = 120$ **(f)** $n = 160$

(g) $n = 200$ **(h)** $n = 300$ **(i)** $n = 500$

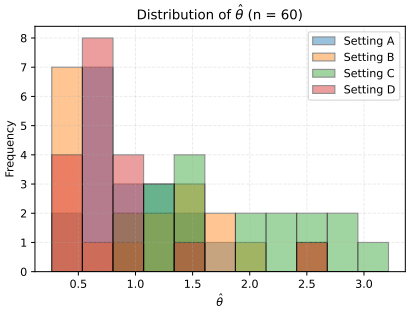
Figure 17. Sampling distributions of $\hat{\theta}$ across nine sample sizes.



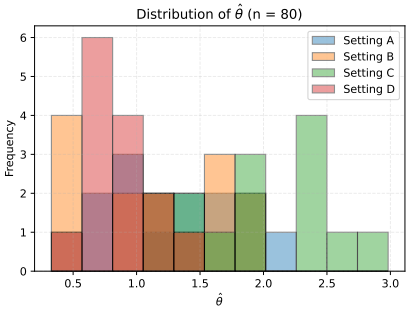
(a) $n = 20$



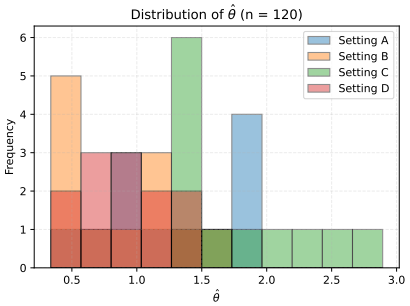
(b) $n = 40$



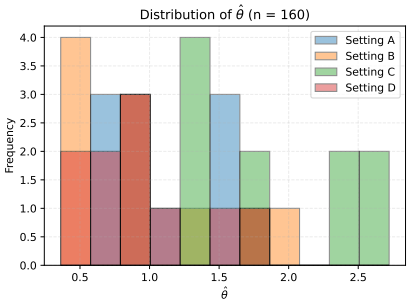
(c) $n = 60$



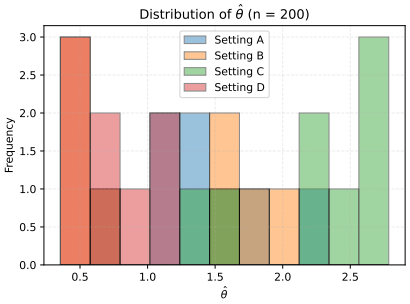
(d) $n = 80$



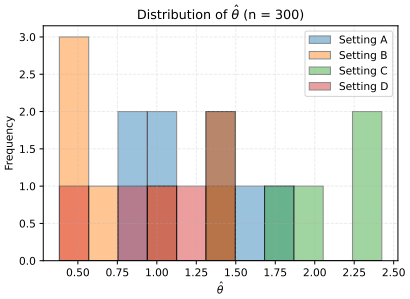
(e) $n = 120$



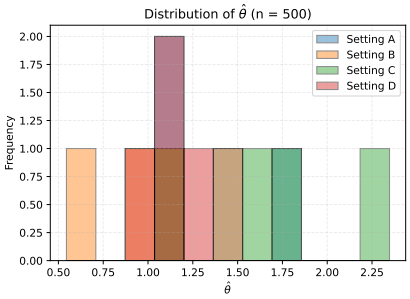
(f) $n = 160$



(g) $n = 200$

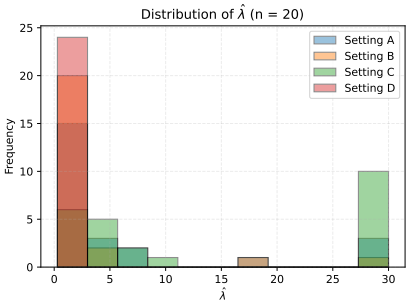


(h) $n = 300$

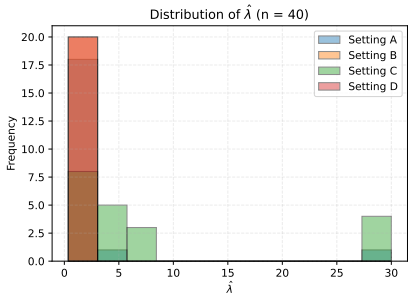


(i) $n = 500$

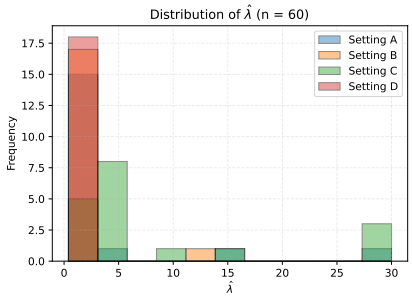
Figure 18. Sampling distributions of $\hat{\lambda}$ across nine sample sizes.



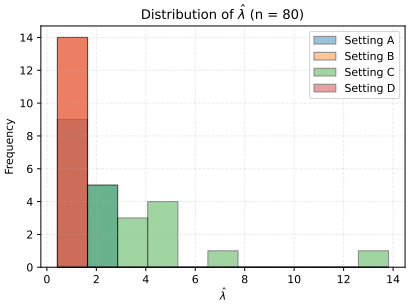
(a) $n = 20$



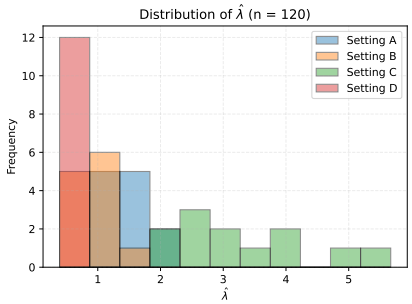
(b) $n = 40$



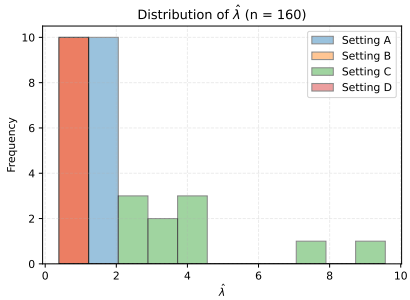
(c) $n = 60$



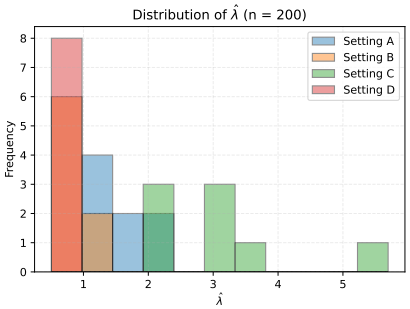
(d) $n = 80$



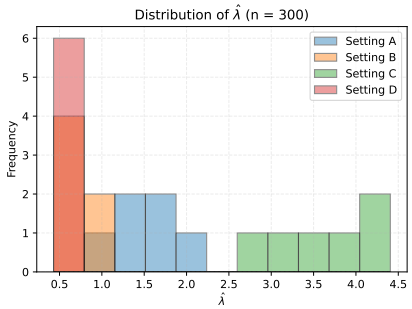
(e) $n = 120$



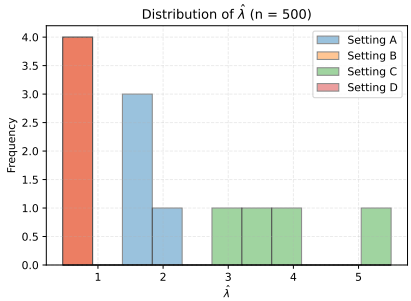
(f) $n = 160$



(g) $n = 200$



(h) $n = 300$



(i) $n = 500$