



Research article

Preservation of relative aging orders under Cumulative-Hazard transformations

Ibrahim A. Nafisah, Mohamed Kayid and Tarek M. AlSayed*

Department of Statistics and Operations Research, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia

* **Correspondence:** Email: tmahrous@ksu.edu.sa; Tel: +966501720701.

Supplement S1. Computational methods and reproducibility

Scope: source data, numerical implementation, grid-density sensitivity, uncertainty calculations and reproducibility. The manuscript contains the mathematical proofs.

S1. Package structure and execution

The archive contains code/, data/, outputs/ and figures/. Run the following commands from its extracted root in a working copy:

```
python -m pip install -r requirements.txt
python -m unittest discover -s code -p "test_*.py" -v
python code/reproduce.py --bootstrap 999
python code/verify_new_results.py
python code/final_revision.py --bootstrap 4999 --workers 4
python code/verify_final_outputs.py
```

The first reproduction step regenerates the original numerical record; final_revision.py then generates the final-review grid checks, the expanded joint-bootstrap intervals, and print-readable versions of Figures 1–5. Both records are retained with distinct filenames. The final-interval files, rather than the original finite-fit-only intervals, correspond to Tables 5–7 in the final manuscript. Scripts resolve paths relative to their own locations and require no private account or network access after dependency installation.

The final run used Python 3.13.5, NumPy 2.3.5, SciPy 1.17.0 and Matplotlib 3.10.8. The complete deterministic test suite contains 26 tests. Platform-dependent runtime and small optimizer-rounding differences are not expected to match byte for byte; scientific results should agree to the printed

precision. The code preserves full numeric precision in CSV files.

S2. Threshold functions and stable computation

For the gamma threshold, write $P_{n-1}(u) = \sum_{j=0}^{n-1} u^j / j!$. Then

$$\Phi_n(u) = e^{-u} P_{n-1}(u), \quad \omega_n(u) = \frac{u^{n-1}}{(n-1)! P_{n-1}(u)}, \quad \Omega_n(u) = -\log \Phi_n(u).$$

The finite polynomial is evaluated on the log scale. With truncated-Poisson weights proportional to $u^j / j!$, the implementation computes

$$\eta_n(u) = \sum_{j=0}^{n-1} (n-1-j) \Pr(J=j), \quad \kappa_n(u) = \frac{u \omega_n(u)}{\Omega_n(u)}.$$

This avoids subtracting nearly equal large terms. Near the lower tail, the regularized incomplete gamma probability and `log1p` avoid cancellation in the cumulative generator; for larger arguments the finite log-polynomial representation is used. The exponential threshold ($n=1$) is evaluated exactly.

For the saturating threshold,

$$\omega_S(u) = 1 - e^{-u}, \quad \Omega_S(u) = u - 1 + e^{-u}, \quad \eta_S(u) = \frac{u}{e^u - 1}.$$

The code uses `expm1` and a small-argument series where needed. The final-review boundary scanner has a separate closed form for the common gamma-2 threshold; regression tests compare it directly with the general implementation.

S3. Synthetic designs and grid-density sensitivity

The linear parameter order is $(a, b, r_0, r_\infty, \gamma)$, with

$$\lambda_Y(t) = a + bt, \quad r(t) = r_\infty - (r_\infty - r_0)e^{-\gamma t}, \quad \lambda_X(t) = r(t)\lambda_Y(t).$$

Table 2 scenarios are A: (0.6, 0.1, 0.4, 0.8, 0.5), B: (1.0, 0.05, 0.85, 0.98, 0.15), and C: (0.3, 0.2, 0.1, 0.95, 1.5), each evaluated for $(m, n) = (2, 1), (3, 2), (5, 3), (10, 5)$. The principal grid has 20,000 equally spaced points from 0.1 to 15, including both endpoints. Synthetic time is dimensionless. Let $x = \Lambda_X$, $y = \Lambda_Y$ and $\rho_i = \lambda_i / \Lambda_i$. The analytical margins are

$$D_{m,n}^{\text{AF}} = r'/r + \rho_X \eta_m(x) - \rho_Y \eta_n(y), \quad D_{m,n}^{\text{AFA}} = \rho_X \kappa_m(x) - \rho_Y \kappa_n(y).$$

No numerical differentiation is used. Table 4 adds Weibull baselines $\lambda_Y(t) = pt^{p-1}$ with $p = 0.75, 1.75$ and $(r_0, r_\infty, \gamma) = (0.35, 0.85, 0.7)$; its cumulative integral is checked independently by quadrature. The gamma pairs are (2,1) and (5,3), and the common saturating threshold is applied to these two Weibull baselines and linear Scenario A. All 38 reported Table 2/Table 4 minima occur at $t = 15$.

For Table 3, fix $(a, b, r_0) = (0.6, 0.1, 0.4)$ and use the common gamma-2 threshold. The parameter grid has 91 limiting ratios from 0.45 to 1.80 and 80 transition parameters from 0.05 to 2.00: 7,280 configurations, of which 2,960 are certified and 4,320 are outside the sufficient region. Each horizon 15, 30, 60, 120 uses its own time grid starting at 0.1; minima below -10^{-10} are recorded as detected violations.

The final density check repeats all of these designs at 5,000, 10,000, 20,000 and 40,000 points. The 38 minima in Tables 2 and 4 are identical across densities, with the same common endpoint. In the full boundary study, no AF or AFA classification changes relative to 40,000 points. The largest 20,000-to-40,000 minimum differences across all horizons/configurations are 4.44×10^{-8} and

1.70×10^{-9} , respectively. Appendix A gives a compact summary; the archive supplies all 116,480 density/configuration/horizon rows and 76 Table 2/Table 4 scenario rows (each with both margins).

These are deterministic grid comparisons, not Monte Carlo preservation probabilities or a continuous-time optimization certificate. Proposition 5.1, rather than grid agreement, establishes eventual failure for the specified outside family. Its high-precision checks retain the normalizations $y^2 D^{\text{AF}}/\lambda_Y$ and $y^2 D^{\text{AFA}}/(\lambda_Y \log y)$, both tending to $r_\infty^{-1} - 1$.

S4. Public data, likelihood and fitting safeguards

The observed/censored record is transcribed from NIST/SEMATECH, e-Handbook of Statistical Methods, Section 8.4.2.1, “An Arrhenius model example”:

<https://www.itl.nist.gov/div898/handbook/apr/section4/apr421.htm>

The source has 100, 50 and 25 units at 85, 105 and 125 °C, with 5, 35 and 24 failures, respectively, and administrative censoring at 1,000 hours. The CSV has one row per unit; censored rows are reconstructed from the reported cell totals, not from invented unobserved failure times. This remains a public handbook example, not a newly collected field dataset or a source-verified physical threshold mechanism.

For group j , assume $\log T \sim N(\mu_j, \sigma^2)$ independently across units. Write $z_i = (\log t_i - \mu_j)/\sigma$ and let d_i be the failure indicator. The likelihood is

$$\ell(\boldsymbol{\mu}, \sigma) = \sum_i \{d_i [\log \phi(z_i) - \log \sigma - \log t_i] + (1 - d_i) \log Q(z_i)\}.$$

The optimizer uses log scale for σ , a stable normal log-survival function and an analytical gradient. Numerical scale safeguards are $0.03 \leq \sigma \leq 5$; all accepted final fits lie strictly inside them. No location ordering is imposed. The shared-scale estimates are (8.1521303, 6.4872486, 5.3327532) and $\hat{\sigma} = 0.7849387$, with negative log-likelihood 476.0904650. The separate-scale fit gives 475.1119058; its likelihood-ratio statistic is 1.9571184 (two asymptotic degrees of freedom, approximate $p = 0.37585$). This diagnostic does not prove scale equality.

The original fitter retains its documented precision-loss restart rule. In the final bootstrap, a further independent BFGS restart is used only when a refit fails the acceptance check; the objective must not deteriorate by more than 10^{-7} and the maximum absolute score must be at most 5×10^{-5} . The largest accepted score component in the delivered final bootstrap is 4.27×10^{-5} . Numerical failures are not silently accepted or resimulated.

S5. Joint bootstrap and uncertainty propagation

The final manuscript uses 4,999 parametric bootstrap tests with random seed 20260925. Each test is generated under the fitted shared-scale model and retains the original cell sizes:

$$\log T_{ij}^{(b)} = \hat{\mu}_j + \hat{\sigma} Z_{ij}^{(b)}, \quad Z_{ij}^{(b)} \sim N(0,1).$$

$$Y_{ij}^{(b)} = \min\{\log T_{ij}^{(b)}, \log 1000\}, \quad d_{ij}^{(b)} = \mathbf{1}\{\log T_{ij}^{(b)} \leq \log 1000\}.$$

The superscript (b) labels a bootstrap replication. Each test is jointly refitted, without order constraints, and each contrast $\Delta_{a,b} = \mu_a - \mu_b$ and both transformed ratios at 500 hours are recomputed from the same fitted parameter vector. This preserves their dependence on the common scale. The central percentile endpoints (2.5% and 97.5%) are taken over all 4,999 estimand draws, with linear order-statistic interpolation. These are approximate pointwise, model-conditional intervals, not simultaneous time bands, familywise intervals across comparisons, or a coverage guarantee under misspecification.

Of the 4,999 samples, 4,982 have finite location MLEs. The remaining 17 have no observed failure at 85 °C; none of the other two cells is all censored. The likelihood for an entirely censored cell has supremum zero in log-likelihood contribution as its location tends to $+\infty$. We therefore fit the informative cells for their locations and the shared scale and retain $\hat{\mu}_{85}^{(b)} = +\infty$ for that cell. Its contrasts against a finite warmer-cell location have limit $+\infty$; for the specified gamma and saturating scenarios, both 85 °C numerator ratios at a finite time have limit zero. No sample is deleted, redrawn or clipped to a large finite location. The boundary locations are not additional proper baselines for the theoretical results.

Every final refit passes the numerical acceptance checks. The 17 boundary cases occupy fewer than 2.5% of the bootstrap distribution; the reported percentile endpoints are consequently finite. The implementation explicitly rejects unresolved numerical optimization and contrasts between two all-censored cells, rather than assigning an arbitrary value. Tests cover these boundary branches, although the latter situation does not occur in the delivered simulation.

The final common-scale interval is [0.6358242, 0.9197931]. Location-contrast estimates and intervals are:

Comparison (°C)	Estimate	95% interval
85 vs 105	1.665	[1.269, 2.212]
105 vs 125	1.154	[0.778, 1.543]
85 vs 125	2.819	[2.360, 3.415]

The 85 °C median estimate is 3,471 hours, with interval 2,457–5,883 hours; the entire interval is beyond the 1,000-hour observation endpoint. Table 6 reports all nine transformed hazard-ratio intervals and all nine transformed cumulative-hazard-ratio intervals, rounded to three significant digits. Full-precision values and all 44,991 ratio-scenario draws are supplied. Finite-only parameter percentiles are additionally retained as a labeled sensitivity comparison; they are not used for the final manuscript intervals.

The bootstrap assumes the common-scale lognormal model and administrative censoring. It propagates parameter uncertainty but does not eliminate model uncertainty, validate a physical threshold or NHPP recurrence mechanism, or turn positive location contrasts into distribution-free order tests. The single-time ratio intervals should not be interpreted as probabilities of global preservation.

S6. Output ledger and verification boundaries

File	Meaning
outputs/table2.csv, table4.csv	Original synthetic minima; unchanged at reported precision.
outputs/table3_horizons.csv, boundary_all_configurations.csv	Original 20,000-point boundary record and summaries.
outputs/grid_density_summary.csv	Sixteen density/horizon summaries, including minimum discrepancies and classification counts.
outputs/grid_density_boundary.csv	All 116,480 density/configuration/horizon rows.
outputs/grid_density_tables2_4.csv	All 76 density/scenario rows for Tables 2 and 4.
outputs/table5_final_intervals.csv	Final locations, medians and their bootstrap intervals.
outputs/table6_final_ratio_intervals.csv	Final Table 6 estimates and 18 pointwise intervals.
outputs/table7_location_contrasts.csv	All three final location contrasts and intervals.
outputs/bootstrap_final_fits.csv	All 4,999 refits with scores, counts and finite/boundary status.
outputs/bootstrap_final_transformed_draws.csv	All 44,991 paired ratio-scenario draws.
outputs/bootstrap_boundary_sensitivity.csv	Explicit finite-only versus retained-boundary comparison.
outputs/final_revision_summary.json	Final environment, seed, sample counts and summary values.
outputs/final_output_verification.json	Fresh cross-checks of retained data and all numerical outputs.

The original `bootstrap_fits.csv` and `run_summary.json` remain as the 999-sample historical reproduction record, not the final interval analysis. The synthetic `numerical_regression_case.npz` is an optimizer test fixture, not observed component data. Passing numerical or symbolic tests supports the tested calculations; it is not an independent proof of every theorem or a model-validity certificate. Read the manuscript, this supplement and the data-source statement together.

Primary data source: NIST/SEMATECH, e-Handbook of Statistical Methods, Section 8.4.2.1, cited as [24] in the manuscript. No observed data or threshold scenario was added in the final reviewer revision.



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 (<https://creativecommons.org/licenses/by/4.0>)