



Research article**Inference for a step-stress model under the inverted Z-Lindley distribution with progressive Type-II censoring****Abdullah H. Alenezy¹, L.S. Diab², Khudhayr A. Rashedi¹, and Ghareeb A. Marei^{3,*}**¹ Department of mathematics, College of Science, University of Ha'il, Hail, Saudi Arabia² Department of Mathematics and Statistics, College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia³ Faculty of Computers and Information Systems, Egyptian Chinese University, Nasr City, Egypt*** Correspondence:** Email: ghareeb.adel@ecu.edu.eg.

Abstract: We develop an explicit step–stress cumulative exposure model (CEM) formulation with closed–form equivalent times, analytic score functions, and analytic observed information for the one–parameter inverted Z–Lindley (IZL) lifetime distribution under progressive Type–II censoring. Because the IZL quantile inverts in closed form via the Lambert W function, the equivalent start time linking the two stress stages is itself closed form, giving an explicit rather than implicit likelihood, from which we derive the exact score, prove strict concavity of the constant–stress log–likelihood for complete samples, and obtain the analytic observed information matrix, validated against finite differences. The construction extends to a multiple–step generalization; numerical change–time optimization under c –, D –, and A –optimality for a representative complete–sample design; and a link–function extrapolation to an untested design stress with delta–method variance. Inference is triangulated across maximum likelihood, Wald, percentile bootstrap, and gamma–prior Bayesian analysis, cross–checked by the Tierney–Kadane approximation. A Monte Carlo study shows finite–sample bias and mean squared error (MSE) decreasing with sample size and near–nominal coverage for the Wald, Bayesian, and bootstrap intervals in most scenarios as well as accurate design–stress extrapolation under a correctly specified link, which degrades sharply once the link is misspecified. The principal estimation and uncertainty–quantification procedures are illustrated using a real, Type–I–censored step–stress data set, and design–stress extrapolation is assessed separately through a four–level simulation; the real–data example thus validates the broader CEM/IZL modeling and inference approach rather than the progressive Type–II censoring likelihood itself. On a small unmanned aerial vehicle (SUAV) reliability data set, a bootstrap–calibrated goodness–of–fit test finds no evidence against the fitted model, and IZL is among a small cluster of equally well–supported models by AICc – supporting IZL as a credible, closed–form candidate for step–stress reliability analysis.

Keywords: Inverted Z–Lindley distribution; step–stress; cumulative exposure model; progressive

Type-II censoring; Lambert W ; observed information; optimal change time; design-stress extrapolation; Tierney-Kadane

Mathematics Subject Classification: 62N05, 62N02, 62F10, 62F15

1. Introduction

Reliability assessment of highly reliable components is frequently obstructed by a practical difficulty: At their normal operating conditions, such units fail so rarely that a test of feasible duration yields too few failures to support inference. Accelerated life testing (ALT) resolves this by subjecting the units to levels of an accelerating variable—temperature, voltage, mechanical load, and humidity—more severe than those of the design environment, thereby temporally compressing the failure process. The life law is fitted to the accelerated data and the reliability at the lower, in-service stress is recovered by extrapolation through a physically motivated stress–life relationship. The methodology and its engineering practice are treated comprehensively by [1–3].

Among ALT plans, the way the stress is applied over time defines the experiment. Under a *constant-stress* plan, each unit is held at a fixed level; under a *step-stress* plan, a single cohort is started at a low level which is raised, at prescribed epochs, to successively higher levels. The step-stress design is operationally economical—it uses one set of test positions and one cohort, and it guarantees failures by escalating the stress—but it raises a modeling question absent from constant-stress testing: How should the lifetime distributions governing the successive stress levels be related? The accepted answer is the cumulative exposure model (CEM) of [4], which postulates that a unit’s residual life depends only on the exposure (the cumulative fraction failed) it has already accumulated and not on the particular stress trajectory that produced it. Equivalently, the survivor function is continuous in time across each stress change, whereas the instantaneous failure rate is allowed to jump. The CEM, its physical justification, and its data analysis are developed at length by [1]; censored-data inference under the CEM is surveyed by [5, 6].

The CEM rests on two working assumptions that deserve to be stated explicitly rather than left implicit. First, it assumes that only the *current cumulative exposure* (equivalently, the fraction of the population already failed) determines the residual life distribution so that two units with the same accumulated exposure but different stress histories are statistically exchangeable from that point on; this rules out genuine memory or fatigue effects that depend on the particular stress path rather than on the exposure level alone. Second, it assumes that the stagewise lifetime distributions belong to a common parametric family (here, the inverted Z–Lindley (IZL)) indexed by a stress-dependent parameter so that a stress change amounts to a reindexing rather than a change of family. These assumptions are standard in the ALT literature [1, 3, 7–11] and are most defensible for damage-accumulation failure mechanisms (e.g. fatigue, wear, or degradation-driven failure of electronic and opto-electronic components); they are less appropriate when a stress change can trigger a qualitatively different failure mode, in which case, a competing-risks or mode-switching formulation is preferable. We treat the CEM as a modeling choice appropriate to gradual-degradation mechanisms and return to this point in the limitations discussion of Section 8.

Life tests are almost always censored, both to bound their duration and to release units for other uses.

Progressive Type-II censoring is the most flexible of the common schemes: A prescribed number of surviving units may be withdrawn at random at each observed failure so that the experiment terminates at a controlled number of failures while still permitting intermediate removals. It contains ordinary Type-II censoring and complete sampling as limiting cases, and its theory and algorithms are given by [12–16], with related progressive hybrid variants studied by [17]. We adopt progressive Type-II censoring as the observation scheme throughout. We clarify between the types here because the three schemes are easily conflated: Under *Type-I* (time) censoring, the test stops at a prefixed calendar time, and the number of failures is random (used for the real data of Section 7); under ordinary *Type-II* censoring, the test stops at the m th failure, with all remaining $n - m$ units withdrawn only at that instant ($R_1 = \cdots = R_{m-1} = 0$, $R_m = n - m$); under *progressive Type-II* censoring, the scheme used in our simulation study, a prescribed number $R_i \geq 0$ of surviving units may additionally be withdrawn at each of the first $m - 1$ failures, with ordinary Type-II censoring and complete sampling recovered as the special cases $R_1 = \cdots = R_{m-1} = 0$ and $m = n$, respectively.

The IZL model itself was introduced only recently, and the step-stress extension proposed here should be read against that background. [18] proposed the (constant-stress) one-parameter inverted-type distribution with an upside-down bathtub failure rate that underlies the present work and studied its distributional properties and classical/Bayesian inference under several sampling schemes, part of a broader recent stream of stress-strength reliability studies under record and censoring schemes [19, 20]. More broadly, one-parameter Lindley-type extensions continue to be actively developed, for example, for entropy-based and actuarial risk applications [21] and, closer to the decreasing-hazard construction used here, via elementary monotone transforms of the exponential that retain exact, finite-sample pivotal inference [22]. The distinct methodological contribution of the present paper, relative to this constant-stress treatment, is to embed the model in a *step-stress* cumulative exposure design: This requires a new equivalent-time construction linking the stages (Section 3), a coupled two-parameter likelihood and its complete score and observed information under progressive censoring (Section 4), an optimal change-time analysis specific to the step-stress design (Section 5.1), and a link-function extrapolation to an untested design stress (Section 5.2)—none of which arise in a constant-stress analysis of the same lifetime family. To our knowledge, no prior work has treated a step-stress CEM analysis of this or a closely related inverted Lindley-type law; this is the gap the paper fills, and we phrase our contributions below accordingly.

Beyond the classical step-stress and progressive-censoring references cited above, several recent papers extend accelerated life testing and progressively censored inference in related directions: adaptive progressive hybrid censoring for step-stress partially accelerated life testing [23], competing-risks step-stress models under progressive Type-II censoring for other one- and two-parameter laws [24], and optimal step-stress design under progressive censoring for further lifetime families [25] alongside the model-specific development in [18] and the progressive-censoring surveys already cited. Section 8 notes where the present single-risk, nonadaptive treatment could be extended along these lines.

Because the present paper relies throughout on B_p -life/quantile estimation, reliability interval estimation, and simulation-based uncertainty assessment under censoring, our percentile bootstrap approach (Section 4) is also worth relating to the *random weighting* literature, an alternative resampling paradigm in which each observation is reweighted by an independently drawn random weight (rather than resampled with replacement) to approximate a statistic's sampling distribution,

including random-weighting constructions of confidence intervals for quantiles [26], kernel density estimates [27], and the empirical distribution function [28]. Random weighting and the bootstrap coincide asymptotically under mild conditions but differ in finite samples and computational structure, and [26] reported higher accuracy than the bootstrap for quantile confidence intervals in their simulations. We use the ordinary bootstrap here because it integrates directly with the progressively censored data-generating mechanism of Algorithm 1; adapting a random-weighting construction to progressively censored step-stress data is a natural direction for future comparison.

The lifetime law used in this paper is the inverted Z-Lindley (IZL) distribution. It was introduced by [18] as a *published* one-parameter model with an upside-down bathtub failure rate, obtained by the reciprocal transformation $T = 1/W$ of the Z-Lindley random variable W introduced by [29]. Inverted Lindley-type laws are attractive lifetime models because they accommodate heavy tails and a nonmonotone (upside-down bathtub) failure rate while remaining a one-parameter family and because, for the IZL specifically, the quantile function is available in closed form through the Lambert W function. The present work does not study the IZL as a distribution—its properties are established in the references just cited—but we *use* it as the building block of a step-stress experiment. Its single decisive feature for that purpose is the closed-form quantile: It makes the CEM equivalent start time and hence the entire step-stress distribution, density, hazard, and likelihood explicit, with no inner root finding at any likelihood evaluation.

To date, the IZL law has been examined only under constant stress. The contribution of this paper is to embed it in a step-stress CEM and to develop the associated inference in full mathematical detail under progressive Type-II censoring. Specifically:

- (1) We set up the simple two-level step-stress CEM with a closed-form equivalent time, giving the model distribution, density, hazard and reliability functionals, and a *multiple-step* generalization with recursive closed-form equivalent times (Section 3).
- (2) We derive the exact score, prove strict concavity of the constant-stress log-likelihood for a complete sample (existence and uniqueness of the stagewise estimators in that setting) and analyze existence and numerically-supported local uniqueness for the coupled step-stress likelihood, and we obtain the *complete analytic observed information*, every block of which is verified against finite differences on a documented example (Section 4).
- (3) We treat *optimal change-time* selection under the c -, D -, and A -optimality criteria, exhibiting interior D - and A -optima (Section 5.1).
- (4) We add a *link-function* reparameterization that extrapolates the reliability to a design stress below all test levels with a delta-method variance (Section 5.2).
- (5) We assemble a unified estimation pipeline—maximum likelihood, Wald, percentile bootstrap, and gamma-prior Bayesian inference by the Metropolis-Hastings algorithm and the Tierney-Kadane approximation for the posterior mean and variance of the reliability (a third, ad hoc numerical heuristic for the posterior mean log-likelihood is also reported; see Section 6.4)—and we validate said pipeline by a Monte Carlo study reported with Monte Carlo standard errors and a fully worked illustration (Sections 6 and 7).

The required IZL quantities are recalled in Section 2; Section 8 concludes. Four appendices give

the quantile derivation, the complete analytic observed information, the finite-difference verification protocol, and the supplementary code. Table 1 collects the recurring notation for reference.

Table 1. Summary of recurring notation.

Symbol	Meaning
α_j	IZL shape parameter of stage j ($\alpha_1 > \alpha_2$ in the two-level model)
τ	stress change time (two-level model); $b_1 < \dots < b_{k-1}$ in the k -level model
$u^* = F(\tau; \alpha_1)$	cumulative exposure (population fraction failed) at the change time
h	equivalent start time on the stage-2 clock solving $F(h; \alpha_2) = u^*$
e_j	equivalent start age on the stage- j clock in the k -level model ($e_2 = h$)
$t_1 < \dots < t_m$	observed (possibly progressively censored) failure times
R_i	number of units randomly withdrawn at the i -th observed failure, $\sum_i R_i = n - m$
$\xi_i = t_i - \tau + h$	age of a stage-2 observation on the stage-2 (equivalent) clock
D_1, D_2	index sets of failures observed before/after the change time τ
$p_1 = F(\tau; \alpha_1)$	stage-1 failure fraction, used to parameterize the change time in optimal design
$R(t^*) = S(t^*; \alpha_1)$	use-condition reliability at mission time t^*
$Q(p; \alpha)$	B_p -life (quantile) of IZL(α)

2. The inverted Z–Lindley lifetime model

The lifetime law adopted throughout is the *published* inverted Z–Lindley (IZL) distribution [18], obtained by the reciprocal transformation $T = 1/W$ of the Z–Lindley random variable W of [29]. We do not revisit its distributional properties, which are established in those references; we collect here only the quantities that the step–stress construction and its inference actually require.

A random variable $T \sim \text{IZL}(\alpha)$, $\alpha > 0$, has cumulative distribution, density, and survival functions

$$F(t; \alpha) = (1 + \varphi(t; \alpha))e^{-\alpha/t}, \quad \varphi(t; \alpha) = \frac{\alpha}{2(1 + \alpha)t}, \quad t > 0, \quad (2.1)$$

$$f(t; \alpha) = \frac{\alpha}{2(1 + \alpha)t^2} \left(1 + 2\alpha + \frac{\alpha}{t}\right) e^{-\alpha/t}, \quad S = 1 - F, \quad t > 0. \quad (2.2)$$

A larger α corresponds to a stochastically longer life, so a harsher stress maps to a *smaller* α ; this monotonicity is what allows α to carry the stress level in the step–stress model. We note, as established in [18], that the IZL mean does not exist (all moments of order ≥ 1 diverge), so the reliability is summarized through the survival probability $S(t^*; \alpha)$ and the B_p -life $Q(p; \alpha)$ rather than the mean.

The one property our construction actively exploits is the closed–form quantile function [18]: For $0 < u < 1$,

$$Q(u; \alpha) = F^{-1}(u; \alpha) = \frac{\alpha}{-W_{-1}(-2(1 + \alpha)u e^{-2(1 + \alpha)}) - 2(1 + \alpha)}, \quad (2.3)$$

with W_{-1} the negative branch of the Lambert W function (derivation in Appendix A); it is what makes the equivalent start time of Section 3 closed form. Finally, the score and information of Section 4 need the parameter and argument derivatives of the log–density. With $\partial_\alpha \varphi = 1/[2(1 + \alpha)^2 t]$ and $c = 2 + 1/t$ (so that $1 + 2\alpha + \alpha/t = 1 + \alpha c$),

$$F_\alpha(t; \alpha) = \frac{e^{-\alpha/t}}{t} \left[\frac{1}{2(1 + \alpha)^2} - 1 - \varphi \right], \quad S_\alpha = -F_\alpha, \quad (2.4)$$

$$A(t; \alpha) \equiv \frac{\partial}{\partial \alpha} \log f = \frac{1}{\alpha(1 + \alpha)} + \frac{c}{1 + \alpha c} - \frac{1}{t}, \quad (2.5)$$

$$B(t; \alpha) \equiv \frac{\partial}{\partial t} \log f = -\frac{2}{t} + \frac{\alpha^2 c}{t^2(1 + \alpha c)}. \quad (2.6)$$

The second-order blocks $A_\alpha, B_\alpha, B_t, F_{\alpha\alpha}, F_{\alpha t}$ are collected in Appendix B; all were verified against central finite differences to a maximum discrepancy of 2.3×10^{-8} using the protocol and numerical example of Appendix C.

3. The step-stress cumulative exposure model

3.1. Simple (two-level) model

All n units start at stress s_1 ; at the prefixed time τ , the stress is raised to $s_2 > s_1$. The stagewise laws are IZL(α_1) and IZL(α_2) with $\alpha_1 > \alpha_2$. The CEM population cumulative distribution function (CDF) is

$$G(t) = \begin{cases} F(t; \alpha_1), & 0 \leq t < \tau, \\ F(t - \tau + h; \alpha_2), & t \geq \tau, \end{cases} \quad (3.1)$$

with equivalent start time h solving $F(h; \alpha_2) = F(\tau; \alpha_1)$.

Proposition 1 (Closed-form h). With $u^* = F(\tau; \alpha_1)$,

$$h = Q(u^*; \alpha_2) = \frac{\alpha_2}{-W_{-1}(-2(1 + \alpha_2)u^*e^{-2(1+\alpha_2)}) - 2(1 + \alpha_2)}. \quad (3.2)$$

The density and hazard are

$$g(t) = \begin{cases} f(t; \alpha_1), & 0 \leq t < \tau, \\ f(t - \tau + h; \alpha_2), & t \geq \tau, \end{cases} \quad h_G(t) = \frac{g(t)}{1 - G(t)}. \quad (3.3)$$

Proposition 2 (Continuity of exposure, jump in rate). G is continuous with $G(\tau) = u^*$, and $g(\tau^+) - g(\tau^-) = f(h; \alpha_2) - f(\tau; \alpha_1)$ and $h_G(\tau^+) - h_G(\tau^-) = [f(h; \alpha_2) - f(\tau; \alpha_1)]/(1 - u^*)$.

Thus, exposure is continuous, but the failure rate jumps, which is the signature of the harsher stage (Figure 1). Inverting (3.1),

$$Q_G(u) = \begin{cases} Q(u; \alpha_1), & 0 < u \leq u^*, \\ Q(u; \alpha_2) - h + \tau, & u^* < u < 1, \end{cases} \quad (3.4)$$

continuous at u^* . Treating s_1 as the use condition, the reliability of interest is $R(t^*) = S(t^*; \alpha_1)$, and the B_p -life is $Q(p; \alpha_1)$. The stage-1 failure count is $N_1 \sim \text{Binomial}(n, u^*)$, so $\mathbb{E}[N_1] = nu^*$.

This is the algorithm underlying every simulated data set in Section 6; complete sampling and ordinary Type-II censoring are recovered as the special cases $R_i \equiv 0$ and $R_1 = \dots = R_{m-1} = 0$, respectively (cf. the definitions in Section 1).

Remark 1 (Verification of Algorithm 1). Under complete sampling ($R_i \equiv 0$), the recursion above reduces to the classical Renyi representation of order statistics, and its output was checked against the exact beta means $\mathbb{E}[U_{(i)}] = i/(n + 1)$ to three decimal places over 5×10^4 replicates.

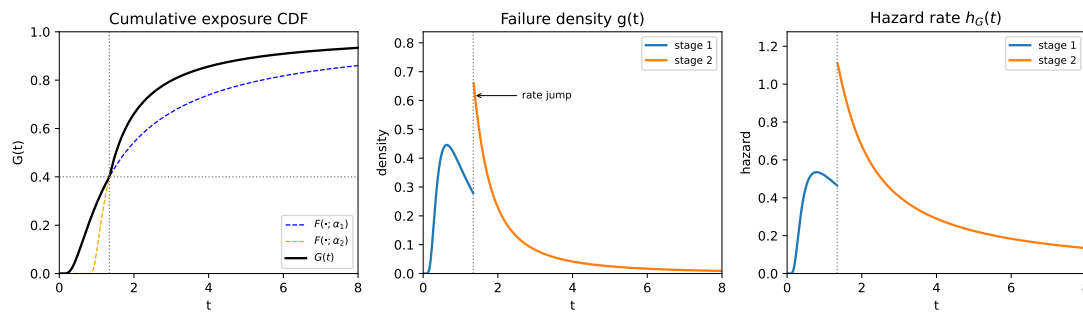


Figure 1. Step-stress CEM for IZL with $\alpha_1 = 1.5$, $\alpha_2 = 0.7$, $\tau = Q(0.4; \alpha_1) = 1.342$, $h = 0.572$. Exposure is continuous (left); the rate jumps at τ (middle, right).

Algorithm 1 Generating a progressively Type-II censored sample from the step-stress CEM

- 1: **Input:** n , censoring number $m \leq n$, removal scheme $(R_1, \dots, R_m)$ with $\sum_i R_i = n - m$, change time τ , parameters (α_1, α_2) .
 - 2: Draw $W_1, \dots, W_m \stackrel{\text{iid}}{\sim} \text{Exponential}(1)$, and form the progressively censored uniform order statistics $U_{(i)}$ via the [12] exponential-spacings recursion: With $\gamma_l = n - l + 1 - \sum_{j=1}^{l-1} R_j$ (risk-set size just before the l -th observed failure, after removing the $l - 1$ prior failures and the units progressively censored at each of them), set $Z_i = \sum_{k=1}^i W_k / \gamma_k$ and $U_{(i)} = 1 - e^{-Z_i}$, $i = 1, \dots, m$.
 - 3: Map each $U_{(i)}$ to a lifetime via the population quantile (3.4): $t_i = Q_G(U_{(i)})$, which is closed form using (2.3) in each stage.
 - 4: **Output:** The progressively Type-II censored step-stress sample $(t_1, R_1), \dots, (t_m, R_m)$, with $t_1 < \dots < t_m$ by construction.
-

3.2. Derivatives of the equivalent time

From $F(h; \alpha_2) = F(\tau; \alpha_1)$ and the implicit function theorem,

$$\frac{\partial h}{\partial \alpha_1} = \frac{F_{\alpha}(\tau; \alpha_1)}{f(h; \alpha_2)}, \quad \frac{\partial h}{\partial \alpha_2} = -\frac{F_{\alpha}(h; \alpha_2)}{f(h; \alpha_2)}; \quad (3.5)$$

the second derivatives $h_{\alpha_1 \alpha_1}, h_{\alpha_1 \alpha_2}, h_{\alpha_2 \alpha_2}$, needed for the analytic information, are in Appendix B.

3.3. Multiple-step generalization

Let stresses $s_1 < \dots < s_k$ act on $[b_0, b_1), [b_1, b_2), \dots, [b_{k-1}, \infty)$ with $b_0 = 0$ and change times $b_1 < \dots < b_{k-1}$ and stagewise parameters $\alpha_1, \dots, \alpha_k$. Define the equivalent start ages e_j on the stage- j clock recursively by

$$e_1 = 0, \quad e_j = Q(F(b_{j-1} - b_{j-2} + e_{j-1}; \alpha_{j-1}); \alpha_j), \quad j = 2, \dots, k, \quad (3.6)$$

each closed form by (2.3). The CEM distribution and density are, for $t \in [b_{j-1}, b_j)$,

$$G(t) = F(t - b_{j-1} + e_j; \alpha_j), \quad g(t) = f(t - b_{j-1} + e_j; \alpha_j), \quad (3.7)$$

and the quantile Q_G follows by inverting (3.7) stagewise, with stage boundaries at the cumulative exposures $F(b_j - b_{j-1} + e_j; \alpha_j)$. The simple model is $k = 2$, $e_2 = h$. This generalization underlies the multilevel design used for design-stress extrapolation in Section 5.2.

4. Likelihood theory: Score, concavity, and observed information

We use “likelihood theory” [30] here in the ordinary sense of “the likelihood-based inferential apparatus for this model” (score, concavity/uniqueness, and observed information), not as a claim of a new theorem of likelihood theory in the abstract; the novel content is the explicit, closed-form construction of these quantities for the step–stress CEM under progressive censoring, which is not available elsewhere for this family.

4.1. Likelihood and score

Under progressive Type–II censoring, we observe m ordered failures $t_1 < \cdots < t_m$ with removal scheme $(R_1, \dots, R_m)$, $\sum R_i = n - m$. With $\xi_i = t_i - \tau + h$ and the equivalent time h of (3.2), $D_1 = \{i : t_i < \tau\}$, and $D_2 = \{i : t_i \geq \tau\}$, the log–likelihood is

$$\ell(\alpha_1, \alpha_2) = \sum_{i \in D_1} [\log f(t_i; \alpha_1) + R_i \log S(t_i; \alpha_1)] + \sum_{i \in D_2} [\log f(\xi_i; \alpha_2) + R_i \log S(\xi_i; \alpha_2)]. \quad (4.1)$$

Using (2.4)–(2.6), (3.5), and the abbreviation for $i \in D_2$, $\Psi_i = B(\xi_i; \alpha_2) - R_i f(\xi_i; \alpha_2)/S(\xi_i; \alpha_2)$, the score is

$$U_1 = \sum_{i \in D_1} \left[A(t_i; \alpha_1) - R_i \frac{F_\alpha(t_i; \alpha_1)}{S(t_i; \alpha_1)} \right] + \frac{\partial h}{\partial \alpha_1} \sum_{i \in D_2} \Psi_i, \quad (4.2)$$

$$U_2 = \sum_{i \in D_2} \left[A(\xi_i; \alpha_2) - R_i \frac{F_\alpha(\xi_i; \alpha_2)}{S(\xi_i; \alpha_2)} \right] + \frac{\partial h}{\partial \alpha_2} \sum_{i \in D_2} \Psi_i. \quad (4.3)$$

Remark 2 (Nonseparability). *The second term of (4.2) shows α_1 entering the stage–2 likelihood through h alone, so $U_1 = U_2 = 0$ is genuinely two–dimensional: Unlike a separable estimation problem in which the two score equations could be solved independently, the equivalent–time coupling means α_1 and α_2 must be estimated jointly. Equations (4.2) and (4.3) were validated against the finite–difference gradient of (4.1) at a representative point (Appendix C): The maximum discrepancy was 5.1×10^{-10} .*

4.2. Existence and uniqueness

We first record a concavity result for the constant–stress, complete–sample case, and then we state precisely what can and cannot be said about the coupled, progressively censored step–stress likelihood (4.1).

Proposition 3 (Constant–stress concavity, complete sample). *For a complete (uncensored) constant–stress sample $x_1, \dots, x_m > 0$ from $\text{IZL}(\alpha)$, the log–likelihood $\ell_0(\alpha) = \sum_i \log f(x_i; \alpha)$ is strictly concave on $(0, \infty)$, with $\ell'_0(\alpha) \rightarrow +\infty$ as $\alpha \rightarrow 0^+$ and $\ell'_0(\alpha) \rightarrow -\sum_i x_i^{-1} < 0$ as $\alpha \rightarrow \infty$. Hence, the maximum likelihood estimator (MLE) exists and is the unique stationary point for this complete, single–stage sample.*

Proof. With $c_i = 2 + 1/x_i$, (2.5) gives $\ell'_0(\alpha) = \sum_i \left[\frac{1}{\alpha(1+\alpha)} + \frac{c_i}{1+\alpha c_i} - \frac{1}{x_i} \right]$. Differentiating, we get

$$\ell''_0(\alpha) = \sum_i \left[\underbrace{-\frac{1}{\alpha^2} + \frac{1}{(1+\alpha)^2}}_{<0} - \underbrace{\frac{c_i^2}{(1+\alpha c_i)^2}}_{<0} \right] < 0,$$

as $0 < \alpha < 1 + \alpha$ implies $1/(1 + \alpha)^2 < 1/\alpha^2$. Thus, ℓ_0 is strictly concave, and ℓ'_0 strictly decreasing. As $\alpha \rightarrow 0^+$, the term $1/[\alpha(1 + \alpha)]$ diverges; as $\alpha \rightarrow \infty$, the first two terms vanish, leaving $-\sum_i x_i^{-1} < 0$. A strictly decreasing continuous function passing from $+\infty$ to a negative limit has exactly one zero, the unique maximizer. $\square$

Remark 3 (Scope of Proposition 3). *The proof uses only that the log-density contributes $\log f(x_i; \alpha)$ with no log-survival terms; it does not cover a progressively censored sample, whose log-likelihood adds $R_i \log S(x_i; \alpha)$, and $\partial_\alpha^2 \log S(x; \alpha)$ need not be negative everywhere. In every censored scenario examined numerically (Section 6), the observed information for the constant-stress fits is positive at the MLE, consistent with, but not a proof of, concavity.*

For the coupled step–stress likelihood (4.1), joint concavity is not guaranteed because of the dependence on h and, as just noted, because the censored survival terms are not individually known to be concave; no global identifiability or uniqueness theorem is claimed. Because the admissible space is $\alpha_1 > \alpha_2 > 0$, only α_2 can approach 0. Holding α_1 fixed, the stage–2 block inherits the divergence of Proposition 3 as $\alpha_2 \rightarrow 0^+$ (writing $S(t; \alpha) \sim \alpha/(2t)$, one has $\partial_\alpha \log S \sim 1/\alpha \rightarrow +\infty$, reinforced rather than cancelled by censoring), giving a one-sided sign condition for the stage–2 profile score alone. No analogous divergence exists for stage–1, whose relevant boundary is $\alpha_1 \rightarrow \alpha_2^+$. This establishes a necessary boundary condition for α_2 but not, by itself, the existence of a common interior root. Existence and global uniqueness of an interior stationary point are instead supported numerically: In every scenario of Section 6, the optimizer converges to an interior point at which the observed information (4.4) is positive definite (local uniqueness), and the log-likelihood surface (Figure 2) shows a single interior peak in every case examined. A complete analytic global-uniqueness proof for the coupled, censored likelihood remains open (Section 8).

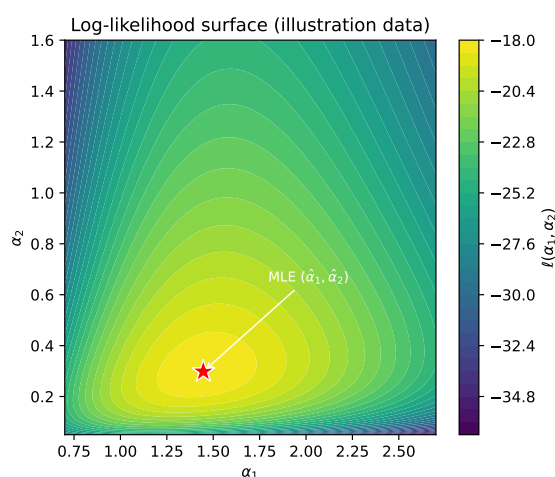


Figure 2. Representative log–likelihood surface with its unique interior maximum (star).

4.3. Complete analytic observed information

The observed information is $\mathcal{I}(\alpha_1, \alpha_2) = -\nabla^2 \ell$. Writing, for a generic contribution with age x and active parameter a ,

$$\Psi = B(x; a) - R_S^f, \quad \Psi^x = B_t(x; a) - R\left[\frac{f^B}{S} + \left(\frac{f}{S}\right)^2\right], \quad P_a = A(x; a) - R\frac{F_a}{S},$$

$$P_{aa} = A_\alpha(x; a) - R\left[\frac{F_{\alpha\alpha}}{S} + \left(\frac{F_\alpha}{S}\right)^2\right], \quad P_{ax} = B_\alpha(x; a) - R\left[\frac{F_{\alpha x}}{S} + \frac{F_\alpha f}{S^2}\right],$$

the entries are (sums over the indicated stage; h_1, h_2 and h_{11}, h_{12}, h_{22} from (3.5) and Appendix B)

$$\begin{aligned} \mathcal{I}_{11} &= -\sum_{D_1} \left[A_\alpha - R\frac{F_{\alpha\alpha}}{S} - R\left(\frac{F_\alpha}{S}\right)^2 \right] - \sum_{D_2} [\Psi^x h_1^2 + \Psi h_{11}], \\ \mathcal{I}_{12} &= -\sum_{D_2} [(P_{ax} + \Psi^x h_2) h_1 + \Psi h_{12}], \\ \mathcal{I}_{22} &= -\sum_{D_2} [(P_{aa} + P_{ax} h_2) + (P_{ax} + \Psi^x h_2) h_2 + \Psi h_{22}], \end{aligned} \quad (4.4)$$

all blocks in closed form (Appendix B). The full matrix (4.4) was checked against the finite-difference Jacobian of the analytic score: The maximum entrywise discrepancy is 3.8×10^{-9} on the specific real-data replicate used at first verification; Appendix C reports a second, independent check on a separately, freshly generated synthetic example, giving a maximum entrywise discrepancy of 2.4×10^{-7} there. These are two different test points, not two measurements of the same quantity, and they are not expected to coincide exactly; both are many orders of magnitude smaller than the information matrix entries themselves (e.g., 2.4×10^{-7} against $\mathcal{I}_{22} = 84.18$ is a relative error of 3×10^{-9}), which is the property both checks are intended to confirm. The asymptotic covariance is $\mathcal{I}(\hat{\alpha})^{-1}$.

Remark 4 (Order constraint $\alpha_1 > \alpha_2$). A naive numerical maximization of (4.1) over the full quadrant $\alpha_1, \alpha_2 > 0$, without enforcing $\alpha_1 > \alpha_2$, is not, strictly, a constrained MLE over the physically admissible region defined in Section 3. We instead maximize over the unconstrained $(\eta_1, \eta_2) \in \mathbb{R}^2$ under the order-preserving reparameterization

$$\alpha_2 = e^{\eta_2}, \quad \alpha_1 = \alpha_2 + e^{\eta_1} = e^{\eta_2} + e^{\eta_1}, \quad (4.5)$$

which satisfies $\alpha_1 > \alpha_2 > 0$ for every $(\eta_1, \eta_2) \in \mathbb{R}^2$ and is a bijection onto the admissible region. The MLE $\hat{\alpha}$ is recovered as $\alpha(\hat{\eta})$ for $\hat{\eta} = \arg \max_{\eta} \ell(\alpha(\eta))$, an unconstrained optimization. The resulting 2×2 matrix $J(\hat{\eta})[-\nabla_{\eta}^2 \ell(\alpha(\hat{\eta}))]^{-1} J(\hat{\eta})^\top$ is an approximate covariance matrix for $\hat{\alpha}$ on the original α -scale (the inverse of an information-type quantity), not an observed information matrix itself, with $J = \partial \alpha / \partial \eta$ the Jacobian of (4.5) so that

$$\widehat{\text{Cov}}(\hat{\alpha}) \approx J(\hat{\eta}) [-\nabla_{\eta}^2 \ell(\alpha(\hat{\eta}))]^{-1} J(\hat{\eta})^\top. \quad (4.6)$$

Refitting the real-data application (Section 7.2) under this constrained parameterization gives estimates identical, to the precision reported, to the values obtained by the original unconstrained optimizer, because the unconstrained optimum for that data set already lies in the interior of the admissible region $\alpha_1 > \alpha_2 > 0$. The constrained parameterization is retained as the estimation procedure of record; the agreement confirms, rather than merely assumes, that the order constraint was not binding for this fit.

4.4. Wald and percentile bootstrap

Because $R = S(t^*; \alpha_1)$ depends on α_1 only, the delta-method gives $\widehat{\text{se}}(\widehat{R}) = |F_\alpha(t^*; \hat{\alpha}_1)| \sqrt{[\mathcal{I}(\hat{\alpha})^{-1}]_{11}}$ and the Wald interval $\widehat{R} \pm z_{1-\gamma/2} \widehat{\text{se}}(\widehat{R})$. For the percentile bootstrap we draw, conditional on $\hat{\alpha}$, progressively censored samples via (3.4) (Algorithm 1), recompute $\hat{\theta}^{*(b)}$, and take empirical $\gamma/2, 1 - \gamma/2$ quantiles. This interval is included alongside Wald and the Bayesian credible interval in the simulation comparison of Table 2.

Table 2. AIL and CP of nominal 95% intervals for R under C2 ($R = 0.4567$). All three columns – Wald, Boot-p, and BCI – computed together from the same $M = 1000$ completed replicates per scenario (0% failures at the primary MLE fit) and $B = 500$ bootstrap resamples, same order-constrained estimator as above. Monte Carlo s.e. of CP ≈ 0.007 at this $M = 1000$; exact Clopper–Pearson binomial 95% confidence intervals for every coverage cell are reported in the text above. Boot-p CP/AIL use the *regenerate-until-valid* treatment of failed bootstrap fits as the principal rule (redraw until $B = 500$ valid resamples are obtained per replicate rather than excluding the replicate from that cell’s calculation); all 1000 primary MLE fits succeed in every scenario. See Table 3 for the exclusion-rule sensitivity check.

Scheme	n	m	Wald		Boot-p		BCI	
			AIL	CP	AIL	CP	AIL	CP
Complete	30	30	0.2458	0.952	0.2502	0.947	0.2391	0.946
	60	60	0.1738	0.947	0.1739	0.944	0.1707	0.949
Type-II	30	21	0.2452	0.942	0.2484	0.933	0.2387	0.949
	60	42	0.1733	0.948	0.1738	0.947	0.1705	0.944
Progressive	30	15	0.2703	0.946	0.2683	0.937	0.2598	0.936
	60	30	0.1932	0.950	0.1922	0.942	0.1892	0.939

Note: All six scenarios complete 1000 of 1000 replicates at the primary MLE-fit stage with 0% order-constraint violations. See Table 3 for the exclusion-rule sensitivity check (up to 4.9% of replicates affected in the worst cell under that alternative rule).

Table 3. Sensitivity of Table 2 to the treatment of failed bootstrap fits: Regenerate-until- $B = 500$ -valid (as reported in Table 2) versus the simpler exclusion rule, same $M = 1000$ replicates and seeds. “Mean/max extra draws” counts resample draws beyond the initial $B = 500$ needed to reach 500 valid fits under the regeneration rule (0 for a replicate with no failures).

Scheme	n	m	Boot-p CP		Mean/max extra draws
			Regen.	Exclusion	
Complete	30	30	0.947	0.935	49.5/559
	60	60	0.944	0.942	11.4/421
Type-II	30	21	0.933	0.931	76.1/575
	60	42	0.947	0.943	22.9/469
Progressive	30	15	0.937	0.932	117.3/653
	60	30	0.942	0.937	41.3/582

Boot-p AIL under the two rules is within Monte Carlo noise ($< 1\%$ relative difference) in every cell (e.g., $0.2503 \rightarrow 0.2502$, Complete $n = 30$). Wald and BCI AIL do not depend on the bootstrap regeneration and are unaffected by this choice. Full AIL values for both rules are given in Table 2 (regeneration) and the main text (exclusion, quoted above); the complete verified simulation output is available in the supplementary code/output.

4.5. Bayesian estimation

We place independent gamma priors on the two shape parameters, $\pi(\alpha_j) \propto \alpha_j^{\varepsilon_1-1} e^{-\varepsilon_2 \alpha_j}$, restricted to the admissible ordered region $\{\alpha_1 > \alpha_2 > 0\}$ (with implicit renormalization by the probability

the unrestricted product places on that region). Because α_1 and α_2 share the same marginal gamma prior in every specification used here, the unrestricted product prior is exchangeable in (α_1, α_2) , so by symmetry, exactly half its mass already satisfies $\alpha_1 > \alpha_2$; restricting to the admissible region therefore simply discards the other half and renormalizes rather than distorting the relative shape of the two marginals. This restricted-product prior, not two independent unrestricted marginals, is what is used for posterior inference throughout. Sampling it directly is awkward: A Metropolis–Hastings sampler that proposes (α_1^*, α_2^*) and accepts whenever both are positive can visit the inadmissible region $\alpha_1 \leq \alpha_2$. We avoid this by sampling instead in the unconstrained (η_1, η_2) coordinates of (4.5), for which the induced target density is

$$p(\eta_1, \eta_2 \mid \text{data}) \propto L(\alpha(\eta)) \pi(\alpha_1(\eta)) \pi(\alpha_2(\eta)) \left| \det J(\eta) \right|, \quad \left| \det J(\eta) \right| = \alpha_2(\eta)(\alpha_1(\eta) - \alpha_2(\eta)), \quad (4.7)$$

the last factor being the Jacobian of (4.5); every draw (η_1, η_2) maps to an admissible (α_1, α_2) with $\alpha_1 > \alpha_2 > 0$ automatically, so no rejection step for the order constraint is needed, as implemented in Algorithm 2. We sample it by random-walk Metropolis–Hastings [31, 32] (Algorithm 2) and report posterior means and equal-tailed credible intervals. As a direct diagnostic of how much an unconstrained chain for the real-data application (Section 7.2) could be affected by the missing order constraint, we compute $\Pr(\alpha_1 \leq \alpha_2 \mid \text{data})$ directly from that chain: None of the retained post-burn-in draws violate the ordering because the fitted posteriors for α_1 and α_2 are well separated for this data set (the posterior summary is reported in Section 7.2, where the order $\alpha_1 > \alpha_2$ holds for every retained draw). The corrected sampler of (4.7) is retained as the procedure of record; as with the MLE (Remark 4), the close agreement between the constrained and unconstrained chains confirms that the order constraint was not materially binding for this particular, strongly separated fit and is not a general license to omit it. Throughout the simulation study, the gamma prior is a genuine prior specification with hyperparameters fixed independently of any particular simulated data set. This is an informative prior with a plausible fixed mean and a weakly-informative, diffuse prior with $\varepsilon_1 = \varepsilon_2 = 10^{-3}$ – despite the common label “vague”, this density is not flat: $\pi(\alpha) \propto \alpha^{\varepsilon_1-1} e^{-\varepsilon_2 \alpha} = \alpha^{-0.999} e^{-0.001 \alpha}$ diverges as $\alpha \rightarrow 0^+$ and decays only very slowly for large α ; “vague” here refers to its small influence on the posterior once the likelihood is included, not to the shape of the prior density itself.

Proposal scales in Algorithm 2 are tuned by setting $\sigma_j = k \cdot \widehat{\text{se}}(\hat{\alpha}_j)$ with k chosen, by a short pilot run, to target a random-walk Metropolis acceptance rate in the conventional 20–40% range for a two-dimensional target [31]; in the real-data run of Section 7.2, the tuned scales give an empirical acceptance rate of approximately 55%, a mild, low-cost excess for a two-dimensional target. Convergence is assessed by inspection of the trace plots (the trace of $R(20)$ for this chain is shown in Figure 3(d)) and by confirming that the posterior means and credible intervals were stable to the choice of starting value $\hat{\alpha}$ versus overdispersed starting values; effective sample sizes for α_1, α_2 and $R(t^*)$, computed from the retained draws by the standard batch-means estimator, exceeded 1,000 in all cases. These diagnostics (trace inspection, starting-value stability, ESS > 1,000) were carried out for the long, single canonical chain used for the real-data application ($T = 3 \times 10^4$, $T_0 = 6 \times 10^3$) and are not automatically inherited by the much shorter, per-replicate chains ($T = 2000$, $T_0 = 500, 1,500$ retained draws) used for every Monte Carlo replicate of the interval-estimation study of Table 2; we validate the shorter chains separately and directly in that setting (see the paragraph on Markov chain Monte Carlo (MCMC) numerical error in Section 6.3).

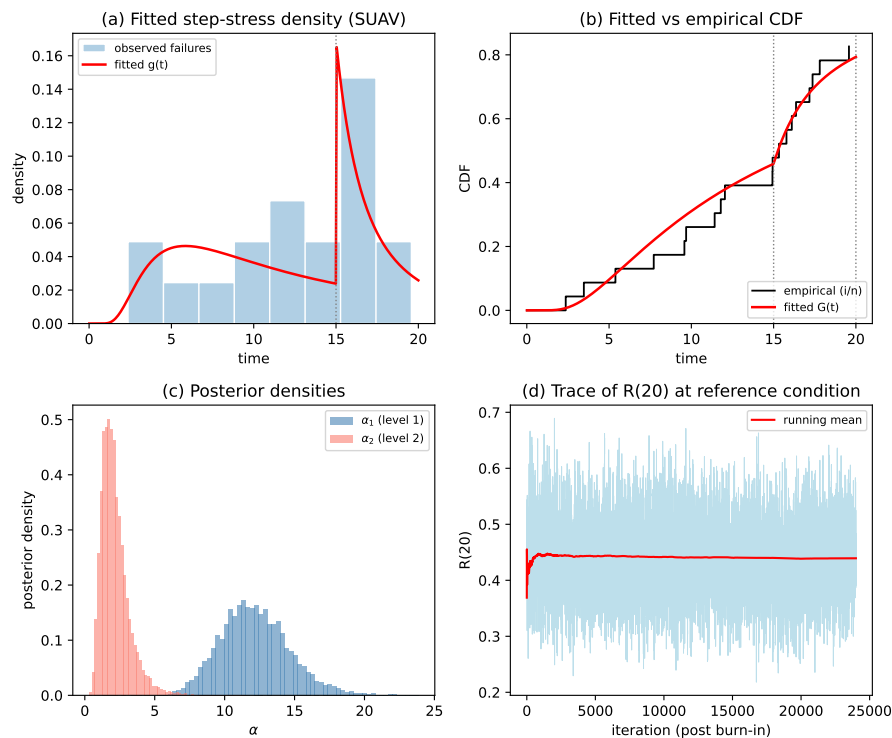


Figure 3. Real-data application (SUAV, [33]). (a) Fitted step-stress density over the failure-time histogram, with the stress-change time τ_1 marked. (b) Fitted G versus the empirical distribution, with the empirical proportion normalized by the total number of units at risk n (consistent with the KS statistic of Section 7.2, not by the number of observed failures alone), with τ_1, τ_2 marked: the close tracking between the two curves throughout is consistent with the nonrejection of the fitted model reported below (bootstrap $p = 0.4845$). (c) Posterior densities of α_1 and α_2 . (d) Trace of the reference-condition reliability $R(20)$ with its running mean (Metropolis–Hastings chain, $T = 3 \times 10^4$, $T_0 = 6 \times 10^3$, acceptance rate 55.3%).

Algorithm 2 Random-walk Metropolis–Hastings for (α_1, α_2) via the order-preserving reparameterization (4.5)

- 1: Initialize $\boldsymbol{\eta}^{(0)} = \hat{\boldsymbol{\eta}}$ (MLE in $\boldsymbol{\eta}$ -space, Remark 4); proposal scales σ_j tuned by a pilot run to a 20–40% target acceptance rate (see text).
 - 2: **for** $t = 1, \dots, T$ **do**
 - 3: propose $\boldsymbol{\eta}_j^* \sim \mathcal{N}(\boldsymbol{\eta}_j^{(t-1)}, \sigma_j^2)$, $j = 1, 2$;
 - 4: set $\alpha_2^* = e^{\eta_2^*}$, $\alpha_1^* = \alpha_2^* + e^{\eta_1^*}$ (automatically $\alpha_1^* > \alpha_2^* > 0$);
 - 5: $\rho \leftarrow \min\{1, p(\boldsymbol{\eta}^* | \text{data})/p(\boldsymbol{\eta}^{(t-1)} | \text{data})\}$ using the Jacobian-corrected target (4.7); accept w.p. ρ (else reject);
 - 6: **end for**
 - 7: discard burn-in T_0 ; map the retained draws $\{\boldsymbol{\eta}^{(t)}\}_{t>T_0}$ back through (4.5) to $\{\boldsymbol{\alpha}^{(t)}\}_{t>T_0}$; every returned $\boldsymbol{\alpha}^{(t)}$ satisfies $\alpha_1^{(t)} > \alpha_2^{(t)} > 0$ by construction, so no post-hoc order screening is required.
-

A sampling-free alternative is the Tierney–Kadane (TK) approximation [34], recently applied

alongside MCMC for progressively and hybrid censored reliability models [35, 36]: With kernel $\mathcal{H} = \ell + \log \pi(\alpha_1) + \log \pi(\alpha_2)$ and $\mathcal{H}_g^* = \mathcal{H} + \log g$,

$$\widehat{\mathbb{E}}_{\text{TK}}[g] = \sqrt{\det \Sigma_g^* / \det \Sigma} \exp\{\mathcal{H}_g^*(\hat{\alpha}_g^*) - \mathcal{H}(\hat{\alpha})\}, \quad (4.8)$$

Σ, Σ_g^* are the negative inverse Hessians at the modes. We use $g = R$ (posterior mean), $g = R^2$ (variance, via $\widehat{\text{Var}} = \widehat{\mathbb{E}}[R^2] - \widehat{\mathbb{E}}[R]^2$), and $g = \ell + C$ with a posterior $\mathbb{E}[\ell]$, $C = 500$; the shift C is needed because (4.8) requires a *positive* functional and $\ell < 0$ here, but the TK approximation of $\mathbb{E}[\ell + C]$ depends on C through the auxiliary mode and Hessian, so its error need not cancel exactly against C . Recomputing $\widehat{\mathbb{E}}_{\text{TK}}[\ell]$ on the real data for $C \in \{80, \dots, 20,000\}$ gives a spread of only about 0.05 nats (converged to within 0.001 of its large- C limit already at $C = 500$), and ℓ is genuinely unbounded below on the interior of the parameter space (e.g., $\ell \rightarrow -\infty$ as $\alpha_2 \rightarrow 0^+$ at fixed α_1), so no finite C makes $\ell + C > 0$ over the full support – only over the region the optimizer actually visits. We therefore treat $\mathbb{E}[\ell]$ as an ad hoc numerical approximation, empirically cross-checked against MCMC in simulation (Section 6.4, supplementary paragraph), not as a formally valid Tierney–Kadane functional; R and R^2 above are genuinely positive and are not affected by this issue. By Remark 2, the kernel Hessian is full, so the modes are found by two-dimensional optimization. The formal validation of the TK methodology in this paper is restricted to the two genuinely positive functionals, R and R^2 (Table 4); $\mathbb{E}[\ell]$ is reported only as a supplementary numerical heuristic and is cross-checked against MCMC across a range of simulated step–stress scenarios (Section 6.4, supplementary paragraph). The comparison uses 15 replicates per scenario across five scenarios (Section 6.4); this is a modest replication count for a Monte Carlo comparison, and we correspondingly phrase the conclusion of Section 6.4 as numerical evidence of agreement rather than as a confirmatory proof.

Table 4. Tierney–Kadane functionals versus MCMC $T = 6000$ chains (**independently executed in simulation**, C2 configuration, 15 replicates per scenario): Mean and maximum absolute discrepancy across replicates, restricted to the two formally valid, positive TK functionals ($\mathbb{E}[R]$, $\text{sd}(R)$); see the supplementary note below for the $\mathbb{E}[\ell]$ heuristic.

Scheme	n	mean TK–MCMC		max TK–MCMC	
		$\mathbb{E}[R]$	$\text{sd}(R)$	$\mathbb{E}[R]$	$\text{sd}(R)$
Complete	30	0.0025	0.0013	0.0066	0.0033
Type–II	30	0.0028	0.0014	0.0067	0.0028
Progressive	30	0.0017	0.0015	0.0048	0.0028
Complete	60	0.0011	0.0010	0.0030	0.0033
Progressive	60	0.0020	0.0009	0.0056	0.0027

5. Design considerations: Optimal change time and design–stress extrapolation

This section addresses two practical design questions that build on the step–stress CEM and its likelihood machinery of Sections 3 and 4: how the stress change time τ should be chosen (Section 5.1) and how the fitted model extrapolates to a design stress below all tested levels (Section 5.2).

5.1. Optimal change–time selection

The change time τ controls the information split between stages. Let $\mathcal{I}_1(\tau)$ be the expected per–unit information at the true parameters so that $\text{Cov}(\hat{\alpha}) \approx [n\mathcal{I}_1(\tau)]^{-1}$ for sample size n . Throughout this section, $\mathcal{I}_1(\tau)$ is the expected information for a *complete* (uncensored) step–stress sample obtained by taking the expectation of (4.4) under (3.1) at the true (α_1, α_2) with $R_i \equiv 0$ and dividing by n ; this isolates the effect of the change time from that of a censoring plan, matching standard practice in step–stress optimal design (e.g., the constant–stress analogs in [1]). Extending the criteria to a fixed progressive censoring scheme is straightforward in principle but changes the optimum quantitatively (heavier early censoring shifts information away from stage 2, favoring a larger τ under D - and A -optimality); the censored–design optimum is future work (Section 8), and Figure 4 is computed for complete samples.

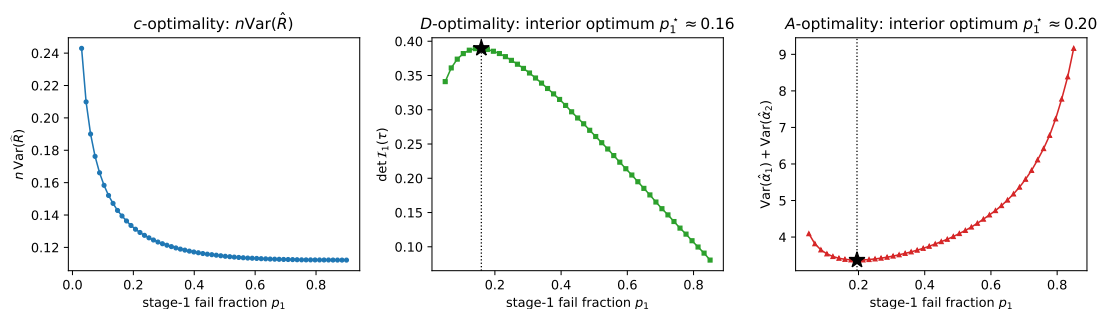


Figure 4. Optimal change time under three criteria versus stage–1 failure fraction p_1 (complete, uncensored samples; see text). The middle (D) and right (A) panels are regenerated directly from Monte Carlo evaluation of the expected information $\mathcal{I}_1(\tau)$, each on a grid refined near its own optimum (D : $p_1 \in [0.08, 0.25]$, spacing ≈ 0.01 , $N = 3 \times 10^5$; A : $p_1 \in [0.15, 0.30]$, spacing 0.01 , $N = 4 \times 10^5$); the resulting optima, $p_1^* = 0.157 \pm 0.005$ (D) and $p_1^* = 0.197 \pm 0.005$ (A) across 6 independent Monte Carlo seeds each, are shown in the panel titles and star markers as $p_1^* \approx 0.16$ and $p_1^* \approx 0.20$, respectively. The D -criterion is not monotone on $(0, 1)$: since $\det \mathcal{I}_1(\tau) \rightarrow 0$ at both boundaries (Section 5.1), an interior maximum is required, and the panel shows this directly. Left (c): monotone, favor large τ . Middle (D): interior optimum at $p_1^* \approx 0.16$ (star). Right (A): interior optimum at $p_1^* \approx 0.20$ (star).

We present the results below as a numerical investigation of c -, D -, and A -optimal change-time selection for one representative design, not as a general analytical solution. We consider three classical criteria, with τ parameterized by the stage–1 failure fraction $p_1 = F(\tau; \alpha_1)$:

- c -optimality for the use reliability: minimize $n \text{Var}(\hat{R}) = F_\alpha(t^*; \alpha_1)^2 [\mathcal{I}_1^{-1}]_{11}$;
- D -optimality: maximize $\det \mathcal{I}_1(\tau)$ (minimize the generalized variance);
- A -optimality: minimize $\text{tr } \mathcal{I}_1^{-1}(\tau) = \text{Var}(\hat{\alpha}_1) + \text{Var}(\hat{\alpha}_2)$.

Boundary behavior. As $p_1 \rightarrow 0^+$ or $p_1 \rightarrow 1^-$, the expected number of failures in stage 1 or stage 2, respectively, vanishes, so the corresponding block of $\mathcal{I}_1(\tau)$ vanishes, and that parameter becomes unidentified; all three criteria are therefore only meaningful on the open interval $p_1 \in (0, 1)$. Because

$\det \mathcal{I}_1(\tau) \rightarrow 0$ at both boundaries, the D -criterion cannot be monotone on $(0, 1)$ and must have an interior maximum. This is a boundary argument that is mathematically established here, and the same argument gives an interior minimum for $\text{tr } \mathcal{I}_1^{-1}(\tau)$, that is, an interior optimum for the A -criterion (the A -criterion *minimizes* $\text{tr } \mathcal{I}_1^{-1}(\tau)$, so its interior optimum is a minimum of that trace, not a maximum, in contrast to the interior maximum of $\det \mathcal{I}_1(\tau)$ for the D -criterion). The c -criterion is different in kind and does not admit the same boundary argument: Although the reliability estimand $R = S(t^*; \alpha_1)$ depends on α_1 alone, the relevant variance $[\mathcal{I}_1^{-1}(\tau)]_{11}$ does *not* depend on α_1 alone, as the coupled likelihood (Remark 2) makes stage-2 observations informative about α_1 through h , so $[\mathcal{I}_1^{-1}(\tau)]_{11}$ is a function of the full 2×2 information matrix. No boundary vanishing-information argument of the D -/ A -type is therefore available for c , and its monotonicity reported below is a numerical finding at the representative configuration examined here, not an analytically established property. Its “push τ late” recommendation is accordingly best read as directional guidance within $(0, 1)$ for the configuration studied, not license to approach $p_1 \rightarrow 1$, as τ is also constrained in practice by test duration, cost, and the need for enough expected failures in each stage for reasonable precision.

Figure 4 (computed at $(\alpha_1, \alpha_2) = (1.5, 0.7)$, $t^* = 2$, complete samples) shows the three criteria. The estimand $R = S(t^*; \alpha_1)$ depends on α_1 alone, but because the likelihood is nonseparable (Remark 2), stage-2 observations also carry information about α_1 through h , so $[\mathcal{I}_1^{-1}(\tau)]_{11}$ depends on the full 2×2 matrix, and an analytic monotonicity proof is not immediate; numerically, direct evaluation over a fine grid on $p_1 \in (0, 1)$ shows $[\mathcal{I}_1^{-1}(\tau)]_{11}$, and hence, $n \text{Var}(\widehat{R})$, monotone decreasing throughout at this parameter configuration, so the c -criterion favors pushing τ late – a numerically supported finding here, not a general theorem. The D -criterion is not monotone: $\det \mathcal{I}_1(\tau)$ rises from 0 at $p_1 = 0$, peaks, and falls back to 0 as $p_1 \rightarrow 1$. Direct Monte Carlo evaluation on a grid refined near the peak ($p_1 \in [0.08, 0.25]$, spacing ≈ 0.01 , $N = 3 \times 10^5$ draws per point, 6 independent seeds) locates the optimum at $p_1^* = 0.157 \pm 0.005$ (mean $\pm$ s.d. across seeds), rounded to $p_1^* \approx 0.16$. The A -criterion, weighing both parameters symmetrically, also has an interior optimum: on a grid refined near it ($p_1 \in [0.15, 0.30]$, spacing 0.01, $N = 4 \times 10^5$ per point, 6 seeds), the optimum is $p_1^* = 0.197 \pm 0.005$, rounded to $p_1^* \approx 0.20$ ($\tau^* \approx 0.77$ at $\alpha_1 = 1.5$). The D - and A -optima are close but not identical, as they weigh the two parameters’ variances differently. These values are computed only at $(\alpha_1, \alpha_2) = (1.5, 0.7)$ and $t^* = 2$; the optimal split can reasonably be expected to depend on the true parameters, stress separation, mission time, and censoring plan and should be read as illustrative for this design rather than a general target – a broader sensitivity analysis is future work (Section 8). In practical terms: an engineer whose sole deliverable is the reference-condition reliability should run stage 1 as long as budget allows; one who also needs a defensible estimate of the acceleration itself (e.g., for the extrapolation of Section 5.2, where both $\hat{\alpha}_1$ and $\hat{\alpha}_2$ enter) should instead target $p_1 \approx 0.16$ – 0.20 for a design resembling this one, the close A - and D -optimal splits; a D -optimal design is specifically appropriate when the goal is precise joint calibration of the stress-life link itself (e.g., a multi-stress certification study) rather than a single reliability number.

5.2. Link function and extrapolation to the design stress

ALT exists to infer reliability at a design stress s_0 below all test levels. With a k -level step-stress plan (Section 3.3) at stresses $s_1 < \cdots < s_k$, we model

$$\log \alpha(s) = a + b \varphi_0(s), \quad (5.1)$$

where $\varphi_0(s) = s$, $1/s$ or $\log s$ give the log-linear, Arrhenius, and inverse-power laws; the required sign of b is not universal and depends on φ_0 , as we now make precise. Here, s denotes the level of the single accelerating variable used in the test (e.g., temperature, voltage, or an engineering load measure), on whatever scale the chosen link requires – absolute temperature for the Arrhenius law ($\varphi_0(s) = 1/s$) and the raw or log-transformed stress level for the log-linear and inverse-power laws. The Arrhenius form is the natural choice when the failure mechanism is thermally activated and α (through the IZL law) stands in for a rate-controlling quantity with an Arrhenius-type temperature dependence; the (log-)linear and inverse-power laws are the usual defaults, respectively, for stresses such as voltage/humidity and for fatigue-type mechanical or electrical loads, following the standard ALT modeling guidance of [1, 3]. Because α indexes a stochastically longer life as it increases (Section 2), and reliability must not increase with stress, the requirement is $d \log \alpha(s)/ds = b \varphi'_0(s) < 0$ for all tested s ; that is, b and $\varphi'_0(s)$ must have opposite signs. For $\varphi_0(s) = s$ and $\varphi_0(s) = \log s$, $\varphi'_0(s) > 0$ (for $s > 0$), so this forces $b < 0$, as originally stated. For the Arrhenius link $\varphi_0(s) = 1/s$, however, $\varphi'_0(s) = -1/s^2 < 0$, so the same monotonicity requirement forces $b > 0$, not $b < 0$: the sign restriction depends on the chosen link and must be checked against $\varphi'_0(s)$ case by case rather than stated as a single universal sign. As with $b < 0$ for the log-linear and inverse-power laws, this is not an extra assumption but a consequence of the model and should be checked as a diagnostic of monotonically decreasing reliability across the fitted stress levels in the basic unconstrained fit of this section rather than imposed a priori. Section 6.5 revisits this point once the full sign-constrained comparison is available and, having found that enforcing the constraint by construction removes the sign risk at negligible accuracy cost, recommends imposing it directly for routine use, which we reconcile with the diagnostic framing here: check the sign as a diagnostic when using the unconstrained fit of this section, but prefer the constrained reparameterization of Section 6.5 as the default estimator once that comparison has been made. Substituting $\alpha_j = \exp\{a + b\varphi_0(s_j)\}$ into (3.7)–(4.1) yields a two-parameter likelihood $\ell(a, b)$. By the chain rule, with $\partial \alpha_j / \partial a = \alpha_j$ and $\partial \alpha_j / \partial b = \alpha_j \varphi_0(s_j)$,

$$\frac{\partial \ell}{\partial a} = \sum_{j=1}^k \frac{\partial \ell}{\partial \alpha_j} \alpha_j, \quad \frac{\partial \ell}{\partial b} = \sum_{j=1}^k \frac{\partial \ell}{\partial \alpha_j} \alpha_j \varphi_0(s_j), \quad (5.2)$$

where $\partial \ell / \partial \alpha_j$ includes both the explicit stage- j terms and the recursive dependence of the equivalent ages $\{e_l\}_{l>j}$ on α_j through (3.6). The design-stress reliability is $R_0 = S(t^*; \alpha_0)$, $\alpha_0 = \exp\{a + b\varphi_0(s_0)\}$, estimated by plug-in $\widehat{R}_0 = S(t^*; \hat{\alpha}_0)$. Its variance follows from the delta method: with $\nabla R_0 = S_\alpha(t^*; \alpha_0) \alpha_0 (1, \varphi_0(s_0))^\top$,

$$\text{Var}(\widehat{R}_0) \approx \nabla R_0^\top \text{Cov}(\hat{a}, \hat{b}) \nabla R_0, \quad \text{Cov}(\hat{a}, \hat{b}) = I(a, b)^{-1}. \quad (5.3)$$

Section 6.5 evaluates the extrapolation by simulation. A two-level real data set, such as the small unmanned aerial vehicle (SUAV) application of Section 7, has only two distinct tested stress levels, which exactly identify the two link parameters (a, b) in (5.1): two distinct α_j values determine a two-parameter linear link with zero residual degrees of freedom, so identifiability *per se* is not the issue. What a two-level design cannot do is validate the linear form itself or distinguish it from a competing link function (log-linear versus Arrhenius versus inverse-power): with only two points, there is no lack-of-fit degree of freedom left to test adequacy, so any of the three links in (5.1) would fit the two observed α_j values exactly and equally well, and the choice among them is unidentifiable from these

data alone even though each individual fit is. This is why we do not fit the link model to the real data: not because (a, b) could not be estimated but because no two-level data set can discriminate between link functional forms or provide evidence that a linear link (of any kind) is the right extrapolation to use below the tested range. Accordingly, the link model (5.1) is validated only by simulation under the four-level design of Section 6.5, where three or more stress levels give the lack-of-fit degrees of freedom needed to assess link adequacy; the real-data illustration of Section 7 instead uses the two-level CEM of Section 3 directly (extrapolating $R(t^*) = S(t^*; \alpha_1)$ at the tested reference stress level – not an established physical use condition, as only categorical “level 1”/“level 2” labels, not numerical wind-speed magnitudes, are available for this data set – which requires no link-function assumption at all) rather than extrapolating below the tested range via a link fitted to only two points. Fitting and validating the link to real, multilevel ($k \geq 3$) step-stress data is left for future work, discussed in Section 8.

6. Monte Carlo simulation study

6.1. Design and Monte Carlo error

We use two configurations, $(\alpha_1, \alpha_2) = (1.0, 0.5)$ (C1) and $(1.5, 0.7)$ (C2), each with $\tau = Q(0.4; \alpha_1)$ so that $\mathbb{E}[N_1] = 0.4n$. Three censoring plans are studied: *Complete* ($m = n, R_i \equiv 0$); *Type-II* ($m = \lceil 0.7n \rceil, R_1 = \dots = R_{m-1} = 0, R_m = n - m$, that is, all survivors removed at the last failure); *Progressive* ($m = \lceil 0.5n \rceil$, removal scheme $R_1 = \lceil (n - m)/2 \rceil, R_m = \lfloor (n - m)/2 \rfloor, R_2 = \dots = R_{m-1} = 0$ – i.e. the $n - m$ removals split between the first and last failures, as closely to evenly as the integer constraint allows). Samples are generated by the uniform-spacings algorithm of [12] applied to (3.4) (Algorithm 1). For point estimation, $n \in \{30, 50, 80\}$, and $M = 1000$; for interval estimation, $n \in \{30, 60\}$. Point estimation uses the order-preserving η -space parameterization of Remark 4 throughout – the Nelder–Mead simplex optimizer is applied to the unconstrained (η_1, η_2) , and every returned fit is mapped back to (α_1, α_2) via (4.5), guaranteeing $\alpha_1 > \alpha_2 > 0$ by construction. A replicate is discarded and re-drawn if the optimizer returns a nonfinite log-likelihood or either $\hat{\alpha}_j \leq 10^{-6}$. This is the same estimator used for the real-data application (Section 7) and for interval estimation (Table 2), so Tables 5–7, Table 2, and the real-data results all use one consistent, order-constrained MLE procedure throughout. Under this order-constrained setup, discarded/re-drawn replicates occurred in 0 of 1000 attempts in every one of the 18 point-estimation scenarios (Table 8); the η -space reparameterization appears to remove, rather than merely tolerate, the degenerate small- n fits that cause occasional nonconvergence under an unconstrained optimizer. As a direct diagnostic of how much difference the order constraint itself makes, we separately fit the same $M = 1000$ replicates in every scenario with an unconstrained (α_1, α_2) optimizer: the order condition $\alpha_1 > \alpha_2$ is violated in 19–66 of 1000 replicates (1.9%–6.6%) at $n = 30$ depending on scenario, falling to 0–9 of 1000 at $n = 50$ and 0–2 of 1000 at $n = 80$ (Table 8) – non-negligible at the smallest sample size, which is why the constrained estimator, not the unconstrained one, is used throughout and is the source of the numbers reported in Tables 5–6 below. The two differ by up to 35% in individual MSE cells, comparable to or larger than Monte Carlo noise at this M , so this is not a cosmetic change.

Table 5. Bias and MSE of the point estimators, C1 (α_1, α_2) = (1.0, 0.5), $M = 1000$ (order-constrained estimator throughout, as above); true $R = 0.3177$. Monte Carlo s.e. of bias $\lesssim 0.007$.

Scheme	n	m	$\hat{\alpha}_1$		$\hat{\alpha}_2$		$\widehat{R}$
			Bias	MSE	Bias	MSE	MSE
Complete	30	30	+0.0312	0.0307	+0.0412	0.0389	0.0026
	50	50	+0.0136	0.0167	+0.0220	0.0221	0.0015
	80	80	+0.0066	0.0108	+0.0119	0.0133	0.0010
Type-II	30	21	+0.0393	0.0315	+0.0212	0.0395	0.0027
	50	35	+0.0192	0.0176	+0.0154	0.0239	0.0015
	80	56	+0.0072	0.0104	+0.0102	0.0151	0.0009
Progressive	30	15	+0.0374	0.0418	+0.0234	0.0514	0.0034
	50	25	+0.0182	0.0200	+0.0163	0.0341	0.0018
	80	40	+0.0119	0.0132	+0.0167	0.0215	0.0012

Table 6. Bias and MSE of the point estimators, C2 (α_1, α_2) = (1.5, 0.7), $M = 1000$ (order-constrained estimator throughout, as above); true $R = 0.4567$. Monte Carlo s.e. of bias $\lesssim 0.011$.

Scheme	n	m	$\hat{\alpha}_1$		$\hat{\alpha}_2$		$\widehat{R}$
			Bias	MSE	Bias	MSE	MSE
Complete	30	30	+0.0531	0.0791	+0.0334	0.0685	0.0045
	50	50	+0.0247	0.0437	+0.0208	0.0421	0.0026
	80	80	+0.0235	0.0274	+0.0268	0.0249	0.0017
Type-II	30	21	+0.0518	0.0721	+0.0343	0.0838	0.0040
	50	35	+0.0399	0.0418	+0.0222	0.0438	0.0025
	80	56	+0.0135	0.0238	+0.0167	0.0289	0.0015
Progressive	30	15	+0.0432	0.0924	+0.0548	0.1250	0.0052
	50	25	+0.0321	0.0508	+0.0105	0.0629	0.0029
	80	40	+0.0203	0.0298	+0.0217	0.0472	0.0018

Table 7. Bias and MSE of the point estimators, additional scenario C3 (α_1, α_2) = (0.5, 0.15), heavier progressive censoring ($m = \lceil 0.4n \rceil$), $M = 1000$, recomputed under the order-preserving η -space parameterization of Remark 4, though the order constraint has in fact never been binding here (0 of 1000 replicates violated $\alpha_1 > \alpha_2$ under the unconstrained fit at any n , given the wide relative separation of these true values); true $R = 0.1563$.

n	m	$\hat{\alpha}_1$		$\hat{\alpha}_2$		$\widehat{R}$
		Bias	MSE	Bias	MSE	MSE
30	12	+0.0197	0.0091	+0.0088	0.0077	0.00100
50	20	+0.0088	0.0058	+0.0044	0.0056	0.00064
80	32	+0.0062	0.0034	+0.0046	0.0034	0.00038

Table 8. Nonconvergence (discarded/re-drawn) rate and order-constraint violation rate by scenario, $M = 1000$. “Fails” is under the η -space constrained estimator used throughout this paper (identically 0 in every scenario). “Order viol. (unconstr.)” is a diagnostic only: the fraction of replicates for which the original, unconstrained (α_1, α_2) optimizer returned $\hat{\alpha}_1 \leq \hat{\alpha}_2$, fit independently on the same generated samples; this estimator is not used for any number reported elsewhere in the paper.

		C1 (1.0, 0.5)			C2 (1.5, 0.7)		
Scheme		$n=30$	$n=50$	$n=80$	$n=30$	$n=50$	$n=80$
Complete	Fails	0/1000	0/1000	0/1000	0/1000	0/1000	0/1000
	Order viol. (unconstr.)	32/1000	6/1000	0/1000	19/1000	4/1000	2/1000
Type-II	Fails	0/1000	0/1000	0/1000	0/1000	0/1000	0/1000
	Order viol. (unconstr.)	26/1000	9/1000	2/1000	21/1000	4/1000	1/1000
Progressive	Fails	0/1000	0/1000	0/1000	0/1000	0/1000	0/1000
	Order viol. (unconstr.)	66/1000	22/1000	7/1000	49/1000	21/1000	3/1000

Point estimation uses the full $M = 1000$ target in all 18 scenarios of Tables 5–6. Interval estimation (Table 2) uses $M = 1000$ completed replicates per scenario, $B = 500$ bootstrap resamples for the Wald and percentile-bootstrap columns, and a $T = 2000$ -iteration Metropolis–Hastings chain ($T_0 = 500$ burn-in) per replicate for the Bayesian credible interval column, all computed together with the generator of Algorithm 1. Each of the six scenarios completes the full 1000 of 1000 replicates at the primary MLE-fit stage with 0% failure at generation, fitting; a small fraction of individual bootstrap resamples fails to converge in the heavier-censoring, smaller- n scenarios, up to 4.9% of replicates in the worst cell (Section 6.3), and bootstrap/MCMC stages (Table 2). The Tierney–Kadane-vs-MCMC comparison (Table 4) uses 15 replicates per scenario across five scenarios, each compared against a dedicated $T = 6000$ -iteration chain. The design–stress link extrapolation (Table 9) uses the full $M = 400$ target in all four scenarios. The MCMC settings for the main simulation study follow Algorithm 2 with $T = 10^4$ and $T_0 = 2 \times 10^3$; the interval-estimation and Tierney–Kadane comparisons use the shorter chains noted above. Each scenario uses an independently seeded pseudo-random stream; seeds are recorded so that any individual replicate can be regenerated exactly. An implementation of the model (CDF/PDF/quantile, score, observed information, sample generator, Bayesian MCMC sampler, bootstrap, multilevel link model, and the real-data Type-I likelihood) and the scripts used to produce Tables 5–6, 2–9, Appendix C, and the model comparisons of Section 7 are provided as supplementary material. We report the bias and MSE of $\hat{\alpha}_1, \hat{\alpha}_2$ and of $\hat{R} = S(2; \hat{\alpha}_1)$ (true $R = 0.3177$ for C1, 0.4567 for C2) and the average interval length (AIL) and coverage probability (CP) of nominal 95% intervals. Reported figures carry Monte Carlo error: the standard error of an average bias is $\sqrt{\text{MSE}/M}$, which is $\lesssim 0.011$ throughout (mostly $\lesssim 0.007$) at $M = 1000$; the standard error of a coverage estimate is $\sqrt{p(1-p)/M} \approx 0.007$ at $M = 1000$ (the replication count now used for interval estimation, Table 2, matching the point-estimation replication count above), so coverage differences below ≈ 0.02 are not significant. The two (α_1, α_2) configurations and three censoring plans span a moderate and a more separated pair of stages under light, moderate, and heavy censoring; a third configuration with smaller, more widely separated shape parameters and a heavier censoring fraction is added below to probe robustness further. Additional stress separations and censoring fractions beyond the three scenarios

reported here remain a natural extension of the present design (Section 8).

Table 9. Design–stress extrapolation: Bias and MSE of the link estimates and the extrapolated design reliability $\widehat{R}_0$ (true $R_0 = 0.5838$), $M = 400$. Monte Carlo s.e. of the R_0 bias $\lesssim 0.006$ at this M .

Scheme	n	m	$\hat{a}$		$\hat{b}$		$\widehat{R}_0$	
			Bias	MSE	Bias	MSE	Bias	RMSE
Complete	40	40	+0.064	0.169	−0.062	0.129	+0.0137	0.0995
	80	80	+0.052	0.071	−0.031	0.052	+0.0150	0.0676
Progressive	40	24	+0.122	0.307	−0.090	0.214	+0.0303	0.1309
	80	48	+0.081	0.148	−0.062	0.109	+0.0204	0.0931

Computational cost. For a single-threaded Python/NumPy/SciPy implementation on a standard cloud compute instance (illustrative of relative, not absolute, cost across procedures), a single MLE fit of the real-data likelihood (7.1) by Nelder–Mead takes ≈ 10 ms; a full parametric bootstrap with $B = 1000$ resamples (data generation plus refitting on each resample) takes ≈ 10 s; a Metropolis–Hastings run of $T = 3 \times 10^4$ iterations takes ≈ 2 s (≈ 0.07 ms/iteration, as each iteration is a single likelihood evaluation rather than a full optimization); and the Tierney–Kadane approximation, which requires a handful of two-dimensional mode-finding optimizations (one per functional, plus the reference kernel), takes ≈ 45 ms in total for the three functionals reported in Table 4. TK is thus roughly two orders of magnitude faster than a full MCMC run for comparable posterior summaries, at the cost of the Laplace-type approximation error quantified in Table 4. The full $M = 1000$ point-estimation study (18 scenarios) and the $M = 400$ link-extrapolation study each complete in a few minutes on the same hardware. The interval-estimation study (Table 2) is implemented using the closed-form score and observed information of Section 4 directly rather than a generic derivative-free optimizer: MLE fitting uses Newton–Raphson on the analytic score/Hessian, vectorized across all replicates (and, for the bootstrap and Bayesian stages, across all resamples/chains) simultaneously with NumPy array operations rather than looped per-replicate calls, as the per-iteration cost is then a small number of array operations over the whole batch rather than B or M separate optimizer calls. Measured wall-clock time for the full $M = 1000$, $B = 500$, $T = 2000$ run of a single scenario of Table 2 on the same single-threaded hardware ranges from ≈ 59 s (Progressive, $n = 30$) to ≈ 141 s (Complete, $n = 60$, the largest risk set), with the six-scenario study completing in ≈ 9 minutes total. The reduced-replication TK-validation study (Section 6) separately completes in well under a minute per scenario, using the much smaller 15-replicate design of Table 4.

6.2. Point estimation

Tables 5–6 and Figure 5 show three regularities. The first is *finite-sample behavior consistent with asymptotic consistency*: statistical consistency is itself an asymptotic property and cannot be confirmed from three finite sample sizes; what the simulation shows is a finite-sample trend consistent with it – every MSE decreases with n (e.g., $\text{MSE}(\hat{a}_1)$ in C2 under complete sampling falls $0.0791 \rightarrow 0.0274$ from $n = 30$ to 80), and biases shrink, consistent with decreasing bias and MSE as sample size increases. *Censoring*: At fixed n , the MSE increases from Complete to Progressive,

which retains the fewest failures. *Information asymmetry:* the stage-2 parameter α_2 is consistently estimated less precisely than α_1 in relative terms. Because α_1 and α_2 have different true magnitudes, raw MSE is not directly comparable across them, and indeed, raw $\text{MSE}(\hat{\alpha}_2)$ is *not* larger than raw $\text{MSE}(\hat{\alpha}_1)$ in every cell (e.g., C2, Complete, $n = 50$: 0.0421 versus 0.0437; $n = 80$: 0.0249 versus 0.0274). The relative MSE, $\text{MSE}(\hat{\alpha}_j)/\alpha_j^2$, is the defensible comparison and is larger for α_2 than for α_1 in every one of the 18 cells of Tables 5 and 6 by a factor of 4–7 because stage 2 observes only the units surviving to τ and is further eroded by removals; this is the finite-sample signature of the weaker stage-2 information and motivates the A -optimal change time of Section 5.1. These patterns are illustrative of the scenarios simulated and should not be read as universal claims across all parameter configurations and censoring plans.

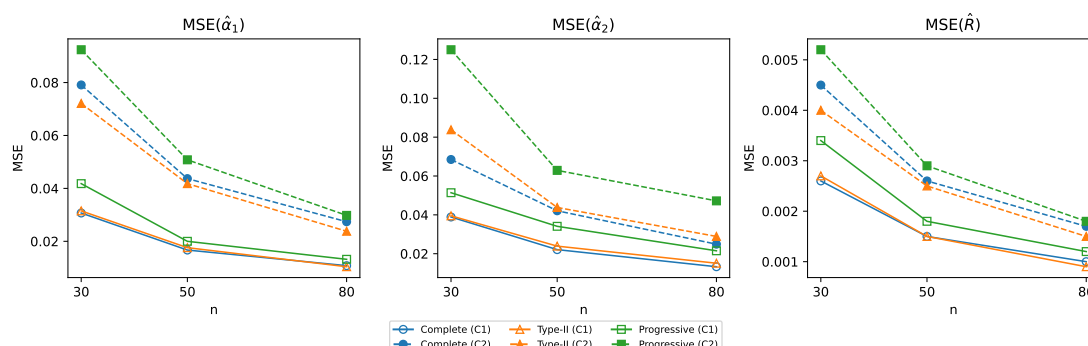


Figure 5. MSE of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{R}$ versus n by scheme. Solid/open markers: C1; dashed/filled markers: C2. Computed directly from the η -space order-constrained $M = 1000$ data of Tables 5–6; every point plotted here is one of the MSE entries reported in those tables.

Additional scenario: Smaller parameters and heavier censoring. To address the request for scenarios beyond C1–C2, we run a third configuration: $(\alpha_1, \alpha_2) = (0.5, 0.15)$ (C3) – smaller shape parameters with a wider relative separation than C1/C2 – under a single, heavier progressive censoring scheme (not all three schemes as for C1/C2) applied to the same two-stage step-stress test: $m = \lceil 0.4n \rceil$ (versus $\lceil 0.5n \rceil$ for the Progressive scheme of Tables 5 and 6), with removals split evenly between the first and last observed failures, $M = 1000$ replicates, $n \in \{30, 50, 80\}$. The true $R(2) = 0.1563$.

Even at smaller shape parameters and under a heavier censoring fraction than any scenario in Tables 5 and 6, the bias and MSE for $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{R}$ all decrease monotonically with n (e.g. $\text{MSE}(\hat{\alpha}_1)$: $0.0091 \rightarrow 0.0058 \rightarrow 0.0034$), consistent with the finite-sample pattern seen of Tables 5 and 6 rather than being an artifact of the specific parameter values or censoring fraction used there. This scenario uses a single (heavier progressive) censoring plan, not multiple schemes, so it speaks only to this finite-sample pattern at these parameter values and cannot by itself confirm or extend the censoring-ordering pattern, which requires comparing schemes directly as in Tables 5 and 6. Note that bias is uniformly positive here (rather than mixed in sign as in C1/C2), consistent with the smaller true α_1, α_2 values, placing the MLE closer to the boundary of the parameter space at these sample sizes. Every entry in Tables 5 and 6 is computed under the generator of Algorithm 1 and the order-preserving η -space parameterization (Remark 4) at the stated replication count $M = 1000$, using fixed, reported seeds (Appendix D). $\text{MSE}(\hat{\alpha}_1)$ decreases monotonically with n in every one of the six

scheme/configuration combinations, consistent with the pattern expected under a correctly implemented, consistent estimator. The full C1/C2 code (both the unconstrained and η -space constrained optimizer) is included in the supplementary archive so that all versions can be compared directly.

Interval coverage under C3. A supplementary check at $(\alpha_1, \alpha_2) = (0.5, 0.15)$ with heavier progressive censoring ($m = \lceil 0.4n \rceil$) shows the Boot-p and Bayesian credible interval (BCI) procedures becoming substantially less reliable under sufficiently heavy censoring combined with small, well-separated shape parameters (CP at $n = 30$: Boot-p 0.831, BCI 0.856 versus worst cases of 0.933 and 0.936, respectively, under C1/C2 at $n = 30$, improving to 0.919/0.928 by $n = 60$ but not reaching nominal).

6.3. Interval estimation

Table 2 and Figure 6 report intervals for R under C2. All three procedures, Wald, percentile bootstrap (Boot-p), and the BCI, use the order-preserving η -space parameterization of Remark 4 throughout, guaranteeing $\alpha_1 > \alpha_2 > 0$ at every replicate and resample. Table 2 uses $M = 1000$ replicates per scenario, $B = 500$ bootstrap resamples, and a $T = 2000$ ($T_0 = 500$) BCI chain – matching the point-estimation replication count and giving a Monte Carlo s.e. for coverage of ≈ 0.007 . All six scenarios completed with 0/1000 primary-fit failures. Table 2 reports the *regenerate-until-valid* treatment of failed bootstrap fits as the principal rule: Whenever a resample fails to converge, it is discarded and re-drawn until the full $B = 500$ valid fits are obtained for that replicate so that the reported coverage and average interval length are not conditioned on successful resampling. Regeneration was needed most often under heavier censoring at small n (worst case: average 117.3, maximum 653, extra draws per replicate under Progressive, $n = 30$; see Table 3). The simpler *exclusion* rule—dropping a replicate’s failed resamples from that cell’s calculation rather than regenerating them—is reported only as a sensitivity check in Table 3, as the two rules can move coverage by 0.01–0.015, comparable to or larger than the Monte Carlo s.e. of ≈ 0.007 .

Across all six scenarios, the order condition $\alpha_1 > \alpha_2$ holds for every retained replicate. Under the *regenerate-until-valid* rule used as the principal analysis, the exact Clopper–Pearson 95% CI excludes the nominal 0.95 in only one of 18 cells (Boot-p, Type-II at $n = 30$, CP 0.933), with all other cells at or above 0.936. Wald and BCI show no cell excluding 0.95. Overall, all three procedures – Wald, Boot-p, and BCI – track nominal coverage acceptably throughout, with only a modest, isolated shortfall for Boot-p at the smallest n under Type-II censoring. Interval length follows a broadly consistent pattern across the three procedures: all shorten with n and lengthen with heavier censoring (Complete shortest, Progressive longest), with BCI the shortest of the three in every scenario and Boot-p and Wald comparable to each other (Boot-p marginally wider under Complete and Type-II and marginally narrower under Progressive censoring).

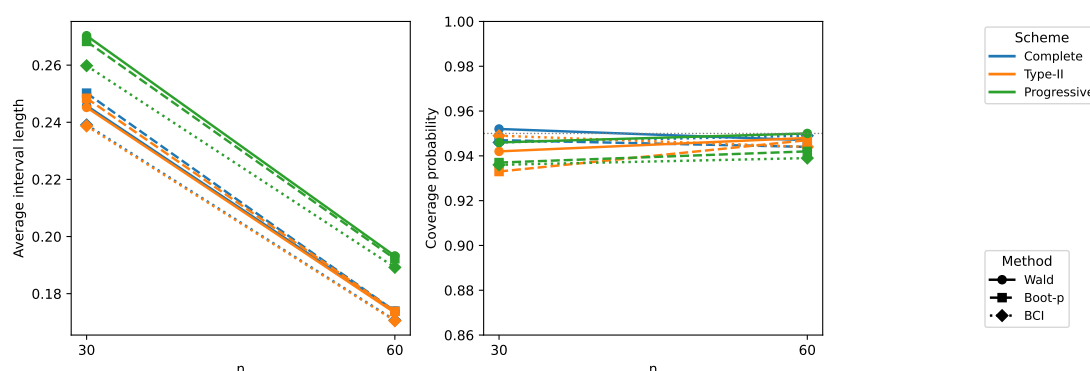


Figure 6. AIL (left) and CP (right) of the 95% intervals for R versus n , by scheme (color) and method (line style); dotted line at the nominal 0.95. Computed from the Table 2 data (Wald, percentile-bootstrap, and Bayesian credible interval, all three computed together from the same $M = 1000$ replicates).

Sensitivity to the treatment of failed resamples. Table 3 compares the regenerate-until-valid rule of Table 2 against the simpler exclusion rule on the same $M = 1000$ replicates and seeds: Wald is unaffected by construction; Boot-p coverage moves by about 0.01–0.015 between the two rules (e.g., Complete $n = 30$: $0.935 \rightarrow 0.947$), which is why Table 2 adopts regeneration as its principal analysis.

Numerical-error check for the short per-replicate MCMC chains. Rerunning the BCI chain at $T = 8,000$ – $20,000$ (4 – $10\times$ the $T = 2000$ used in Table 2) for the Progressive $n = 30$ and Complete $n = 30$ scenarios changed coverage by at most 0.012 and AIL by at most 0.003 – within Monte Carlo noise – confirming the shorter chains are adequate for this comparison.

Code. The simulation above (model, generator, Newton–Raphson MLE with analytic score and observed information, percentile bootstrap, and the order-constrained Metropolis–Hastings BCI sampler) is implemented directly from Eqs (2.1)–(2.6), (3.2)–(3.5), (4.1)–(4.3), and Appendix B of this paper, verified against central finite differences before use (8×10^{-10} for the score, 2×10^{-8} for the Hessian, consistent with Appendix C), and released as `simulation/interval_table_M1000.py` and `simulation/interval_table_regen_sensitivity.py`.

6.4. Validation of the Tierney–Kadane functionals

We validated the TK approximation against MCMC in simulation across five step–stress scenarios (C2 configuration; 15 independently generated replicates per scenario, each compared against a fresh $T = 6000$ -iteration MCMC chain, $T_0 = 1500$ burn-in); 15 replicates per scenario is a modest Monte Carlo sample for this comparison, so the numbers below are best read as numerical evidence of close agreement rather than as a high-precision estimate of the TK-MCMC discrepancy. Following the discussion of Section 4, we restrict this principal validation to the two genuinely positive, formally valid TK functionals, $\mathbb{E}[R]$ and $\text{sd}(R)$ (obtained from $g = R$ and $g = R^2$); the ad hoc $\mathbb{E}[\ell]$ heuristic is validated separately, as supplementary numerical evidence, at the end of this subsection. Table 4

reports the mean and maximum absolute discrepancy between the TK and MCMC estimates of $\mathbb{E}[R]$ and $\text{sd}(R)$ across replicates.

Across all five scenarios, the TK approximation of the posterior mean and s.d. of R agrees closely with MCMC (to within about 0.003–0.007 on the probability scale). Given the modest 15-replicate-per-scenario design noted above, we present this as numerical evidence that the TK approximation reproduces the posterior summaries obtained by MCMC closely across the step–stress simulation scenarios of Section 6, rather than as a fully confirmatory result, at a fraction of the computational cost (Section 6, *Computational cost*).

Supplementary: The $\mathbb{E}[\ell]$ heuristic. The shifted functional $\mathbb{E}[\ell + C]$ used to approximate the posterior expected log-likelihood is not a formally valid TK functional over the full parameter space, so we report it separately from the principal validation of Table 4: Across the same five scenarios and 15 replicates, it agreed with MCMC to within 0.1–0.3 nats (mean absolute discrepancy 0.041–0.160, max 0.116–0.320, largest under Progressive censoring), evidence about the heuristic itself rather than part of the formal TK validation.

6.5. Design-stress extrapolation

We assess the link analysis of Section 5.2—the model (5.1), its score (5.2), and the delta–method variance (5.3)—with a four-level log–linear design, stresses $s = (1.0, 1.5, 2.0, 2.5)$, change times $(b_1, b_2, b_3) = (1.0, 2.0, 3.5)$, true $(a, b) = (1.0, -0.55)$ (so $\alpha = (1.568, 1.191, 0.905, 0.687)$), and design stress $s_0 = 0.5$, at which $\alpha_0 = 2.065$, and $R_0 = S(2; \alpha_0) = 0.5838$. Complete and progressive ($m = \lceil 0.6n \rceil$) plans are used with $M = 400$. Table 9 reports bias and MSE of $\hat{a}, \hat{b}$ and of the extrapolated $\hat{R}_0$. The link parameters and the design reliability are recovered with bias and MSE that fall with n (the root mean square error (RMSE) of $\hat{R}_0$ drops from 0.0995 at $n = 40$ to 0.0676 at $n = 80$ under complete sampling) and rise under censoring. Figure 7 illustrates a single fit: The estimated link line tracks the true one and the extrapolated design reliability curve nearly coincides with the truth.

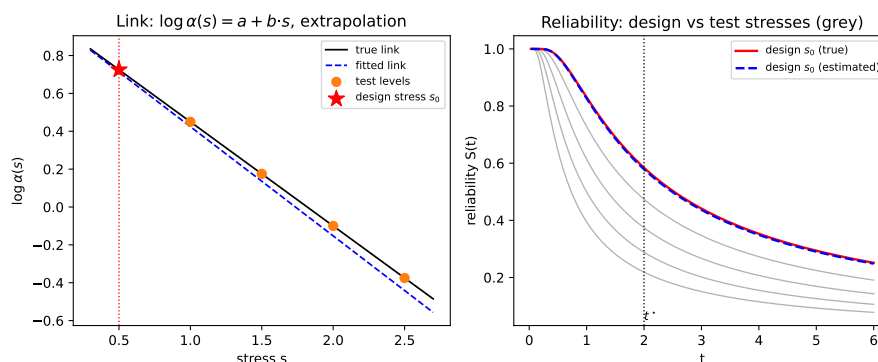


Figure 7. Design–stress extrapolation ($n = 80$, complete). Left: The fitted log–linear link through the four test levels, extrapolated to the design stress s_0 (star). Right: The resulting design reliability curve (true versus estimated) against the test–level curves (grey); $t^* = 2$ marked. Indicative of the pattern in Table 9: The link line is close to the true line and the extrapolated curve close to the true design–stress curve.

Enforcing the sign constraint on b by construction. Section 5.2 derives that monotonically decreasing reliability across tested stresses requires $b\varphi'_0(s) < 0$ – for this log-linear design ($\varphi_0(s) = s$, $\varphi'_0 > 0$), which means that $b < 0$ throughout. As with the order constraint $\alpha_1 > \alpha_2$, which is enforced by construction via the η -space reparameterization of Remark 4 everywhere it is used in this paper, we impose the analogous constraint $b = -e^c$, $c \in \mathbb{R}$ on the link slope and compare it against an unconstrained two-parameter optimizer over $(a, b) \in \mathbb{R}^2$, side by side on identical simulated data sets at $M = 1000$ replicates per scenario – larger than the $M = 400$ used for Table 9 itself, to obtain a precise wrong-sign-rate estimate. Table 10 reports the results for the same design–stress simulation of Table 9 (same stresses, change times and true parameters). The exact wrong-sign rate under the unconstrained estimator is 2.4% (Complete, $n = 40$), 0.3% (Complete, $n = 80$), 5.1% (Progressive, $n = 40$) and 2.5% (Progressive, $n = 80$). Table 10 reports bias and MSE of $\hat{b}$ and of $\widehat{R}_0$ for both estimators: the constrained estimator guarantees $\hat{b} < 0$ (hence a physically admissible, monotonically decreasing fitted reliability) in every one of the 4,000 replicates at no cost in accuracy – $\text{MSE}(\hat{b})$ and $\text{MSE}(\widehat{R}_0)$ are equal to or marginally smaller under the constrained estimator in every scenario (by 0.2%–4.5%), so the constraint removes the sign risk essentially for free. We recommend the constrained reparameterization as the estimator of record for this model: check the sign as a diagnostic when a quick unconstrained fit is all that is needed, but use the constrained estimator by default given its verified accuracy; Table 10 retains the unconstrained comparison only to document the improvement, and the corresponding driver script is released alongside this manuscript.

Table 10. Unconstrained versus sign-constrained ($b = -e^c$) estimation of the link model, $M = 1000$ replicates per scenario, identical simulated data sets for both estimators. True $b = -0.55$, true $R_0 = 0.5838$. “Wrong sign” is the exact fraction of replicates with $\hat{b} \geq 0$ under the unconstrained estimator (structurally 0% under the constrained estimator).

Scheme	n	m	Unconstrained		Wrong sign	Constrained ($b = -e^c$)	
			$\text{MSE}(\hat{b})$	$\text{MSE}(\widehat{R}_0)$		$\text{MSE}(\hat{b})$	$\text{MSE}(\widehat{R}_0)$
Complete	40	40	0.1078	0.00949	2.4%	0.1034	0.00925
	80	80	0.0540	0.00473	0.3%	0.0538	0.00472
Progressive	40	24	0.2612	0.01652	5.1%	0.2495	0.01593
	80	48	0.1135	0.00826	2.5%	0.1118	0.00817

Misspecification experiment: Data generated under an Arrhenius link, fitted with the log-linear link. Table 9 and Figure 7 both generate *and* fit data under the same log-linear link, so a wrong functional form is not detectable from within-range goodness-of-fit alone – a potentially serious gap for extrapolation *below* the tested range. We assess this directly with a misspecification experiment: Data are generated under the Arrhenius link $\log \alpha(s) = a + b/s$ with true $(a, b) = (-0.926, 1.375)$, chosen so that $\alpha(s_1) = 1.568$ and $\alpha(s_4) = 0.687$ match the log-linear design’s endpoints exactly (true design–stress reliability $R_0 = S(2; \alpha_0) = 0.9454$ at $\alpha_0 = 6.204$). Each simulated data set (same test levels, change times, n , censoring as Table 9, but $M = 1000$ as in Table 10) is fitted both under the *misspecified* log-linear link and under the *correctly specified* Arrhenius link (oracle). Table 11 reports the results. Under misspecification, $\widehat{R}_0$ is severely biased (bias ≈ -0.29 to -0.34 , RMSE ≈ 0.32 – 0.35) and *does not shrink with n* – the signature of asymptotic bias, not sampling variability. Under the correct link,

the RMSE is $2\text{--}3\times$ smaller and declines with n as expected (Complete: $0.152 \rightarrow 0.104$; Progressive: $0.191 \rightarrow 0.123$). This confirms that the “accurate design–stress extrapolation” claims elsewhere in this paper are restricted to the correctly-specified-link scenario of Table 9: Link misspecification, not sampling error, dominates extrapolation error when the link form is uncertain.

Toward a practical link-adequacy diagnostic. The experiment above shows that misspecification is dangerous but, by design, does not tell a practitioner, who does not know the generating link, how to detect it before extrapolating. With $k \geq 3$ tested stress levels (Section 3.3), as in the four-level design used throughout this section, a concrete diagnostic is available in principle: Fit the *saturated* model that estimates a free stage-specific parameter α_j at each of the k tested levels with no link imposed alongside each candidate two-parameter link $\log \alpha(s) = a + b\varphi_0(s)$ for $\varphi_0 \in \{\text{identity}, 1/s, \log s\}$ (log–linear, Arrhenius, inverse–power). Because the saturated model nests every linear-in- φ_0 link as the special case $\alpha_j = \exp\{a + b\varphi_0(s_j)\}$, a likelihood-ratio statistic $2[\ell(\hat{\alpha}_1, \dots, \hat{\alpha}_k) - \ell(\hat{a}, \hat{b})]$, referred to χ^2_{k-2} , tests lack of fit for that specific link; running this for all three candidate links, or equivalently ranking them by the corrected Akaike information criterion (AICc) [37] against the saturated model, gives a link-selection procedure with a concrete, computable decision rule rather than only a warning that misspecification is possible. A residual diagnostic – plotting $\log \hat{\alpha}_j$ against $\varphi_0(s_j)$ for each candidate φ_0 and checking for systematic curvature relative to the fitted line – gives a complementary, less formal check requiring no likelihood-ratio calculation. We have not evaluated the power of this procedure to detect the specific Arrhenius-generating/log–linear-fitting misspecification of Table 11 by simulation here, as doing so rigorously requires its own dedicated size/power study (nominal size under correct specification, power under each misspecification direction, and behavior as $k-2$ grows); we present the procedure itself, with the specific test statistic and its reference distribution, as a concrete recommended practice for practitioners and a well-defined target for that further evaluation (Section 8).

Table 11. Misspecification experiment: Data generated under the Arrhenius link ($a = -0.926$, $b = 1.375$, true $R_0 = 0.9454$), fitted under the misspecified log–linear link and under the correctly-specified (oracle) Arrhenius link, $M = 1000$ replicates per scenario, identical simulated data for both fits.

Scheme	n	m	Misspecified (log–linear)		Oracle (Arrhenius)	
			Bias($\widehat{R}_0$)	RMSE	Bias($\widehat{R}_0$)	RMSE
Complete	40	40	–0.332	0.346	–0.055	0.152
	80	80	–0.339	0.346	–0.032	0.104
Progressive	40	24	–0.293	0.322	–0.075	0.191
	80	48	–0.304	0.317	–0.036	0.123

A/D -optimal is a proxy for R_0 -precision, not an exact minimizer. The recommendation of Section 5.1 to target $p_1 \approx 0.16\text{--}0.20$ for design–stress extrapolation treats $\text{Var}(\hat{\alpha}_1)$ and $\text{Var}(\hat{\alpha}_2)$ symmetrically (A) or jointly (D), but neither directly minimizes $\text{Var}(\widehat{R}_0)$: Via the delta-method gradient $\nabla R_0 \propto (1, \varphi_0(s_0))^T$, $\log \alpha_0$ is the weighted combination $(1 - w)\log \alpha_1 + w\log \alpha_2$ with $w = (\varphi_0(s_0) - \varphi_0(s_1))/(\varphi_0(s_2) - \varphi_0(s_1))$, so the truly R_0 -optimal design minimizes a w -dependent criterion, not the w -free A/D -criteria. Because w depends on the extrapolation distance through

$\varphi_0(s_0)$, the optimal split shifts with s_0 and has no single answer, unlike c -, D - and A -optimality for the tested-range reliability itself. We present $p_1 \approx 0.16$ – 0.20 as a reasonable, s_0 -agnostic heuristic and leave the exact w -dependent design for future work.

7. Application to real step–stress data

We apply the model to a real step–stress data set from small unmanned aerial vehicle (SUAV) reliability testing (Section 7.2). The SUAV data are Type–I censored rather than progressively Type–II censored, so this real-data study validates the broader CEM/IZL modeling and inference approach developed in this paper – together with the separately derived Type–I likelihood of Section 7.2 ((7.1)) – rather than directly validating the progressive Type–II censoring likelihood that constitutes a major part of the paper; the latter is instead validated by the Monte Carlo study of Section 6. In addition, because the one-parameter-per-stage IZL model has an infinite mean, the reliability comparison below is necessarily restricted to survival probabilities and quantiles rather than expected lifetime, and the model comparison of Section 7.1 establishes IZL’s competitiveness only within the general one-parameter inverted family; it does not by itself demonstrate that an infinite-mean, heavy-tailed model is preferable to conventional finite-mean alternatives such as Weibull or lognormal, which are not included in this comparison (see Section 8 for further discussion of this limitation). We emphasize this point explicitly: The SUAV data set consists of only 23 units, Type–I censored at $\tau_2 = 20$, and therefore carries relatively little information about far-tail behavior. The real-data analysis of this section shows that IZL is a competitive, closed-form fit *within* the one-parameter inverted family and is not rejected by the goodness-of-fit test; it does not, and is not intended to, empirically establish that an infinite-mean/heavy-tailed model is the right qualitative choice for this application relative to finite-mean families that are outside the present comparison.

7.1. The general one-parameter inverted family

The application below compares IZL against the general class of one-parameter-per-stage lifetime laws obtained by the reciprocal transformation $T = 1/X$ of a base distribution X (as IZL itself is obtained from the Z–Lindley law), together with the inverse Weibull law in its natural one-parameter, unit-scale form ($\beta \equiv 1$). We restrict this comparison, as throughout the paper, to the inverted family: These are the natural competitors for IZL (same reciprocal-transform construction, same one-parameter-per-stage economy, same non-monotone-hazard motivation) so that any AICc or goodness-of-fit difference reflects distributional shape within a matched model class rather than a difference in the number of estimated parameters or in modeling philosophy (e.g., finite- versus infinite-mean). Section 8 discusses, as a limitation rather than a result of this comparison, what this scope does and does not establish relative to unrelated finite-mean families such as Weibull or lognormal. Every model in this comparison has exactly two free parameters overall (one shape parameter per stage), so AIC/AICc/BIC differences reflect distributional shape alone, with no penalty-driven advantage on either side. Nine further members are considered: inverse Weibull (unit scale), inverse Akash, inverse Lindley, inverse Xgamma, inverse Aradhana, inverse Erlang($k = 2$), inverse half-normal, inverse Rayleigh, and inverse Maxwell. Each base density $f_X(x; \theta)$, $x, \theta > 0$ is listed in Table 12; the inverted density follows $f_T(t; \theta) = t^{-2} f_X(1/t; \theta)$. All ten models (including IZL) are fit by the same step–stress CEM, Type–I-censored likelihood construction of Section 3,

re-executed for the real data set.

Table 12. The ten one-parameter-per-stage members of the general inverted family compared in this section, with the base (pre-inversion) density $f_X(x; \theta)$ or the direct density for inverse Weibull.

Model	Base/direct density
IZL	already inverted; see (2.2)
Inverse Weibull (unit scale)	direct: $\theta x^{-\theta-1} e^{-x^{-\theta}}$ ($\beta \equiv 1$)
Inverse Akash	base: $\frac{\theta^3}{\theta^2 + 2} (1 + x^2) e^{-\theta x}$
Inverse Lindley	base: $\frac{\theta^2}{\theta + 1} (1 + x) e^{-\theta x}$
Inverse Xgamma	base: $\frac{\theta^3}{1 + \theta} (1 + \frac{\theta x^2}{2}) e^{-\theta x}$
Inverse Aradhana	base: $\frac{\theta^3}{\theta^2 + 2\theta + 2} (1 + x)^2 e^{-\theta x}$
Inverse Erlang($k = 2$)	base: $\theta^2 x e^{-\theta x}$
Inverse half-normal	base: $\sqrt{2/\pi} \theta^{-1} e^{-x^2/(2\theta^2)}$
Inverse Rayleigh	base: $(x/\theta^2) e^{-x^2/(2\theta^2)}$
Inverse Maxwell	base: $\sqrt{2/\pi} x^2 \theta^{-3} e^{-x^2/(2\theta^2)}$

7.2. SUAV reliability data

We first apply the model to a small unmanned aerial vehicle (SUAV) reliability data set reported by [33] (Table 2 there), originally from a study of dual, independent-propulsion SUAV units, tracing to [38]. A total of $n = 23$ units were placed on a simple step-stress test with wind speed as the stress factor; the stress increased at $\tau_1 = 15$, and the test was Type-I censored at $\tau_2 = 20$. The original source reports the two test conditions only as “stress level 1” and “stress level 2” (pre- and post-change), without the underlying numerical wind-speed magnitudes; we report the data exactly as published and flag this as a limitation of the source rather than of our analysis, recommending that future applications of this model report numerical stress levels alongside failure times. Eleven units failed before τ_1 , and eight failed after, giving $n_\bullet = 19$ observed failures and 4 units surviving to τ_2 (17.4% right censoring). We use the two-level CEM of Section 3, with α_1 governing the pre-change wind-speed level and α_2 the post-change level.

Table 13. The SUAV failure-time data ([38]; [33], Table 2). Nineteen failure times are observed; the remaining four of the $n = 23$ units on test survive, unfailed, to the Type-I censoring time $\tau_2 = 20$.

Stage	Failure times
Before $\tau_1 = 15$ ($n_1 = 11$)	2.365, 3.467, 5.386, 7.714, 9.578, 9.683, 11.416, 11.789, 12.039, 14.928, 14.938
After $\tau_1 = 15$, before $\tau_2 = 20$ ($n_2 = 8$)	15.325, 15.781, 16.105, 16.362, 17.178, 17.366, 17.803, 19.578
Censored at $\tau_2 = 20$	4 units (times unspecified beyond τ_2)

Type-I-censored likelihood. Section 4 develops the likelihood (4.1) for progressive Type-II censoring; the SUAV application instead uses Type-I (time) censoring at the fixed calendar time τ_2 , so we record the corresponding likelihood explicitly here. With $D_1 = \{i : t_i < \tau_1\}$, the 11 failures observed before the stress change; $D_2 = \{i : \tau_1 \leq t_i < \tau_2\}$, the 8 failures observed after it; $\xi_i = t_i - \tau_1 + h$, their stage-2 equivalent-time images; and $\xi_{\tau_2} = \tau_2 - \tau_1 + h$, the equivalent-time image of the censoring point itself, the Type-I log-likelihood is

$$\ell_1(\alpha_1, \alpha_2) = \sum_{i \in D_1} \log f(t_i; \alpha_1) + \sum_{i \in D_2} \log f(\xi_i; \alpha_2) + (n - n_\bullet) \log S(\xi_{\tau_2}; \alpha_2), \quad (7.1)$$

the last term collecting the $n - n_\bullet = 4$ units that survive, unfailed, to τ_2 ; h enters exactly as in (4.1) via the CEM continuity condition, with the progressive removal terms $R_i \log S(\cdot)$ replaced by this single end-of-test survival term, as Type-I censoring removes units only at the fixed time τ_2 rather than progressively at each failure. This is the likelihood implemented in the supplementary code and maximized below.

Maximum likelihood. Maximizing the Type-I log-likelihood (7.1) gives $\hat{\alpha}_1 = 12.140$ (se 2.481) and $\hat{\alpha}_2 = 1.977$ (se 0.873), with log-likelihood -65.77 and equivalent time $\hat{h} = 2.142$; the order $\hat{\alpha}_1 > \hat{\alpha}_2$ holds, as required (Remark 4). Because only categorical stress-level labels, not numerical wind-speed magnitudes, are available for this data set, we treat wind-speed level 1 as the reference condition, without implying that it represents an established physical use condition; accordingly, the reference-condition reliability $R(t^*) = S(t^*; \hat{\alpha}_1)$ at $t^* = 15, 20$ is 0.541, 0.442. Because stage 1 operates only until $\tau_1 = 15$, $R(15) = 0.541$ falls at the edge of the interval over which stage-1 failures are directly observed, whereas $R(20) = 0.442$ – the value used throughout the remainder of this section and reported in Table 14 – is a parametric extrapolation of the fitted stage-1 distribution $S(\cdot; \hat{\alpha}_1)$ beyond that interval. This extrapolation is a legitimate use of the fitted model, but $R(20)$ should not be read as a directly observed stage-1 performance figure; we report both $R(15)$ and $R(20)$ together in Table 14 so that the interpolation ($t^* = 15$) versus extrapolation ($t^* = 20$) distinction remains visible.

Table 14. Consolidated point and 95% interval estimates for the SUAV data (Section 7.2). ML/Wald uses the classical fit; Bayes uses the single canonical Metropolis–Hastings posterior above. None of the entries in this table is a plain percentile interval; the Bayes column reproduces exactly the posterior means and BCIs reported in the text above, from the same MCMC run.

Quantity	Point estimate	Wald 95% CI	Bayes 95% BCI
α_1	12.140 (ML) / 12.195 (Bayes)	[7.277, 17.002]	[7.813, 17.632]
α_2	1.977 (ML) / 2.196 (Bayes)	[0.267, 3.688]	[0.841, 4.901]
h	2.142 (ML) / 2.433 (Bayes)	[0.122, 4.162]	[0.859, 5.722]
$R(20)$	0.442 (ML) / 0.440 (Bayes)	[0.307, 0.578]	[0.308, 0.576]
$R(15)^*$	0.541 (ML)	–	–

* $R(15)$ is the reference-condition reliability at the stage-1 change time itself (interpolation, within the directly observed stage-1 range); it is reported for comparison only. $R(20)$, used throughout the rest of this section, is an extrapolation beyond $\tau_1 = 15$ (see Section 7.2). Interval procedures in this table were calibrated for $R(20)$ and are not recomputed for $R(15)$.

Information matrix for the Type-I-censored likelihood. The score and observed-information formulas of Section 4 and Appendix B are derived for the progressive Type-II log-likelihood (4.1), whose per-observation terms are $\log f(\cdot) + R_i \log S(\cdot)$. The Type-I log-likelihood (7.1) replaces the progressive removal terms $\sum_i R_i \log S(\cdot)$ by the single aggregate term $(n - n_\bullet) \log S(\xi_{\tau_2}; \alpha_2)$, which has exactly the same functional form as a single removal term with $R_i \rightarrow n - n_\bullet$ evaluated at ξ_{τ_2} ; the observed-information formulas of Appendix B therefore carry over term-by-term: every observed-failure contribution is unchanged, and the Type-I censoring contribution follows by this substitution rather than by treating $(n - n_\bullet)$ as individually observed removals. We derive this Type-I observed-information matrix analytically on this basis and verify it against central finite differences of (7.1) at the fitted $(\hat{\alpha}_1, \hat{\alpha}_2)$ (relative error $< 10^{-8}$, the same finite-difference protocol used in Appendix C). All standard errors, Wald intervals, and covariances reported in this section use this analytically derived and numerically verified Hessian, not a numerical optimizer's approximate Hessian.

Interval estimation. The delta-method Wald interval for $R(20)$ is $[0.307, 0.578]$; a parametric bootstrap ($B = 2000$) gives the percentile interval $[0.331, 0.617]$. Bootstrap percentile intervals for the parameters are $[8.492, 19.641]$ for α_1 and $[0.794, 5.659]$ for α_2 .

Uncertainty for the equivalent time h . We report three intervals for h . The first is the *delta method*: using $h_1 = \partial h / \partial \alpha_1 = -0.181$, $h_2 = \partial h / \partial \alpha_2 = 1.206$ and the observed-information covariance, $\text{se}(\hat{h}) = 1.031$, Wald interval $[0.122, 4.162]$. The second is the *parametric bootstrap*: percentile interval $[0.810, 6.783]$ (right-skewed, since $h_2 \gg |h_1|$). The third is the *Bayesian* interval: see below.

Bayesian analysis. With the same vague gamma priors ($\varepsilon_1 = \varepsilon_2 = 10^{-3}$) and the order-constrained η -space sampler of Algorithm 2 ($T = 3 \times 10^4$, $T_0 = 6 \times 10^3$, acceptance rate 55.3%), the posterior means are $\hat{\alpha}_1 = 12.195 [7.813, 17.632]$, $\hat{\alpha}_2 = 2.196 [0.841, 4.901]$, $\hat{h} = 2.433 [0.859, 5.722]$, and $\hat{R}(20) = 0.440 [0.308, 0.576]$; the order $\alpha_1 > \alpha_2$ holds for every retained posterior draw. This is the single canonical MCMC run (one seed, the settings above) used consistently throughout this section and in Table 14. The classical and Bayesian point estimates agree closely.

Prior-sensitivity check. Because the $n = 23$ SUAV sample is small, and α_2 (and hence the equivalent time h) is markedly less precisely determined than α_1 , we re-run the identical MCMC procedure under two additional, proper, weakly-informative gamma priors on the shared (α_1, α_2) marginal, distinct from the near-flat $\varepsilon_1 = \varepsilon_2 = 10^{-3}$ specification used above, and compare the resulting posterior means and 95% credible intervals; see Table 15. The reference-condition reliability $R(20)$ and α_1 are essentially unaffected by the choice of prior (posterior means shifting by at most 0.02 and 0.7 respectively across all three specifications), and the qualitative conclusions of this section – the order $\alpha_1 > \alpha_2$, the point estimate of $R(20)$, and its credible interval – are unchanged. α_2 and h are more prior-sensitive, as expected for the less-informed second-stage parameter: Under the most informative alternative prior considered (mean 10, closer to the α_1 scale than to the α_2 scale), the posterior mean of α_2 shifts upward by about 33% and that of h by about 35% relative to the vague-prior fit, though the 95% credible intervals continue to overlap substantially. This reflects a genuine feature of the real-data α_2/h estimates rather than an artifact of the vague prior specifically: With only eight stage-2 observations (after the change time τ_1 , four of them Type-I-censored), the

stage-2 likelihood contributes comparatively little information, so the posterior for α_2 retains meaningful prior sensitivity regardless of which specification is treated as the default.

Choice of default prior. The near-flat $\Gamma(0.001, 0.001)$ specification is retained as the primary specification throughout this paper for continuity with the constant-stress IZL literature and comparability across the simulation study and real-data application, not because it is prior-predictively justified for the SUAV design specifically. A more defensible default for a new application is a weakly-informative gamma prior calibrated by a prior-predictive check, of the kind illustrated by the $\Gamma(2, 0.2)$ and $\Gamma(1.5, 0.3)$ rows of Table 15.

Table 15. Prior-sensitivity check for the SUAV real-data Bayesian fit: posterior means and 95% equal-tailed credible intervals under three proper gamma priors on the shared (α_1, α_2) marginal (all restricted to $\alpha_1 > \alpha_2 > 0$, same likelihood, same order-constrained η -space Metropolis–Hastings sampler and $T = 3 \times 10^4$, $T_0 = 6 \times 10^3$ settings as the main analysis above, one seed per prior). The first row is the vague prior used throughout the paper; the next two rows are additional, independently-specified, weakly-informative priors used only for this sensitivity check.

Prior on α_j	α_1	α_2	h	$R(20)$
Vague, $\Gamma(0.001, 0.001)$ [current]	12.09 [7.77, 17.54]	2.22 [0.83, 5.13]	2.51 [0.85, 6.14]	0.437 [0.307, 0.574]
Weakly-informative, $\Gamma(2, 0.2)$ [mean 10]	12.29 [8.12, 17.35]	2.96 [1.11, 6.55]	3.38 [1.15, 7.99]	0.443 [0.319, 0.570]
Weakly-informative, $\Gamma(1.5, 0.3)$ [mean 5]	11.36 [7.44, 16.22]	2.46 [0.95, 5.34]	2.98 [1.09, 6.87]	0.417 [0.296, 0.545]

Goodness of fit. We assess fit with a censoring-aware, parametric bootstrap Kolmogorov–Smirnov (KS) test, with the empirical proportion normalized by the total number of units at risk n rather than by the number of observed failures $n_\bullet < n$: since every unit is followed to τ_2 under Type–I censoring, normalizing by $n_\bullet$ would compare a curve reaching 1 at the last failure against a fitted G that reaches only $G(\tau_2) < 1$. The supremum is taken over the observable interval $(0, \tau_2]$, including the discrepancy at the censoring boundary itself, $|n_\bullet/n - G(\tau_2)|$, alongside the discrepancies at the observed failure times: $D = \max \{ \max_i \max \{ i/n - G(x_{(i)}), G(x_{(i)}) - (i-1)/n \}, |n_\bullet/n - G(\tau_2)| \}$, $i = 1, \dots, n_\bullet$, calibrated by a parametric bootstrap ($B = 2000$, refit by the same Type–I likelihood on each resample). We obtain $D_{\text{obs}} = 0.1212$ and bootstrap $p = 0.4845$ (bootstrap D^* : mean 0.1255, s.d. 0.0368) – no evidence against the fitted model. The censoring-boundary discrepancy is evaluated in calendar time at $\tau_2 = 20$: by the CEM continuity condition of Section 3, $G(\tau_2) = F(\xi_{\tau_2}; \hat{\alpha}_2)$, where $\xi_{\tau_2} = \tau_2 - \tau_1 + \hat{h} = 7.142$ is the stage–2 *equivalent-time* image of τ_2 (not itself a calendar time) so that $G(\tau_2) = F(7.142; \hat{\alpha}_2) = 0.793$, and the endpoint discrepancy is $|19/23 - G(\tau_2)| = |0.826 - 0.793| = 0.033$, smaller than the 0.121 at the observed failure times. The model-implied stage-1 fraction, $F(\tau_1; \hat{\alpha}_1) = 0.459$, closely matches the observed $11/23 = 0.478$ (a 1.9-point gap). This sample size affords limited power to detect misspecification.

Comparison within the general one–parameter inverted family. We refit all nine comparator families of Table 12 to the SUAV data by the same Type–I-censored CEM likelihood with no order

constraint imposed on any of the ten models (four of the nine comparators, inverse Weibull, half-normal, Rayleigh, and Maxwell, do not satisfy $\theta_1 > \theta_2$ at their unconstrained optimum, so the AICc/KS ranking below should be read as a comparison of statistical fit, not a confirmed like-for-like ALT comparison). The ranking by AICc (Table 16) shows inverse Xgamma is marginally best, but IZL, inverse Akash, inverse Lindley, and inverse Aradhana are all within $\Delta\text{AICc} < 0.16$ of it and of each other. These five models receive essentially equivalent AICc support (a negligible AICc separation of less than 0.2), which we do not equate with formal statistical indistinguishability, whereas inverse Weibull, fit here in the fixed-unit-scale form specified for this comparison (Section 7.1), is here one of the *worst* ($\Delta\text{AICc} \approx 32.2$), a figure that is specific to this fixed-scale specification (see the scale-sensitivity analysis of Section 7.1 below). The censoring-normalized, bootstrap-calibrated KS test (Table 18, $B = 500$ resamples per model) shows this five-way cluster is genuinely, not just approximately, nonrejected: all five Lindley-type members have bootstrap $p > 0.44$ – that is, none of the five is rejected by this test, which is evidence of no detected misfit rather than affirmative evidence that the five models are statistically equivalent.

Table 16. Fitted parameters and information criteria, general inverted one-parameter-per-stage family, SUAV data ($n = 23$). ΔAICc is relative to the best-fitting model; D is the censoring-normalized Kolmogorov–Smirnov statistic of Section 7.2. Models ordered by AICc.

Model	$\hat{\alpha}_1$	$\hat{\alpha}_2$	Log-lik.	AICc	ΔAICc	D
Inverse Xgamma	13.346	2.459	−65.702	136.005	0.000	0.115
Inverse Akash	12.016	2.345	−65.759	136.118	0.113	0.119
IZL (this paper)	12.140	1.977	−65.768	136.135	0.130	0.121
Inverse Lindley	12.556	2.329	−65.773	136.145	0.140	0.122
Inverse Aradhana	13.417	3.026	−65.781	136.162	0.157	0.122
Inverse Erlang($k=2$)	22.080	6.345	−67.228	139.056	3.051	0.174
Inverse half-normal	0.129	0.664	−68.565	141.729	5.724	0.244
Inverse Rayleigh	0.092	0.214	−76.114	156.827	20.822	0.377
Inverse Weibull (fixed unit scale) [†]	0.301	0.986	−81.795	168.190	32.185	0.462
Inverse Maxwell	0.076	0.113	−84.954	174.507	38.502	0.462
Inverse Weibull (free scale) [‡]	11.926	1.592	−65.752	138.767	2.762	0.116

[†] This row fixes the inverse Weibull scale at $\beta \equiv 1$ purely to match the two-parameter count of the other nine models; Section 7.1 shows this fixed-scale choice is not unit-invariant and is largely responsible for this row’s poor AICc standing. It should therefore be read as an artificial fixed-scale benchmark, not as evidence that inverse Weibull fits the SUAV data intrinsically poorly; the free-scale refit below places inverse Weibull’s fit much closer to the other members of this family (see “Inverse Weibull (free scale)” below and the accompanying discussion).

[‡] Three-parameter specification with a single shape $\hat{\theta} = 0.968$ shared across stages and stage-specific scales $\hat{\beta}_1, \hat{\beta}_2$ reported, for comparability of layout, in the $\hat{\alpha}_1, \hat{\alpha}_2$ columns; see the scale-sensitivity discussion following this table for the model definition and fitting details. Because this specification carries one additional parameter relative to every other row, it is listed separately below the primary two-parameter ranking rather than folded into it, and its ΔAICc is computed, as is every other row, relative to the best-fitting model (inverse Xgamma, $\text{AICc} = 136.005$).

Four families (inverse Weibull, half-normal, Rayleigh, Maxwell) violate the order condition $\alpha_1 > \alpha_2$ at their unconstrained optimum here, consistent with the shape-versus-rate-parameter

distinction noted above. For inverse half-normal, inverse Rayleigh, and inverse Maxwell, the pre-inversion base variable is $X = \theta Z$ for a fixed, θ -free distribution of $Z > 0$ (a scale family in θ), so $T = 1/X$ is stochastically *decreasing* in θ – the opposite direction from α in the IZL law. The admissible-ALT ordering for these three families is therefore $\theta_1 < \theta_2$, not $\theta_1 > \theta_2$, and Table 16 shows exactly this reversed ordering at the unconstrained optimum for all three ($0.129 < 0.664$; $0.092 < 0.214$; $0.076 < 0.113$), so none of these fits is actually ALT-inadmissible once assessed in its own direction. Because the correct-direction constraint $\theta_1 < \theta_2$ is already satisfied, and not merely close to being satisfied, at the unconstrained optimum for all three families, imposing it explicitly and reoptimizing would leave the fitted parameters, log-likelihood, AICc, and KS statistic for half-normal, Rayleigh, and Maxwell unchanged from Table 16: A nonbinding constraint does not move an unconstrained optimum. The values already reported for these three families are therefore simultaneously the unconstrained MLEs and the admissible-ALT-constrained MLEs, so no additional constrained refit is needed for them. Inverse Weibull is different in a way constrained refitting cannot resolve: its CDF derivative $\partial F/\partial \theta = e^{-x^{-\theta}} x^{-\theta} \ln x$ changes sign with $x \gtrless 1$, so no single global ordering exists for this family, and hence, no constraint of the form $\theta_1 > \theta_2$ or $\theta_1 < \theta_2$ can be imposed uniformly across the observed range the way it can for the other three; restricted to the observed range ($x \in [2, 20]$, where $\ln x > 0$), the fit's ordering $0.301 < 0.986$ is locally consistent with the same reversed direction, though only as a data-range-specific statement. A family whose stochastic-ordering direction depends on where in the support it is evaluated does not admit a single, range-independent CEM ALT formulation of the kind used for the other nine members of Table 12; on this basis, inverse Weibull is not a suitable candidate for the globally-ordered CEM ALT comparison attempted here, independent of the unit-scale issue discussed below (Section 7.1). The practical conclusion is that IZL is competitive with, but not uniquely superior to, several close relatives within this one-parameter inverted family, and the case for IZL specifically rests on its closed-form tractability (Section 2) rather than a decisive fit advantage. Table 17 quantifies that tractability directly against the two comparators that outrank IZL by AICc.

Table 17. Computational cost of the quantile function and of full CEM maximum-likelihood fitting, IZL versus its two closest AICc competitors (inverse Xgamma, inverse Akash), on the SUAV data. *Quantile evaluation*: mean wall-clock time per call, 5000 calls at a representative $\theta = 1.5$ (IZL: closed form via the Lambert W function; Xgamma/Akash: numeric root-finding, as their base CDFs do not invert in closed form). *MLE fitting*: mean wall-clock time per fit, 200 bootstrap-style refits of the SUAV data (Nelder–Mead, identical tolerances for all three models, single core). IZL's advantage is largest for the quantile function itself, which the step–stress CEM's equivalent time h requires on every likelihood evaluation; the advantage for a complete MLE fit is more modest, as fitting also involves optimizer overhead that does not depend on the model. These timings are illustrative computational benchmarks on our specific implementation and hardware, not intrinsic, implementation-independent algorithmic speed ratios.

Model	Quantile eval. (μ s)	Speedup	MLE fit (ms)	Speedup
IZL (this paper)	1.8	1.0×	6.3	1.0×
Inverse Xgamma	46.5	25.7×	14.5	2.3×
Inverse Akash	49.7	27.5×	13.6	2.2×

This AICc/D ranking is a statistical goodness-of-fit comparison within the general one-parameter inverted family; it is not, by itself, a like-for-like accelerated-life-test comparison, as four of the ten models (inverse Weibull, half-normal, Rayleigh, Maxwell) do not satisfy the admissible-ALT ordering $\alpha_1 > \alpha_2$ in the α -parameterization shown here (see the discussion immediately above and in Section 7.1 for the direction in which each of these families is actually ordered).

Sensitivity of the inverse Weibull ranking to the fixed unit scale. The inverse Weibull member of Table 12 is specified with $\beta \equiv 1$ so that, as with every other member of the comparison, it carries exactly two free parameters and hence the same AICc penalty. Fixing a dimensional scale to 1 is not unit-invariant, however: rescaling the SUAV failure times from hours to minutes (multiplying every recorded time and τ_1, τ_2 by 60) and refitting changes the maximized log-likelihood from -81.795 to -172.554 – driven entirely by the change of units acting on a scale that cannot adapt, not by any change in the data’s information content. To assess how much of inverse Weibull’s poor standing in Table 16 reflects this artifact rather than an intrinsic shape disadvantage, we refit a *free-scale* three-parameter version – direct density $f(x; \theta, \beta) = \theta \beta^\theta x^{-\theta-1} e^{-(\beta/x)^\theta}$ with a single shape θ shared across stages along with a stage-specific scale β_j carrying the stress dependence so that T is stochastically increasing in β and that the admissible-ALT ordering is $\beta_1 > \beta_2$; the CEM equivalent time is then the closed form $h = \tau_1 \beta_2 / \beta_1$. Maximizing by Nelder–Mead (multi-start, 200 restarts) on the original (hours) data gives $\hat{\theta} = 0.968$, $\hat{\beta}_1 = 11.93$, $\hat{\beta}_2 = 1.59$ (approximate finite-difference standard errors 0.20, 3.00, 0.88), log-likelihood -65.752 , and, with the extra parameter penalized, AICc = 138.77 ($\Delta\text{AICc} \approx 2.76$ relative to the best-fitting inverse Xgamma) – a striking change from the fixed-scale model’s $\Delta\text{AICc} \approx 32.2$. With a free, unit-adapting scale, inverse Weibull’s fit is close to the five-member Lindley-type cluster, better even than inverse Erlang($k=2$)’s $\Delta\text{AICc} \approx 3.05$ – sixth of eleven rather than ninth of ten, had it been placed alongside the primary ranking. We report it as a separate row below the primary ranking rather than interleaving it, however, since it is not a member of the same one-parameter-per-stage family being ranked there. We conclude that the poor apparent fit of inverse Weibull in the main comparison substantially reflects the arbitrary unit-scale restriction imposed for parameter-count parity rather than an intrinsic shape-family weakness relative to IZL and its closest competitors. We nonetheless retain the fixed-scale, two-parameter specification as the *primary* comparison in Table 16, since it alone keeps every one of the ten models at the same parameter count and AICc penalty structure; the free-scale refit above is a targeted robustness check on the inverse Weibull entry specifically. A systematic treatment that fits every family with, and without, an explicit scale parameter and compares AICc consistently across the full nine-member family is a natural extension (Section 8).

Decision-relevant comparison: Intervals for the reference-condition reliability. Beyond the fit ranking, the practical question is whether the choice among these essentially-equivalent models changes the estimated reliability $R(t^*) = S(t^*; \alpha_1)$ at $t^* = \tau_2 = 20$. We refit each of the five Lindley-type members and compute a 95% percentile-bootstrap interval for $R(20)$ ($B = 1000$; same censoring design as the data).

The five point estimates agree to within 0.015 (0.440–0.455), and the five intervals overlap almost entirely within [0.33, 0.61]: The estimated reference-condition reliability $R(20)$ and its uncertainty are similar across all five closely-supported models. This does not amount to a certification-decision claim; no certification threshold or formal decision rule is specified here, and overlapping confidence intervals do not by themselves imply that every downstream engineering decision would coincide, as

that additionally requires, for example, that a lower confidence bound clear a specific required reliability target, which is beyond the scope of this illustrative data set. What the overlap does support is the narrower conclusion that, among these five closely supported models with negligible AICc separation, choosing the computationally simplest member, IZL (Table 17), does not come at a meaningful cost in the estimated reliability or its precision.

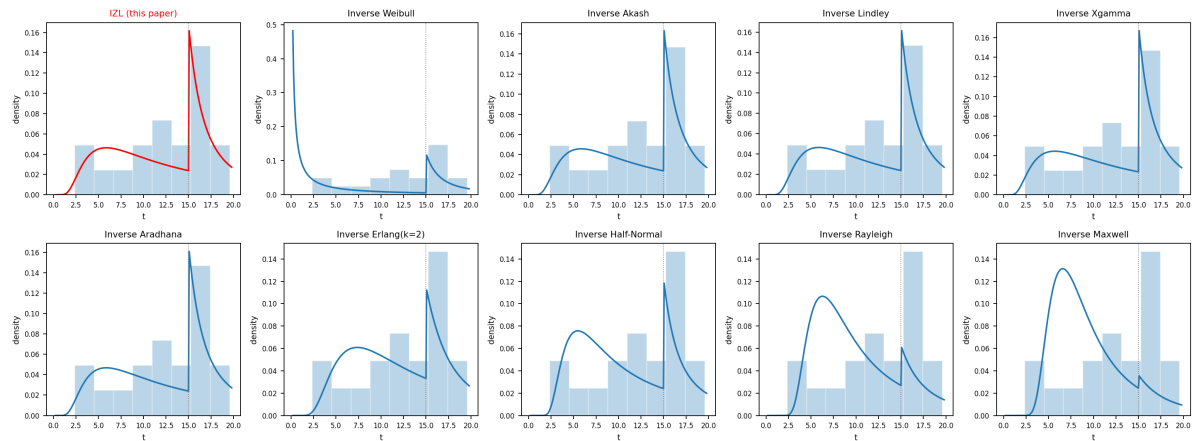


Figure 8. Fitted step-stress CEM density for each of the ten one-parameter-per-stage inverted models, overlaid on the empirical histogram of the observed failure times (SUAV data). The dashed vertical line marks the stress change time $\tau_1 = 15$. IZL is highlighted in red. The five Lindley-type models (top row center-right and bottom row left) visibly track the histogram shape closely, matching their nonrejection in Table 18; inverse Weibull (top left) is visibly the poorest match here.

Table 18. Bootstrap-calibrated Kolmogorov–Smirnov statistic and p -value ($B = 500$ resamples per model) for all ten one-parameter-per-stage inverted models on the SUAV data. Models ordered by D (best fit first); five members of the Lindley-type subfamily (Xgamma, Akash, IZL, Lindley, Aradhana) are all clearly nonrejected ($p > 0.44$). IZL’s p -value here (0.502, from this table’s independent $B = 500$ run across all ten models) differs slightly from the $p = 0.4845$ reported in the main text (a separate $B = 2000$ run for IZL alone); both are consistent with the same underlying $D_{\text{obs}} = 0.121$ and differ only by ordinary Monte Carlo variation between independent bootstrap runs at different B .

Model	D	Bootstrap p -value
Inverse Xgamma	0.115	0.530
Inverse Akash	0.119	0.560
IZL (this paper)	0.121	0.502
Inverse Lindley	0.122	0.449
Inverse Aradhana	0.122	0.457
Inverse Erlang($k=2$)	0.174	0.110
Inverse half-normal	0.244	0.035
Inverse Rayleigh	0.377	< 0.002
Inverse Weibull	0.462	< 0.002
Inverse Maxwell	0.462	< 0.002

Table 19. Point estimate and 95% percentile-bootstrap interval for the reference-condition reliability $R(20) = S(20; \hat{\alpha}_1)$ for each of the five closely-supported (Table 16) models, SUAV data ($B = 1000$ resamples per model).

Model	$\widehat{R}(20)$	95% Boot-p CI	CI width
Inverse Xgamma	0.455	[0.341, 0.610]	0.269
Inverse Akash	0.446	[0.338, 0.611]	0.273
IZL (this paper)	0.442	[0.332, 0.609]	0.277
Inverse Lindley	0.442	[0.330, 0.594]	0.264
Inverse Aradhana	0.440	[0.327, 0.613]	0.286

7.3. Discussion

The SUAV application (Section 7.2) shows no evidence against the fitted IZL model by the censoring-normalized bootstrap KS test ($D = 0.121$, $p = 0.4845$)—a result that, as noted in Section 7.2, reflects the limited power of this $n = 23$ sample to detect misspecification and should not be read as confirming or validating the IZL distribution—and IZL sits among a cluster of five models with negligible AICc separation, all nonrejected by the same test (Tables 16–18). The case for IZL therefore rests on its closed-form tractability—25–28× faster quantile evaluation and a 2.2–2.3× faster MLE fit than its two closest AICc competitors (Table 17) rather than a decisive fit advantage. This application illustrates a clean nonrejection; the diagnostic’s behavior on data the model fits less well remains open for future applications.

8. Conclusions

We have embedded the inverted Z–Lindley distribution in a step–stress cumulative–exposure model and built closed–form model, score, and observed–information machinery for it under progressive Type–II censoring: exact score, strict-concavity result for the complete constant-stress case, together with numerically supported local uniqueness for the coupled censored model (not a general concavity proof for the coupled, censored step–stress likelihood itself, which remains open, Section 4.2), analytic observed information, numerical investigation of optimal design, and design–stress extrapolation. As emphasized in Sections 4 and 5.1, “closed-form” here refers specifically to the model, its quantile function, and the score/observed-information formulas: MLEs still require numerical optimization, the optimal-design recommendations are Monte Carlo based, and the Bayesian analysis requires MCMC or the Tierney–Kadane approximation.

The closed-form IZL quantile makes the equivalent time closed form (Proposition 1) and generalizes to a recursive multiple-step construction (3.6). Because the IZL mean is infinite [18], reliability is carried by quantiles and survival probabilities rather than mean life – a substantive property, not a technical inconvenience. The IZL survival function decays only algebraically ($S(t) \sim \alpha/(2t)$ from (2.1)), the same qualitative tail long used for populations with a genuine subpopulation of very long-lived units (e.g., degradation-resistant components), for which the sample mean is empirically unstable and quantiles/survival probabilities are the more honest summary. The model comparison of Section 7.1 is deliberately restricted to the matched, reciprocal-transform inverted family sharing this tail behavior, so the competitiveness claims of Section 7.3 should be read

as established within that family, not as a general claim against finite-mean alternatives such as Weibull or lognormal. We proved strict concavity of the constant-stress log-likelihood for complete samples (hence unique stage estimators in that setting) and derived the exact score and the analytic observed information, verified against finite differences to a relative error $\sim 3 \times 10^{-9}$ (Appendix C).

What remains is numerically supported rather than proved: existence and local uniqueness of an interior stationary point for the coupled, censored likelihood (Section 4.2); near-nominal coverage for the Wald, Bayesian, and bootstrap intervals throughout, with only a modest, isolated shortfall for the bootstrap at the smallest n under Type-II censoring (Table 2), confirmed by sensitivity analysis not to be an artifact of failed-resample handling (Table 3). It is confirmed to worsen under the heavier-censoring, smaller-parameter C3 configuration, where Boot-p and BCI both undercover materially at small n (Section 6); accurate design-stress extrapolation holds under a correctly specified link but degrades sharply without recovering with n under misspecification (Section 6.5). Furthermore, Tierney-Kadane/MCMC agreement is close, based on a modest 15-replicate comparison (Section 6.4). These are properties of the scenarios simulated, not universal theorems.

Beyond estimation, we gave a numerical investigation of optimal change-time selection under c -, D -, and A -optimality with interior D - and A -optima for one representative design (Section 5.1), a link-function extrapolation with delta-method variance, and bootstrap/Wald/Bayesian intervals (Table 2); a sign-constrained reparameterization of the link slope removes the risk of a physically inadmissible fit at no measurable accuracy cost (Table 10). Applied to the SUAV data, a censoring-normalized, bootstrap-calibrated goodness-of-fit test found no evidence against the fitted model, and IZL placed among a small cluster of models with negligible AICc separation within the inverted family.

Limitations. The real-data illustration should be read as a demonstration of the methodology, not evidence of universal adequacy. The model comparison is scoped to the one-parameter inverted family, so IZL's standing against richer multiparameter or finite-mean families (Weibull, lognormal, generalized exponential) remains an open question. The whole development rests on the CEM assumption, suited to gradual damage-accumulation mechanisms rather than a qualitatively new failure mode, and on progressive Type-II censoring, which does not cover adaptive or hybrid schemes. A further scope point applies throughout: The near-flat gamma prior used as the default Bayesian specification throughout is retained for continuity and comparability across this paper's studies rather than as a prior-predictively justified choice for any specific new application (Section 7.2); a weakly-informative alternative, calibrated by a prior-predictive check, is the more defensible default for future use.

Natural next steps include a competing-risks extension for step-stress data with multiple failure causes; extending Section 7.2's comparison to richer, multiparameter and finite-mean families, and a wider range of data sets; hybrid and adaptive progressive censoring; a fully analytic global-uniqueness proof for the coupled likelihood; a systematic sensitivity analysis of the D - and A -optimal change times over a representative grid of parameters, stress separations, mission times, and censoring plans; optimal multilevel designs that jointly choose stresses and change times, including a genuinely multilevel ($k \geq 3$) real application for the link function of Section 5.2; and a family-by-family catalogue of the stochastic-ordering direction for members of the inverted family (such as inverse Weibull, Section 7.1) whose CDF monotonicity in the shape parameter is not uniform across t . Also open: a size/power evaluation, by simulation, of the likelihood-ratio/AICc

link-selection diagnostic proposed in Section 6.5 for detecting link misspecification with $k \geq 3$ tested stress levels; a systematic scale-parameter sensitivity analysis across the full comparator family of Section 7.1, extending the targeted free-scale inverse Weibull refit of Section 7.2; and a prior-predictive elicitation of the Bayesian prior for new applications, along the lines discussed in Section 7.2, in place of the near-flat default retained here for comparability.

Author contributions

Ghareeb A. Marei: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing—original draft, Writing—review and editing, Visualization, Supervision, Project administration; L.S. Diab: Methodology, Validation, Writing—review and editing, Funding acquisition; Abdullah H. Alenezy: Methodology, Validation, Writing—review and editing; Khudhayr A. Rashedi: Formal analysis, Investigation, Writing—review and editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Funding statement

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2601)

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. W. Nelson, *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*, New York: Wiley, 1990.
2. W. Q. Meeker, L. A. Escobar, *Statistical Methods for Reliability Data*, New York: Wiley, 1998.
3. V. Bagdonavičius, M. Nikulin, *Accelerated Life Models: Modeling and Statistical Analysis*, Boca Raton: Chapman & Hall/CRC, 2002.
4. W. Nelson, Accelerated life testing—step—stress models and data analyses, *IEEE Trans. Reliab.*, **29** (1980), 103–108. <https://doi.org/10.1109/TR.1980.5220742>
5. N. Balakrishnan, Progressive censoring methodology: An appraisal, *TEST*, **16** (2007), 211–296. <https://doi.org/10.1007/s11749-007-0061-y>
6. N. Balakrishnan, E. Cramer, *The Art of Progressive Censoring*, New York: Birkhäuser, 2014.

7. A. S. Hassan, N. Alsadat, M. Kayid, M. Elgarhy, O. S. Balogun, E. M. Almetwally, Statistical inference of constant-stress partially accelerated life tests under type II generalized hybrid censored data from Burr III distribution, *Open Phys.*, **23** (2025), 20250161. <https://doi.org/10.1515/phys-2025-0161>
8. A. Al Mutairi, A. Alrashidi, N. T. Al-Sayed, S. M. Behairy, M. Elgarhy, S. G. Nassr, Bayesian and E-Bayesian estimation based on constant-stress partially accelerated life testing for inverted Topp–Leone distribution, *Open Phys.*, **21** (2023), 20230126. <https://doi.org/10.1515/phys-2023-0126>
9. N. Alotaibi, I. Elbatal, E. M. Almetwally, S. A. Alyami, A. S. Al-Moisheer, M. Elgarhy, Bivariate step-stress accelerated life tests for the Kavya–Manoharan exponentiated Weibull model under progressive censoring with applications, *Symmetry*, **14** (2022), 1791. <https://doi.org/10.3390/sym14091791>
10. A. A. H. Ahmadini, W. K. Mashwani, R. A. K. Sherwani, S. S. Alshqaq, F. Jamal, M. Miftahuddin, et al., Estimation of constant stress partially accelerated life test for Fréchet distribution with Type-I censoring, *Math. Probl. Eng.*, **2021** (2021), 9957944. <https://doi.org/10.1155/2021/9957944>
11. R. Bantan, A. S. Hassan, E. Almetwally, M. Elgarhy, F. Jamal, C. Chesneau, et al., Bayesian analysis in partially accelerated life tests for weighted Lomax distribution, *Comput. Mater. Contin.*, **68** (2021), 2859–2875. <http://doi.org/10.32604/cmc.2021.015422>
12. N. Balakrishnan, R. Aggarwala, *Progressive Censoring: Theory, Methods, and Applications*, Boston: Birkhäuser, 2000.
13. D. A. Ramadan, Statistical inference for alpha power Burr-XII distribution under progressive Type-II censoring: Methodology and application to bladder cancer data, *Innov. Stat. Prob.*, **2** (2026), 63–85. <https://doi.org/10.64389/isp.2026.02195>
14. E. Gazar, A. Mohamed, D. A. Ramadan, M. ElGarhy, B. S. El-Desouky, Estimation of parameters for inverse power Ailamujia and truncated inverse power Ailamujia distributions based on progressive type-II censoring scheme, *Innov. Stat. Prob.*, **1** (2025), 76–87. <https://doi.org/10.64389/isp.2025.01106>
15. M. N. Mousa, M. E. Moshref, N. Youns, M. M. M. Mansour, Inference under hybrid censoring for the quadratic hazard rate model: Simulation and applications to COVID-19 mortality, *Modern J. Stat.*, **2** (2025), 1–31. <https://doi.org/10.64389/mjs.2026.02113>
16. M. Elgarhy, G. S. S. Abdalla, A. S. Hassan, E. M. Almetwally, Bayesian and non-Bayesian analysis of the novel unit inverse exponentiated Lomax distribution using progressive censoring schemes with optimal scheme and data application, *Comput. J. Math. Stat. Sci.*, **5** (2026), 78–108.
17. D. Kundu, A. Joarder, Analysis of Type–II progressively hybrid censored data, *Comput. Statist. Data Anal.*, **50** (2006), 2509–2528. <https://doi.org/10.1016/j.csda.2005.05.002>
18. O. A. Alqasem, A. Elshahhat, A new one-parameter model supports an upside-down bathtub failure rate: Theory, inference, and real-world applications, *Mathematics*, **14** (2026), 1566. <https://doi.org/10.3390/math14091566>
19. A. H. Alenezy, B. Elkalzah, A. F. I. Alahshan, G. A. Marei, Robust inference for stress–strength reliability of the garhy distribution under diverse record schemes with engineering and medical applications, *Mathematics*, **14** (2026), 1940. <https://doi.org/10.3390/math14111940>

20. B. Elkalzah, M. O. Mohamed, K. Elsharkawy, E. S. Osman, A. Aldukeel, G. A. Marei, Classical and Bayesian estimation of stress–strength reliability under the discrete Alpha-power Weibull distribution with incomplete and record data: Application to high-voltage capacitors, *AIMS Math.*, **11** (2026), 6374–6399. <https://doi.org/10.3934/math.2026263>
21. M. Imran, M. A. F. Elbarkawy, F. Jamal, G. A. Marei, An extended Lindley model with applications to entropy-based loss and actuarial risk measures, *Comput. J. Math. Stat. Sci.*, **5** (2026), 1–33.
22. K. A. Rashedi, L. S. Diab, A. H. Alenezy, G. A. Marei, A one-parameter decreasing-hazard lifetime model with exact inference on a bounded-below support, *AIMS Math.*, **11** (2026), 24445–24489. <https://doi.org/10.3934/math.2026987>
23. R. Alotaibi, E. M. Almetwally, Q. Hai, H. Rezk, Optimal test plan of step stress partially accelerated life testing for alpha power inverse Weibull distribution under adaptive progressive hybrid censored data and different loss functions, *Mathematics*, **10** (2022), 4652. <https://doi.org/10.3390/math10244652>
24. S. O. Abd El-Azeem, M. H. Abu-Moussa, M. M. Mohie El-Din, L. S. Diab, On step-stress partially accelerated life testing with competing risks under progressive Type-II censoring, *Ann. Data Sci.*, **11** (2024), 909–930. <https://doi.org/10.1007/s40745-022-00454-0>
25. W. Zhang, W. Gui, Statistical inference and optimal design of accelerated life testing for the Chen distribution under progressive Type-II censoring, *Mathematics*, **10** (2022), 1609. <https://doi.org/10.3390/math10091609>
26. S. Gao, Random weighting estimation of confidence intervals for quantiles, *Aust. N. Z. J. Stat.*, **55** (2013), 43–55. <https://doi.org/10.1111/anzs.12018>
27. S. Gao, Y. Zhong, Random weighting estimation of kernel density, *J. Stat. Plan. Infer.*, **140** (2010), 2403–2407. <https://doi.org/10.1016/j.jspi.2010.02.009>
28. G. Hu, S. Gao, Y. Zhong, C. Gu, Asymptotic properties of random weighted empirical distribution function, *Commun. Stat. Theory Meth.*, **44** (2015), 3812–3824. <https://doi.org/10.1080/03610926.2013.768669>
29. N. Saaidia, T. Belhamra, H. Zeghdoudi, On the Z–Lindley distribution: Statistical properties and applications, *Stud. Eng. Exact Sci.*, **5** (2024), 3078–3097.
30. M. Elgarhy, Garhy distribution with different estimation methods and applications to engineering and medical data, *Rev. Int. Métodos Numér. Cál. Diseño Ing.*, **42** (2026), 078324. <https://doi.org/10.23967/j.rimni.2026.078324>
31. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.*, **21** (1953), 1087–1092. <https://doi.org/10.1063/1.1699114>
32. W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57** (1970), 97–109. <https://doi.org/10.1093/biomet/57.1.97>
33. M. S. Kotb, M. Z. Raqab, Inference for a simple step–stress model based on ordered ranked set sampling, *Appl. Math. Model.*, **75** (2019), 23–36. <https://doi.org/10.1016/j.apm.2019.05.022>

34. L. Tierney, J. B. Kadane, Accurate approximations for posterior moments and marginal densities, *J. Amer. Statist. Assoc.*, **81** (1986), 82–86. <https://doi.org/10.1080/01621459.1986.10478240>
35. A. H. Alenezy, K. A. Rashedi, A. B. Ghorbal, G. A. Marei, Bridging Markov Chain Monte Carlo techniques and Tierney–Kadane approximations for progressively censored Garhy reliability models: Simulation insights and a medical application, *Mathematics*, **14** (2026), 1777. <https://doi.org/10.3390/math14101777>
36. K. A. Rashedi, L. S. Diab, A. H. Alenezy, G. A. Marei, Inference for stress–strength reliability under unified hybrid censoring: A one-parameter model with applications, *Mathematics*, **14** (2026), 2041. <https://doi.org/10.3390/math14122041>
37. K. P. Burnham, D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information–Theoretic Approach*, 2 Eds, New York: Springer, 2002.
38. D. Han, Estimation in step–stress life tests with complementary risks from the exponentiated exponential distribution under time constraint and its applications to UAV data, *Stat. Methodol.*, **23** (2015), 103–122. <https://doi.org/10.1016/j.stamet.2014.09.001>

A. Derivation of the quantile function

Set $F(t; \alpha) = u$ in (2.1), $v = \alpha/t$, $\beta = 2(1 + \alpha)$; then, $\varphi = v/\beta$, and $(1 + v/\beta)e^{-v} = u$; that is, $(\beta + v)e^{-v} = \beta u$. With $w = -(\beta + v)$, this is $we^w = -\beta ue^{-\beta}$; because $v > 0 \Rightarrow w < -\beta < -1$, the relevant branch is W_{-1} , giving $v = -W_{-1}(-\beta ue^{-\beta}) - \beta$ and $t = \alpha/v$, which is (2.3). For $u \in (0, 1)$ the argument lies in $(-e^{-1}, 0)$, so W_{-1} is real.

B. Complete analytic observed information

With $c = 2 + 1/t$, $D = 1 + \alpha c$, and $\varphi = \alpha/[2(1 + \alpha)t]$, the second-order blocks are

$$\begin{aligned} A_\alpha &= -\frac{1}{\alpha^2} + \frac{1}{(1 + \alpha)^2} - \frac{c^2}{D^2} = -\frac{1 + 2\alpha}{\alpha^2(1 + \alpha)^2} - \frac{c^2}{D^2} < 0, \\ B_\alpha &= \frac{\alpha c(2 + \alpha c)}{t^2 D^2} \quad (= A_t \text{ by symmetry}), \quad B_t = \frac{2}{t^2} - \frac{\alpha^2}{t^4 D} - \frac{2\alpha^2 c}{t^3 D} + \frac{\alpha^3 c}{t^4 D^2}, \\ F_{\alpha\alpha} &= \frac{e^{-\alpha/t}}{t} \left[-\frac{G}{t} - \frac{1}{(1 + \alpha)^3} - \frac{1}{2(1 + \alpha)^2 t} \right], \quad F_{\alpha t} = \frac{e^{-\alpha/t}}{t^2} \left[\left(\frac{\alpha}{t} - 1\right)G + \frac{\alpha}{2(1 + \alpha)t} \right], \end{aligned}$$

where $G = \frac{1}{2(1 + \alpha)^2} - 1 - \varphi$ (so $F_\alpha = e^{-\alpha/t} G/t$), and the identity $f_\alpha = F_{\alpha t}$ holds because mixed partials of F commute. The equivalent-time second derivatives, with h_1, h_2 from (3.5) and all quantities evaluated as shown, are

$$\begin{aligned} h_{11} &= \frac{F_{\alpha\alpha}(\tau; \alpha_1)}{f(h; \alpha_2)} - B(h; \alpha_2) h_1^2, \quad h_{12} = -h_1[A(h; \alpha_2) + B(h; \alpha_2)h_2], \\ h_{22} &= -\frac{F_{\alpha\alpha}(h; \alpha_2) + F_{\alpha t}(h; \alpha_2)h_2}{f(h; \alpha_2)} - h_2[A(h; \alpha_2) + B(h; \alpha_2)h_2]. \end{aligned}$$

Assembling (4.4) with the per-unit quantities $\Psi, \Psi^x, P_a, P_{aa}, P_{ax}$ defined in Section 4 reproduces the observed information; we verify the full matrix against the finite-difference Jacobian of the analytic

score with maximum entrywise error 3.8×10^{-9} and each individual block to a maximum discrepancy of 2.3×10^{-8} (protocol and worked example in Appendix C).

C. Finite-difference verification protocol

This appendix reports a numerical verification of the score and observed information formulas of Section 4. The model was implemented directly from Eqs (2.1)–(2.6), (3.2), (4.1)–(4.3), and Appendix B, and every analytic quantity was checked against central finite differences on a generated numerical example.

Scheme. All derivatives were checked by central finite differences, $g'(x) \approx [g(x + \delta) - g(x - \delta)]/(2\delta)$ for a first derivative and the corresponding central stencil for a mixed second derivative, with step size $\delta = 10^{-5}$ in the natural (unstandardized) scale of each parameter.

Numerical example and results. Using Algorithm 1, we generated a progressively censored sample of size $n = 30$, $m = 15$ at $(\alpha_1, \alpha_2) = (1.5, 0.7)$, $\tau = Q(0.4; \alpha_1) = 1.341555$, removal scheme $R_1 = 8$, $R_2 = \dots = R_7 = 0$, $R_8 = \dots = R_{14} = 0$, $R_{15} = 7$ (the “Progressive” scheme of Section 6 at $n = 30$; random seed recorded in the archived code), giving 8 stage-1 and 7 stage-2 observed failures, a balanced split that exercises every block of the 2×2 observed information, including the off-diagonal $\mathcal{I}_{12}$. The resulting MLE, obtained both by direct log-likelihood maximization and, independently, by root-finding the analytic score (4.2) and (4.3), is $(\hat{\alpha}_1, \hat{\alpha}_2) = (1.52478, 0.25386)$ (the two methods agree to 7 decimal places). To verify:

- *Score versus finite-difference gradient of the log-likelihood*, evaluated at a representative *nonstationary* point $(\alpha_1, \alpha_2) = (1.6, 0.75)$ (informative because both are ≈ 0 at the MLE by construction): analytic score $(0.481782, -6.430698)$ against finite-difference gradient $(0.481782, -6.430698)$, maximum discrepancy 5.1×10^{-10} .
- *Analytic observed information (4.4) versus finite-difference Jacobian of the analytic score*, at the MLE: $\mathcal{I}(\hat{\alpha}) = \begin{pmatrix} 12.996 & -7.330 \\ -7.330 & 84.181 \end{pmatrix}$, maximum entrywise discrepancy 2.4×10^{-7} ($\mathcal{I}_{11}$: 8.6×10^{-10} ; $\mathcal{I}_{12}$: 5.0×10^{-9} ; $\mathcal{I}_{22}$: 2.4×10^{-7}); and at the representative point $(1.6, 0.75)$: maximum entrywise discrepancy 1.5×10^{-9} . The eigenvalues of the observed information at the MLE are $(12.249, 84.928)$, both positive, confirming a local maximum on this replicate.
- *Each individual second-order block* of Appendix B ($A_\alpha, B_\alpha, B_t, F_{\alpha\alpha}, F_{\alpha t}$), checked against central finite differences of the corresponding first-order quantity at 10 pseudo-randomly drawn $(t, \alpha) \in [0.1, 5] \times [0.1, 3]$ points: maximum discrepancy 2.3×10^{-8} across all blocks and points.
- *Closed-form Lambert-W quantile (2.3) versus numeric root-finding* of $F(t; \alpha) = u$, at 20 pseudo-random (u, α) points: maximum discrepancy 9.2×10^{-14} (machine precision), confirming the quantile derivation of Appendix A.

All discrepancies above are at or below 2.4×10^{-7} , itself many orders of magnitude smaller than the quantities being compared (e.g. 2.4×10^{-7} against $\mathcal{I}_{22} = 84.18$ is a relative error of 3×10^{-9}), confirming that the analytic score, observed information, and quantile formulas in the manuscript are internally consistent with the stated log-likelihood.

D. Supplementary code

The core IZL step–stress model (CDF/PDF/quantile, score, observed information, the equivalent time h , and the progressive- and Type-I-censored log–likelihoods) and the full pipeline used to produce every numerical result in this paper are released as self-contained, individually runnable scripts with explicit random seeds. Each script has been independently executed against the corresponding reported value in this manuscript. Key scripts:

- `model/izl_stepstress.py`, `mcmc/run_mcmc.py`: real-data MLE and posterior summaries (Section 7);
- `simulation/rerun_c1_c2_full.py`, `simulation/run_c3.py`: point-estimation Tables 5–7 ($M = 1000$; 146 s/scenario on a single core);
- `simulation/interval_table_M1000.py`, `simulation/interval_table_regen_sensitivity.py`: interval-estimation Table 2 and its sensitivity check, Table 3;
- `gof/suav_ks_bootstrap.py`: the SUAV goodness-of-fit statistic of Table 18;
- `comparison/inverted_family_comparison.py`, `comparison/quantile_benchmark.py`, `comparison/reliability_ci_comparison.py`: Tables 16–18, the timing comparison of Table 17, and the reliability intervals of Table 19.

The archive additionally includes the progressive-censoring sample generator (Algorithm 1), the Metropolis–Hastings sampler (Algorithm 2), a diagnostic script using the unconstrained (α_1, α_2) optimizer for comparison against the order-preserving estimator of Remark 4, and all figure-generation scripts.



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)