mathematics

Article

A Scale-Invariant Adaptive Test for IFR Alternatives Based on Cumulative Residual Entropy Under Proportional Hazards

Mashael A. Alshehri

Department of Quantitative Analysis, College of Business Administration, King Saud University, Riyadh 11362, Saudi Arabia; mealshehri@ksu.edu.sa

Abstract

Testing exponentiality against increasing failure rate alternatives is central to reliability theory and lifetime data analysis. This paper develops the Adaptive Cumulative Residual Entropy Test under Proportional Hazards (Adaptive CRE-PH Test), a scale-invariant nonparametric procedure that unifies a Tsallis-entropy departure functional with a proportional-hazards transformation. For each tuning value, the fixed-q statistic admits a normalized-spacing representation. Under exponentiality, the spacing proportions follow a Dirichlet distribution, yielding exact finite-sample means, covariances, and a joint null characterization of the adaptive maximum. Joint asymptotic normality and consistency under fixed alternatives are established for the complete dependent score vector. Structural analysis of the score family motivates a prespecified moderate grid governed by endpoint stability, directional diversity, and multiplicity economy, rather than retrospective power optimization. Monte Carlo experiments demonstrate accurate size control, explicitly quantify calibration stability, and show power close to the best fixed-q component under linear failure rate, Makeham, and Weibull alternatives. A dedicated power experiment confirms substantial detection capability against an IFRA-but-not-IFR benchmark, showing that the broader population sign condition has practical as well as theoretical relevance. Additional DFR and bathtub experiments clarify directional specificity: rejection provides evidence against exponentiality in the IFR-sensitive direction but does not, without shape-specific inference, establish a globally increasing hazard. Three real-data applications illustrate the practical importance of distinguishing formal directional inference from exploratory Q–Q and total-time-on-test diagnostics.

Keywords: nonparametric inference; exponentiality testing; increasing failure rate; adaptive testing; Tsallis entropy; lifetime data

MSC: 62G10; 62N05; 94A17; 62E20

Academic Editors: Maurizio Brizzi and Sung Nok Chiu

Received: 15 June 2026
Revised: 3 August 2026
Accepted: 5 August 2026
Published: 11 August 2026

1. Introduction

The exponential distribution is the canonical benchmark for lifetime modeling in reliability theory, survival analysis, and life testing. Its constant hazard rate represents the absence of aging and gives rise to the memoryless property, whereby the residual lifetime, conditional on survival to any age, has the same distribution as the lifetime of a new unit. This unique combination of structural simplicity and analytical tractability makes the exponential distribution a natural reference model for systems whose failures are driven primarily by random shocks rather than by cumulative degradation or wear [1].

 https://doi.org/10.3390/math14162902

In many practical applications, however, the assumption of a constant hazard rate is unrealistic. Mechanical components, electronic devices, industrial systems, and biological organisms are subject to fatigue, corrosion, cumulative stress, material deterioration, and wear, all of which contribute to an increasing risk of failure over time. The increasing-failure-rate (IFR) class, characterized by a non-decreasing hazard function, therefore provides one of the fundamental models of positive aging. Detecting IFR behavior has important implications for preventive maintenance, replacement scheduling, warranty design, spare-parts management, and reliability assessment of engineering systems [2,3]. Consequently, testing exponentiality against IFR alternatives is more than a conventional goodness-of-fit problem; failing to detect aging may result in adopting a non-aging model for a population that is, in fact, undergoing progressive deterioration.

The IFR class is a proper subclass of the broader increasing failure rate average (IFRA) class. Although every IFR distribution is IFRA, the converse does not generally hold. This distinction is central to the theoretical framework developed in this paper. While the primary reliability motivation and comparative evaluation focus on IFR alternatives, the non-negativity condition underlying the proposed population departure functional is, in fact, equivalent to the IFRA property. Consequently, the formal domain of positive-aging alternatives covered by the proposed procedure extends beyond the strict IFR class. This broader scope is examined both theoretically and through a dedicated IFRA-but-not-IFR benchmark.

Testing exponentiality against positive-aging alternatives has attracted sustained attention in the statistical literature. Early contributions include the rank-based procedure of Proschan and Pyke [2] together with the asymptotic investigations of Bickel and Doksum [4] and Bickel [5]. Barlow and Doksum [6] subsequently proposed an asymptotically minimax test based on the signed area between the total-time-on-test transform and the diagonal. Other notable developments include the U-statistic tests of Ahmad [7] and Deshpande and Kochar [8], the scaled total-time-on-test procedure of Klefsjö [9], and the IFR-plot methodology of Aly [10], including its extension to randomly censored data. More recently, Mitra and Anis [11] and Anis [12] developed generalized classes of L-statistics and established their exact and asymptotic distributions, together with consistency properties.

More recently, information-theoretic approaches have emerged as a powerful framework for lifetime testing. In particular, entropy and cumulative-residual-entropy (CRE) functionals provide interpretable measures of survival uncertainty, residual-life information, and departures from the memoryless property. Within this framework, Alqefari [13] proposed nonparametric tests based on fractional generalized cumulative residual entropy and derived their exact finite-sample null distributions using normalized spacings. These contributions highlight the flexibility of entropy-based procedures while also revealing an inherent limitation: their detection performance can depend strongly on the choice of a tuning parameter, which determines the regions of the lifetime distribution receiving the greatest statistical emphasis.

This dependence is both beneficial and challenging. Different tuning values emphasize different regions of the lifetime distribution and are therefore sensitive to different forms of departure from exponentiality. However, the underlying aging mechanism is unknown prior to testing, making it impossible to identify an optimal tuning value in advance. A choice that performs well under a linear-failure-rate model may be less effective for a Makeham alternative, whereas a value well suited to Weibull-type wear-out may exhibit reduced sensitivity to milder or more localized departures. Consequently, a fixed-tuning procedure may perform exceptionally well for certain alternatives while suffering a substantial loss of power for others. The central methodological challenge is therefore to preserve the flexibility of entropy-based weighting while reducing dependence on an arbitrarily selected tuning parameter, without sacrificing finite-sample validity or interpretability.

To address these challenges, we develop the Adaptive Cumulative Residual Entropy Test under Proportional Hazards (Adaptive CRE-PH Test), a scale-invariant nonparametric procedure for testing exponentiality against positive-aging alternatives. The proposed methodology is built upon a cumulative residual Tsallis-entropy departure functional embedded within a proportional-hazards transformation. This construction gives rise to a family of scale-invariant nonparametric test statistics indexed by a tuning parameter that governs their regional and tail sensitivity. Unlike a purely technical smoothing parameter, this index has a clear structural interpretation: it determines the geometry of the induced score function and, consequently, the regions of the lifetime distribution that receive the greatest statistical emphasis.

The tuning parameter q has two mathematically linked interpretations. Algebraically, the power-survival transformation $S_q(t) = [S(t)]^q$ generates a proportional-hazards family with transformed hazard $qh(t)$. Information-theoretically, the same exponent is the Tsallis order governing the non-extensive weighting in the cumulative-residual-entropy functional. Thus, q is neither an estimated regression coefficient nor an unknown physical hazard ratio. It is a prespecified transformation index whose variation changes the geometry of the score function and thereby the directional sensitivity of the resulting test.

The contributions of this paper are fourfold. First, we develop a scale-invariant family of cumulative-residual-entropy statistics derived from a population departure functional whose sign characterization extends from the IFR class to the broader IFRA class. Second, the exactly standardized fixed-q statistics are integrated through a max-type adaptive procedure over a prespecified finite grid constructed according to the principles of endpoint stability, directional diversity, and multiplicity economy, thereby separating structural adaptation from retrospective power optimization. Third, the complete dependent score vector is characterized under exponentiality through normalized Dirichlet spacings, yielding exact finite-sample means, covariance structure, and the exact joint null distribution of the adaptive maximum, together with joint asymptotic normality and consistency under fixed alternatives. Finally, the proposed methodology is supported by reproducible calibration, extensive Monte Carlo validation, a complete mathematical development, and illustrative R scripts for representative simulation studies and graphical diagnostics.

Unlike many adaptive procedures that rely primarily on empirical calibration, the theoretical development presented here explicitly characterizes the adaptive statistic itself. For each fixed value of q, the proposed statistic admits a normalized-spacing representation. Under exponentiality, the normalized-spacing proportions follow a Dirichlet distribution, providing exact finite-sample moments together with the exact joint null characterization of the dependent fixed-q score vector and its adaptive maximum. For any prespecified finite tuning grid, joint asymptotic normality follows from multivariate L-statistic theory combined with the multivariate delta method, while consistency is established whenever at least one component possesses a strictly positive population departure. Monte Carlo simulation is employed solely to evaluate the exact joint Dirichlet probabilities numerically; it does not replace the theoretical characterization of the null distribution.

A comprehensive simulation study evaluates the finite-sample calibration and power of the proposed procedure under linear failure rate, Makeham, and Weibull alternatives using exact null centering and scaling. Repeated independent calibration runs further assess the numerical stability of the adaptive critical values. A dedicated experiment based on a two-component parallel exponential system examines performance under an IFRA-but-not-IFR alternative, demonstrating that the broader theoretical scope of the proposed methodology is accompanied by meaningful finite-sample detection capability. Additional experiments involving DFR and piecewise-exponential bathtub alternatives highlight the directional specificity of the test. Finally, three real-data applications demonstrate the practical value of distinguishing formal IFR-oriented statistical inference from exploratory Q–Q and total-time-on-test (TTT) diagnostics.

Collectively, these theoretical, computational, and empirical developments establish the Adaptive CRE-PH Test as a unified framework for testing exponentiality against positive-aging alternatives, combining exact finite-sample validity, rigorous asymptotic justification, computational transparency, and practical applicability within a single coherent methodology.

The remainder of the paper is organized as follows. Section 2 introduces the population departure functional and develops the fixed-q statistics underlying the Adaptive CRE-PH Test. Section 3 establishes the exact joint finite-sample null distribution, asymptotic properties, consistency, and adaptive calibration of the proposed procedure. Section 4 investigates the sensitivity of the fixed-q statistics and evaluates the structural adequacy of the proposed tuning grid. Section 5 assesses the performance of the Adaptive CRE-PH Test through simulation studies, comparisons with existing procedures, and investigations under IFRA-but-not-IFR and off-direction alternatives. Section 6 presents three real-data applications illustrating the practical performance of the proposed methodology. Finally, Section 7 summarizes the main findings, discusses the limitations of the proposed approach, and outlines several directions for future research.

2. Construction of the Adaptive CRE-PH Test

This section develops the fixed-parameter component on which the Adaptive CRE-PH Test is built. For later use, we denote the cumulative hazard function by

$$H(t) = \int_0^t h(s)ds = -\log S(t), \quad t \geq 0.$$

Since the exponential scale family and the IFR class have already been specified in Section 1, the testing problem considered in this paper is written as

$$H_0 : F \in \mathcal{E}_0, \quad H_1 : F \in \mathcal{F}_{\mathrm{IFR}}, \quad F \notin \mathcal{E}_0.$$

The purpose of this section is to develop the Tsallis-entropy-weighted population departure measure and the corresponding scale-invariant fixed-q statistic that form the theoretical foundation of the Adaptive CRE-PH Test.

2.1. A Local IFR Departure Kernel

The IFR property is naturally expressed through local comparisons of the hazard rate. If h is non-decreasing, then $h(t) \geq h(s)$ whenever $0 < s \leq t$. This motivates the local IFR departure kernel

$$\eta(t) = \int_0^t [h(t) - h(s)]ds = th(t) - H(t), \quad t \geq 0.$$

Under exponentiality, the hazard rate is constant, so $\eta(t) = 0$ for all $t > 0$. Under strict IFR behavior, $\eta(t)$ is non-negative and becomes positive in regions where the hazard rate increases. Hence, $\eta(t)$ provides a pointwise measure of departure from the exponential boundary in the IFR direction. The next step is to aggregate this local departure over the support of the distribution using an information-based survival weight.

2.2. Population Departure Functional for the Adaptive CRE-PH Test

For a fixed $q > 0$, consider the power-survival transformation $S_q(t) = [S(t)]^q$. This transformation belongs to a proportional-hazards family because its hazard is qh(t). Here q is used as a deterministic transformation index, not as an estimated hazard ratio from a covariate model. The same q is also the Tsallis order in the resulting cumulative-residual-entropy functional. Thus, the PH and Tsallis interpretations are two mathematically linked views of one index: the PH representation supplies the transformed lifetime model, whereas the Tsallis representation explains the associated non-extensive weighting and score sensitivity. Define the population departure functional underlying the Adaptive CRE-PH Test by

$$D_q(F) = \int_0^\infty [S(t)]^q \eta(t)dt = \int_0^\infty th(t)[S(t)]^q dt - \int_0^\infty [S(t)]^q H(t)dt. \tag{1}$$

For any fixed $q > 0$, the functional in (1) is zero under the exponential null model. If F is strictly IFR and the integrals in (1) are finite, then $D_q(F) > 0$. Thus, $D_q(F)$ is a directional functional for detecting departures from exponentiality toward IFR alternatives. The parameter q is not merely a technical constant; it determines how the survival curve is weighted and is therefore central to the adaptive strategy developed later.

The preceding positivity statement can be sharpened. Let $H(t)$ denote the cumulative hazard function and $h(t)$ its derivative at points of differentiability. The sign condition governing the proposed departure functional is expressed through the following quantity:

$$\eta(t) = t\,h(t) - H(t).$$

$$\frac{d}{dt}\left\{\frac{H(t)}{t}\right\} = \frac{t\,h(t) - H(t)}{t^2} = \frac{\eta(t)}{t^2} \geq 0.$$

Thus, non-negativity of the departure kernel is equivalent to monotonicity of the average hazard $\frac{H(t)}{t}$, which is the IFRA condition. Since every IFR distribution is IFRA, all results established for IFR alternatives remain valid, while the functional also covers a broader class of positive-aging distributions. Throughout the paper, the term IFR-oriented refers to the principal reliability interpretation and comparison framework, whereas the formal alternative domain induced by the population functional includes IFRA distributions.

The first term in (1) has a proportional-hazards interpretation. Let X_q denote the lifetime random variable with survival function $S_q(t) = [S(t)]^q$, and let $\mu_q = E(X_q)$. Since the density associated with X_q is $f_q(t) = qf(t)[S(t)]^{q-1}$, it follows that

$$\int_0^\infty th(t)[S(t)]^q dt = \int_0^\infty tf(t)[S(t)]^{q-1} dt = \frac{\mu_q}{q}. \tag{2}$$

The second term in (1) is connected with the cumulative residual entropy of the transformed lifetime. Let

$$\mathcal{E}(X_q) = -\int_0^\infty S_q(t)\log S_q(t)dt$$

be the cumulative residual entropy of X_q. Since $\log S_q(t) = q\log S(t) = -qH(t)$, we obtain

$$\mathcal{E}(X_q) = q\int_0^\infty [S(t)]^q H(t)dt.$$

Combining this identity with (1) and (2) yields the compact representation

$$D_q(F) = \frac{1}{q}\left[\mu_q - \mathcal{E}(X_q)\right]. \tag{3}$$

Representation (3) clarifies the information-theoretic meaning of the proposed departure measure. It compares the proportional-hazards mean with a cumulative-residual-entropy-type quantity computed under the same power-transformed survival model. Accordingly, the phrase PH parameter refers to the algebraic effect of q on the transformed hazard, while the Tsallis tuning parameter refers to its information-weighting role. These are not competing interpretations or two separately estimated parameters; they are complementary consequences of the single transformation $S_q = S^q$. Different choices of q therefore generate different score geometries and sensitivity profiles, which motivates the adaptive max-type construction.

Example 1. *Weibull benchmark. Let X have Weibull survival function* $S(x) = exp(-x^{\alpha}),\quad x > 0,$ $\alpha > 0$. *The case* $\alpha = 1$ *corresponds to exponentiality, whereas* $\alpha > 1$ *corresponds to an IFR lifetime model. Direct calculation gives*

$$\mu_q = \int_0^{\infty} [S(x)]^q dx = q^{-1/\alpha}\Gamma\left(1 + \frac{1}{\alpha}\right),$$

and

$$\mathcal{E}\left(X_q\right) = \frac{q^{-1/\alpha}}{\alpha}\Gamma\left(1 + \frac{1}{\alpha}\right).$$

Therefore, from (3),

$$D_q(F) = q^{-(1+\alpha)/\alpha}\Gamma\left(1 + \frac{1}{\alpha}\right)\left(\frac{\alpha - 1}{\alpha}\right).$$

This expression is zero when $\alpha = 1$ *and strictly positive when* $\alpha > 1$. *The example confirms that the proposed functional is aligned with the IFR direction and shows how q affects the magnitude of the departure measure.*

Example 2. *Parallel exponential system: an IFRA but non-IFR benchmark.*

Let the system lifetime be the maximum of two independent exponential component lifetimes with unequal rates. The model and its survival function are

$$X = \max(X_1, X_2), \qquad X_i \sim \mathrm{Exp}(\lambda_i), \quad \lambda_1 \neq \lambda_2.$$

$$\overline{F}_X(t) = e^{-\lambda_1 t} + e^{-\lambda_2 t} - e^{-(\lambda_1+\lambda_2)t}.$$

$$H_X(t) = -\log\overline{F}_X(t).$$

$$h_X(t) = \frac{\lambda_1 e^{-\lambda_1 t} + \lambda_2 e^{-\lambda_2 t} - (\lambda_1 + \lambda_2)e^{-(\lambda_1+\lambda_2)t}}{e^{-\lambda_1 t} + e^{-\lambda_2 t} - e^{-(\lambda_1+\lambda_2)t}}.$$

$$\frac{d}{dt}\left\{\frac{H_X(t)}{t}\right\} = \frac{t h_X(t) - H_X(t)}{t^2} \geq 0.$$

For unequal component rates, the system hazard may increase over an initial interval and then decrease toward the smaller component rate, so it need not be globally non-decreasing. Nevertheless, the average hazard $H_X(t)/t$ *is non-decreasing, and hence the lifetime is IFRA. This provides a natural benchmark outside the strict IFR class for assessing the wider scope of the proposed departure functional. In the numerical benchmark below, the scale-invariant choice of rates is* $\lambda_1 = 1,\ \lambda_2 = 2$.

2.3. L-Statistic Representation and Empirical Statistic

The functional in (3) admits an L-statistic representation. Let $Q(u) = F^{-1}(u)$, $0 < u < 1$, denote the quantile function of F. By rewriting the deviation in terms of $F(x)$, we obtain

$$D_q(F) = \int_0^\infty x J_q(F(x)) dF(x),$$

where the score function is

$$J_q(u) = (1-u)^{q-1}[2 + q\log(1-u)], \quad 0 < u < 1. \tag{4}$$

Equivalently,

$$D_q(F) = \int_0^1 Q(u) J_q(u) du. \tag{5}$$

The score function $J_q(u)$ is induced by the Tsallis-weighted population deviation. It has a single zero at $u_q = 1 - \exp(-2/q)$; hence, $J_q(u)$ is positive below u_q and negative above it. This sign change does not contradict the IFR direction of $D_q(F)$, because the latter integrates the non-negative local kernel $\eta(t)$ against $S(t)^q$. It does, however, affect finite-sample signal-to-noise behavior. For $q < 1$, the score is unbounded near the upper endpoint and endpoint quadrature may converge slowly; for larger q, cancelation occurs over a wider upper-tail region. These facts require a prespecified, numerically stable grid rather than an unrestricted search over q.

The tuning parameter q should not be interpreted as an estimated proportional-hazards coefficient. Within the proposed test family, it serves as a prespecified Tsallis order whose equivalent power-survival representation scales the transformed hazard by a factor of q. This single index determines the geometry of the induced score function and, consequently, the regions of the lifetime distribution that receive the greatest directional emphasis. Smaller values of q concentrate sensitivity more strongly toward the upper tail, whereas larger values distribute the weighting more broadly across the support and increase the range over which positive and negative score contributions may offset one another. Consequently, the construction of the adaptive tuning grid is guided by a balance between geometric diversity and numerical stability rather than by interpreting q as an unknown physical or scientific hazard ratio.

As illustrated in Figure 1, very small values of q produce increasingly steep endpoint behavior, making the resulting statistic more sensitive to finite-sample plotting positions and upper-tail variability. Conversely, progressively larger values of q generate broader cancelation regions and increasingly correlated neighboring score components. These structural considerations motivate the use of a moderate collection of representative score geometries, specified a priori, before examining the performance of the procedure under any particular alternative.

Let $X_{1:n}, \ldots, X_{n:n}$ be the order statistics from a random sample $X_1, \ldots, X_n$. Because the score in (4) is defined on the open interval $(0,1)$, endpoint-stabilized plotting positions are used. The fixed-q empirical statistic is defined by

$$V_n(q) = \frac{1}{n}\sum_{i=1}^{n} J_q(p_{i,n}) X_{i:n}, \quad p_{i,n} = \frac{i}{n+1}, \quad i = 1, \ldots, n.$$

To remove the effect of the unknown scale parameter, the statistic is normalized by the sample mean $\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$, producing the scale-free statistic

$$V_n^\star(q) = \frac{V_n(q)}{\overline{X}}, \quad q > 0. \tag{6}$$

The statistic in (6) is scale-invariant. Because endpoint stabilization induces a finite-sample null mean that need not be exactly zero, inference is based on exact null centering and scaling derived from the joint spacing representation in Section 3. Large positive standardized values provide evidence in the IFR direction. The development is for complete samples; censored-data extensions require a separate spacing representation and calibration.

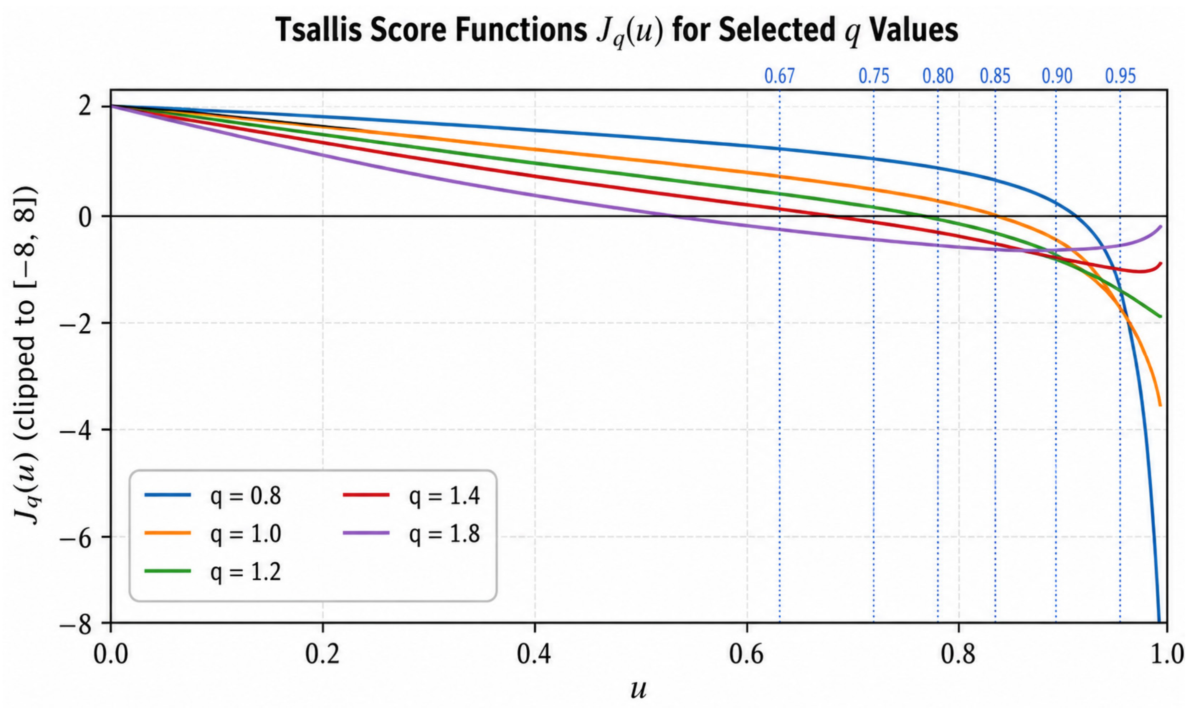

Figure 1. Score functions $J_q(u)$ for representative tuning values; vertical dotted lines mark $u_q = 1 - \exp(-2/q)$. Blue, orange, green, red, and purple solid curves correspond to q = 0.8, 1.0, 1.2, 1.4, and 1.8, respectively.

3. Exact Null Distribution, Asymptotic Theory, and Adaptive Calibration

This section develops the distributional theory underlying the proposed adaptive procedure. We first establish an exact finite-sample characterization of the fixed-q statistic by expressing it in terms of normalized spacings under the exponential model. We then derive its asymptotic properties using classical L-statistic theory. Finally, the exactly standardized fixed-q statistics are combined over a prespecified finite tuning grid to construct the adaptive max-type statistic employed throughout the simulation study and real-data applications.

3.1. Exact Finite-Sample Null Distribution for a Fixed Tuning Parameter

Let the ordered sample be denoted by the order statistics corresponding to a random sample of size n. The initial value is taken to be zero. The normalized spacings are defined by

$$D_r = (n - r + 1)(X_{r:n} - X_{r-1:n}), \qquad r = 1, \ldots, n. \tag{7}$$

For a fixed positive tuning value, the endpoint-stabilized score used in Section 2 gives the following spacing representation of the scale-invariant statistic:

$$V_n^{\star}(q) = \frac{\sum_{r=1}^{n} b_{r,n}(q) D_r}{\sum_{r=1}^{n} D_r} \tag{8}$$

The coefficients in (8) are given by

$$b_{r,n}(q) = \frac{1}{n-r+1}\sum_{i=r}^{n} J_q(p_{i,n}), \quad p_{i,n} = \frac{i}{n+1}, \quad r = 1,\ldots,n. \tag{9}$$

The endpoint-stabilized plotting positions defined in (9) are consistent with the construction introduced in Section 2 and avoid evaluating the score function at u = 1.

Theorem 1. *Suppose that the observations are independent and follow an exponential distribution. For fixed q and n, assume that the coefficients in (9) are pairwise distinct. Then, under the null hypothesis,*

$$Pr\{V_n^\star(q) \le x\} = 1 - \sum_{\ell=1}^{n}\left[\prod_{\substack{m=1\\ m\neq\ell}}^{n} \frac{b_{\ell,n}(q) - x}{b_{\ell,n}(q) - b_{m,n}(q)}\right] I\{x \le b_{\ell,n}(q)\}. \tag{10}$$

If some of the coefficients coincide, the distribution is obtained by the corresponding limiting form.

Proof of Theorem 1. From the definition of the normalized spacings, the weighted statistic in Section 2 can be written as a linear combination of the normalized spacings. Specifically,

$$nV_n(q) = \sum_{r=1}^{n} b_{r,n}(q) D_r = \sum_{i=1}^{n} J_q(p_{i,n}) X_{i:n}. \tag{11}$$

Division by the total normalized spacing yields the statistic in (8). Under the null hypothesis, the normalized spacings are independent exponential random variables sharing a common scale parameter. Since the statistic is scale-invariant, this common scale has no effect on its distribution. Equivalently, after a simple scale normalization, the spacings may be represented as independent chi-square random variables with two degrees of freedom, a representation that is convenient for the subsequent theoretical development. Therefore,

$$Pr\{V_n^\star(q) \le x\} = 1 - Pr\left\{\sum_{r=1}^{n}[b_{r,n}(q) - x]D_r > 0\right\}. \tag{12}$$

Applying the standard distributional result for the ratio of a linear combination of independent chi-square random variables to their sum immediately yields (10). The unknown exponential scale parameter plays no role in this derivation because it cancels in the scale-invariant ratio defining the statistic. □

3.2. Exact Joint Finite-Sample Null Structure and Standardization

The fixed-q law in Theorem 1 extends naturally to the complete vector used by the adaptive procedure. Let $\mathcal{Q} = \{q_1,\ldots,q_K\}$ be a fixed finite grid, let $B_n = [b_{r,n}(q_j)]$ be the $n \times K$ coefficient matrix, and define $P_{r,n} = D_r / \sum_{s=1}^{n} D_s$.

$$\mathbf{P}_n = (P_{1,n},\ldots,P_{n,n})^{\mathrm{T}} \sim \mathrm{Dirichlet}_n(1,\ldots,1), \qquad H_0.$$

Theorem 2. *Under H_0, the vector of scale-free fixed-q statistics satisfies*

$$\boldsymbol{V}_n^\star = B_n^{\mathrm{T}} \boldsymbol{P}_n. \tag{13}$$

Consequently, its exact mean vector and covariance matrix have the elements given in (14). $m_n \boldsymbol{\Omega}_n$

$$m_{n,j} = \frac{1}{n}\sum_{r=1}^{n} b_{r,n}(q_j), \quad \omega_{n,jk} = \frac{n\sum_{r=1}^{n} b_{r,n}(q_j) b_{r,n}(q_k) - (\sum_{r=1}^{n} b_{r,n}(q_j))(\sum_{r=1}^{n} b_{r,n}(q_k))}{n^2(n+1)}. \tag{14}$$

Proof of Theorem 2. Under exponentiality, the normalized spacings $D_1, \dots, D_n$ are independent exponential variables with a common scale. Dividing them by their sum gives a Dirichlet$(1, \dots, 1)$ vector, independent of the total spacing. The vector representation follows from (8), and the stated moments follow from the standard Dirichlet moments.

Define $s_{n,j} = \sqrt{\omega_{n,jj}}$, $Z_n(q_j) = \{V_n^\star(q_j) - m_{n,j}\}/s_{n,j}$, and $T_n^{\max} = \max_{1 \le j \le K} Z_n(q_j)$. The finite-sample null law of the adaptive statistic is therefore fully determined by the Dirichlet probability

$$Pr\{T_n^{\max} \le t\} = (n-1)! \int_{\Delta_{n-1}} \mathbf{1}\Big\{B_n^{\mathrm{T}} \mathbf{p} \le m_n + t\mathbf{s}_n\Big\}\, d\mathbf{p},$$

where Δ_{n-1} is the unit simplex and the displayed inequality is componentwise. The integral generally has no convenient closed form, but it is an exact theoretical characterization. Simulation from standard exponential spacings, or directly from the Dirichlet law, is a numerical method for evaluating this probability. □

3.3. Joint Asymptotic Distribution and Consistency

The finite-sample result is complemented by a joint large-sample limit. Put $\theta_j(F) = D_{q_j}(F)$, $\mu = E(X)$, $d_j(F) = \theta_j(F)/\mu$, and $J_0(u) = 1$. For score functions a and b, define

$$\Gamma_F(a, b) = \int_0^\infty \int_0^\infty a(F(x)) b(F(y)) \{F(\min(x, y)) - F(x)F(y)\}\, dx\, dy.$$

Theorem 3. *Let the tuning grid be fixed and finite. Assume that F is continuous, 0 < μ = E(X) < ∞, E(X*^{2}*) < ∞, and that every score function in the grid satisfies the regularity and integrability conditions for the multivariate central limit theorem for L-statistics. In particular, all covariance-kernel integrals displayed below are finite, and the resulting covariance matrix is nonsingular [14]. Then*

$$\sqrt{n}\{\mathbf{V}_n^\star - \mathbf{d}(F)\} \xrightarrow{\mathcal{L}} N_K(\mathbf{0}, \boldsymbol{\Sigma}_F),$$

where

$$(\Sigma_F)_{jk} = \mu^{-2}\Big[\Gamma_F\Big(J_{q_j}, J_{q_k}\Big) - d_k \Gamma_F\Big(J_{q_j}, J_0\Big) - d_j \Gamma_F(J_0, J_{q_k}) + d_j d_k \Gamma_F(J_0, J_0)\Big].$$

Proof of Theorem 3. Apply the multivariate central limit theorem for L-statistics to the K weighted statistics jointly with the sample mean [14] and then apply the multivariate delta method to the componentwise ratios. Under H_0, the population departures are zero for all grid components, so the covariance simplifies to the null expression displayed above. □

Corollary 1. *If the finite-sample correlation matrices converge to* $\boldsymbol{R}_0$ *under* H_0 *, then the standardized vector converges to* $N_K(\mathbf{0}, \boldsymbol{R}_0)$ *and*

$$T_n^{max} \xrightarrow{\mathcal{L}} \max_{1 \le j \le K} Z_j, \quad Pr\Big\{\max_j Z_j \le t\Big\} = \Phi_K(t\mathbf{1}_K; \boldsymbol{R}_0).$$

Theorem 4. *For any fixed alternative F such that $d_j(F) > 0$ for at least one $q_j \in \mathcal{Q}$, the Adaptive CRE-PH Test is consistent; specifically, $T_n^{max} \to \infty$ in probability and its rejection probability tends to one.*

Proof of Theorem 4. For a component with $d_j(F) > 0$, consistency of the L-statistic gives $V_n^{\star}(q_j) \to d_j(F) > 0$ in probability. Moreover, the exact null centering satisfies $m_{n,j} \to 0$, while the null standard deviation satisfies $s_{n,j} = O\left(n^{-1/2}\right)$. Therefore, $Z_n(q_j) \to \infty$ in probability. Since $T_n^{\max} \geq Z_n(q_j)$, the same conclusion follows for the maximum. □

Remark 1 (Adaptive local asymptotic). *The consistency result established above applies to fixed alternatives. A rigorous treatment of contiguous alternatives would require a joint local asymptotic expansion of the complete standardized score vector prior to the maximization step. Under such local alternatives, the noncentral mean vector, covariance structure, and even the identity of the maximizing component may depend on the underlying local direction. Consequently, the local asymptotic power of the adaptive maximum cannot be derived by applying Pitman-efficiency arguments separately to the individual fixed-q statistics. Instead, it requires a dedicated analysis of the limiting noncentral multivariate Gaussian distribution together with the distribution of its maximum. The exact joint null theory and multivariate asymptotic representation developed in this paper provide the probabilistic foundation for such an extension. A comprehensive investigation of local asymptotic power and asymptotic relative efficiency is beyond the scope of the present work and is left for future research.*

3.4. A Stability-Based Design Principle for Adaptive Tsallis Grids

The adaptive maximum requires only a finite collection of representative score geometries rather than a dense discretization of the tuning-parameter space. To preserve the structural nature of the proposed procedure and avoid retrospective tuning, the adaptive grid is specified a priori, independently of the power comparisons presented in Section 4. Its construction is guided by the following three design principles.

Principle 1 (Endpoint stability). *Values exhibiting excessive endpoint sensitivity are excluded to ensure that the endpoint-stabilized score representation remains numerically stable for the finite sample sizes considered.*

Principle 2 (Directional diversity). *Successive grid points are chosen to represent genuinely distinct score geometries, characterized by different zero-crossing locations and weighting profiles. Closely spaced tuning values contribute highly similar score functions and therefore provide little additional directional information.*

Principle 3 (Multiplicity economy). *The tuning grid is enlarged only when an additional component contributes substantial geometric diversity. Otherwise, highly correlated score components merely inflate the critical value of the adaptive maximum without providing a commensurate increase in independent directional information.*

These principles lead to the four-point tuning grid adopted throughout the paper. The resulting grid is determined entirely by structural and numerical considerations and should not be interpreted as an empirically optimized or minimax-optimal choice. Rather, it provides a parsimonious collection of representative score geometries that supports stable and interpretable adaptive inference.

Robustness with respect to the choice of the tuning grid is ensured at the calibration stage because Theorem 2 applies to any prespecified finite collection of admissible tuning values. A smaller grid may fail to capture representative score geometries, whereas enlarging the grid with closely spaced or more extreme values may introduce endpoint instability or inflate the adaptive critical value through multiplicity without providing a commensurate gain in independent directional information. Consequently, the adopted grid should be viewed as a principled, moderate design rather than a uniquely optimal choice. Any alternative prespecified finite grid can be calibrated exactly using the same Dirichlet-based procedure; throughout this paper, we adopt $\mathcal{Q}_A = \{0.8, 1.0, 1.2, 1.4\}$.

For each sample size, the exact null centering and scaling constants are computed from Theorem 2. The observed adaptive maximum is then compared with the upper-tail critical value obtained from Monte Carlo simulation of the exact Dirichlet null representation. Equivalently, the Monte Carlo p-value is estimated as the proportion of simulated null maxima that are at least as large as the observed adaptive maximum. Because all fixed-q components are evaluated using the same Dirichlet samples, their dependence structure is preserved automatically throughout the calibration. This calibration strategy preserves the structural a priori construction of the tuning grid while keeping it entirely separate from the subsequent assessment of power under specific alternatives.

Supplementary File S1 contains illustrative R scripts for a representative fixed-q simulation under a linear-failure-rate alternative, together with exponential Q–Q and TTT graphical diagnostics. The complete adaptive calibration and inference procedure is developed and specified mathematically in Sections 3 and 5 of the main text.

4. Monte Carlo Validation of the Stability-Based Grid Design

This section evaluates the finite-sample performance of the fixed-q statistics developed in Sections 2 and 3. Its objective is not to select or refine the adaptive tuning grid retrospectively. Instead, the simulation study provides an independent empirical assessment of whether the a priori structural design captures genuinely distinct sensitivity profiles while avoiding unnecessary redundancy among neighboring components. The investigation therefore examines both directional power and the extent to which no single fixed-q statistic performs uniformly best across a representative collection of IFR alternatives.

4.1. Simulation Design

The simulation study is conducted under the exponential null model together with three representative IFR alternatives: the linear failure rate (LFR), Makeham, and Weibull models. The broader exploratory tuning grid is defined as $\mathcal{Q}_S = \{0.8, 1.0, 1.2, 1.4, 1.6, 1.8\}$; where values below 0.8 are excluded because the corresponding endpoint-stabilized score functions exhibit slow finite-sample convergence. The adaptive procedure investigated in Section 5 is based on the more parsimonious grid $\mathcal{Q}_A = \{0.8, 1.0, 1.2, 1.4\}$, which was selected a priori according to the structural design principles established in Section 3.

The LFR distribution is generated from the survival function

$$S_1(x) = e^{-(x+0.5\alpha x^2)}, \, x > 0, \alpha > 0.$$

The Makeham distribution is generated from

$$S_2(x) = \exp\{-x - \alpha\left(x + e^{-x} - 1\right)\}, \qquad x > 0, \, \alpha > 0.$$

The Weibull distribution is generated from

$$\mathrm{S}(x) = e^{-x^{\alpha}}, \, x > 0, \alpha > 0.$$

In each case, the exponential model is recovered at the no-aging boundary, while larger alpha values represent stronger IFR departures. Fixed-q experiments use $n = 20$, 30, 40, and 50. For each n and q, upper-tail critical values are estimated from 50,000 standard exponential samples using the exact finite-sample centering and scaling in Theorem 2. Each reported power is based on 20,000 independent samples. The adaptive experiments use 100,000 null samples for critical values and 20,000 samples per alternative.

All tests are calibrated at the nominal 5% level. A fixed-q test rejects for a large standardized component, while the Adaptive CRE-PH Test rejects for a large value of the maximum. Random-number streams were prespecified and separated by task: calibration samples, empirical-size samples, and alternative samples were generated with independent seeds. All tables and figures were produced under the same implementation protocol.

To further assess finite-sample calibration, empirical rejection probabilities under the exponential null were monitored for every sample size considered. Across all settings, the observed rejection frequencies remained close to the nominal 5% significance level, indicating accurate finite-sample size control. Since the empirical sizes were estimated from 20,000 independent Monte Carlo samples, the Monte Carlo standard error at the nominal level is approximately $\sqrt{0.05(0.95)/20,000}$. Consequently, deviations of the order observed in the simulation study are consistent with ordinary Monte Carlo sampling variability and do not indicate systematic miscalibration of the proposed procedure.

Accordingly, the simulation results corroborate the exact finite-sample calibration established in Section 3 and demonstrate that the proposed adaptive procedure maintains the nominal significance level consistently across the range of sample sizes considered.

4.2. Fixed-q Power Under LFR Alternatives

Table 1 and Figure 2 report the fixed-q power under LFR alternatives. Power increases with n and α. Across the examined settings, $q = 1.2$ is typically best, with $q = 1.0$ and $q = 1.4$ close behind. The endpoints $q = 0.8$ and $q = 1.8$ are generally weaker, showing that neither the smallest nor the largest candidate is uniformly preferable.

Table 1. Empirical power of the exactly standardized fixed-q statistic for the LFR distribution at $\gamma = 0.05$.

Dist.	α	q **= 0.8**	q **= 1.0**	q **= 1.2**	q **= 1.4**	q **= 1.6**	q **= 1.8**
	1.2	0.3265	0.3341	0.3352	0.3307	0.3236	0.3118
	1.5	0.3679	0.3769	0.3786	0.3740	0.3649	0.3527
$n = 20$	2.0	0.4251	0.4339	0.4348	0.4300	0.4193	0.4048
	2.5	0.4723	0.4833	0.4878	0.4845	0.4763	0.4583
	3.0	0.5080	0.5212	0.5244	0.5203	0.5100	0.4950
	1.2	0.4477	0.4611	0.4621	0.4566	0.4430	0.4188
	1.5	0.5047	0.5217	0.5241	0.5202	0.5070	0.4812
$n = 30$	2.0	0.5749	0.5935	0.5964	0.5936	0.5788	0.5545
	2.5	0.6370	0.6594	0.6646	0.6606	0.6464	0.6220
	3.0	0.6813	0.7006	0.7053	0.7005	0.6894	0.6634
	1.2	0.5585	0.5735	0.5761	0.5617	0.5429	0.5150
	1.5	0.6297	0.6455	0.6462	0.6333	0.6115	0.5828
$n = 40$	2.0	0.7104	0.7272	0.7311	0.7175	0.6997	0.6716
	2.5	0.7658	0.7863	0.7927	0.7811	0.7641	0.7401
	3.0	0.8083	0.8238	0.8279	0.8175	0.8030	0.7772
	1.2	0.6562	0.6809	0.6865	0.6753	0.6454	0.6070
	1.5	0.7229	0.7470	0.7543	0.7446	0.7155	0.6794
$n = 50$	2.0	0.7991	0.8220	0.8282	0.8199	0.7956	0.7648
	2.5	0.8519	0.8720	0.8764	0.8699	0.8491	0.8221
	3.0	0.8858	0.9016	0.9079	0.9032	0.8864	0.8620

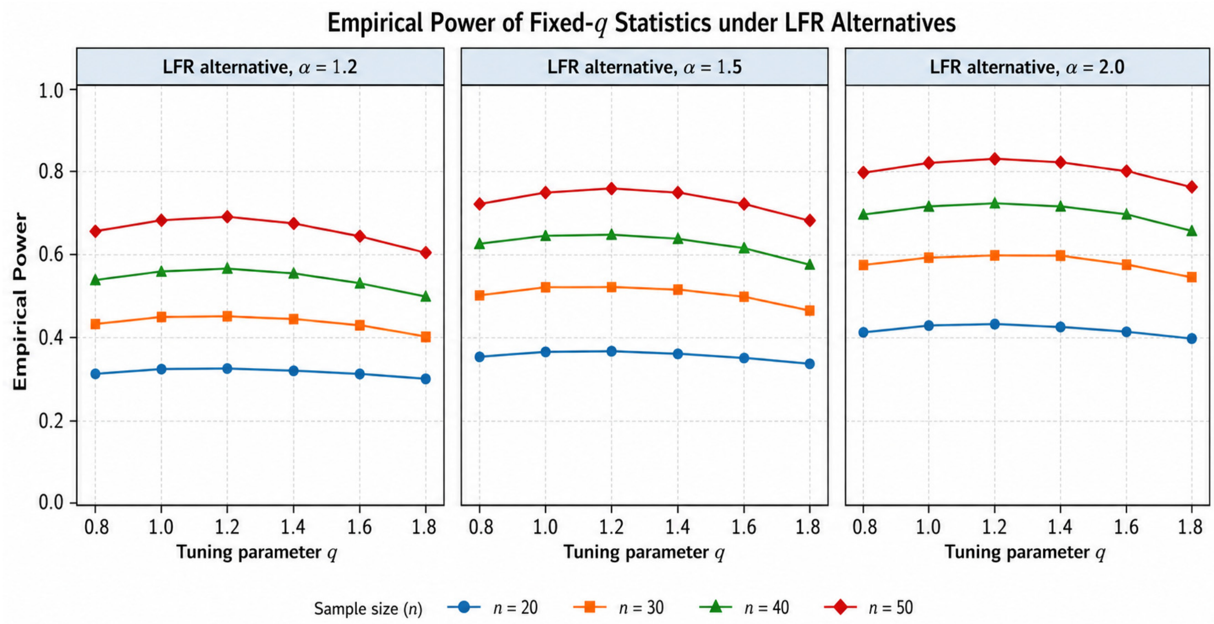

Figure 2. Empirical power over $q \in \{0.8, 1.0, \ldots, 1.8\}$ for $n = 20$, 30, 40, and 50 under the LFR distribution, with $\alpha = 1.2, 1.5,$ and 2.0. Blue circles, orange squares, green triangles, and red diamonds denote n = 20, 30, 40, and 50, respectively; solid lines connect the empirical power values across q.

The LFR experiment therefore favors a moderate tuning value. For example, at $n = 50$ and $\alpha = 2.0$, powers for $q = 0.8$, 1.0, 1.2, 1.4, 1.6, and 1.8 are 0.799, 0.822, 0.828, 0.820, 0.796, and 0.765, respectively. This relatively flat peak around $q = 1.2$ supports a compact adaptive grid centered on moderate values.

4.3. Fixed-q Power Under Makeham Alternatives

Table 2 and Figure 3 present the Makeham results. These alternatives are more difficult to detect, especially for small n and weak aging. The best performance shifts toward $q = 1.4$ or $q = 1.6$, although neighboring values remain close. Thus, the preferred tuning value differs from the LFR case.

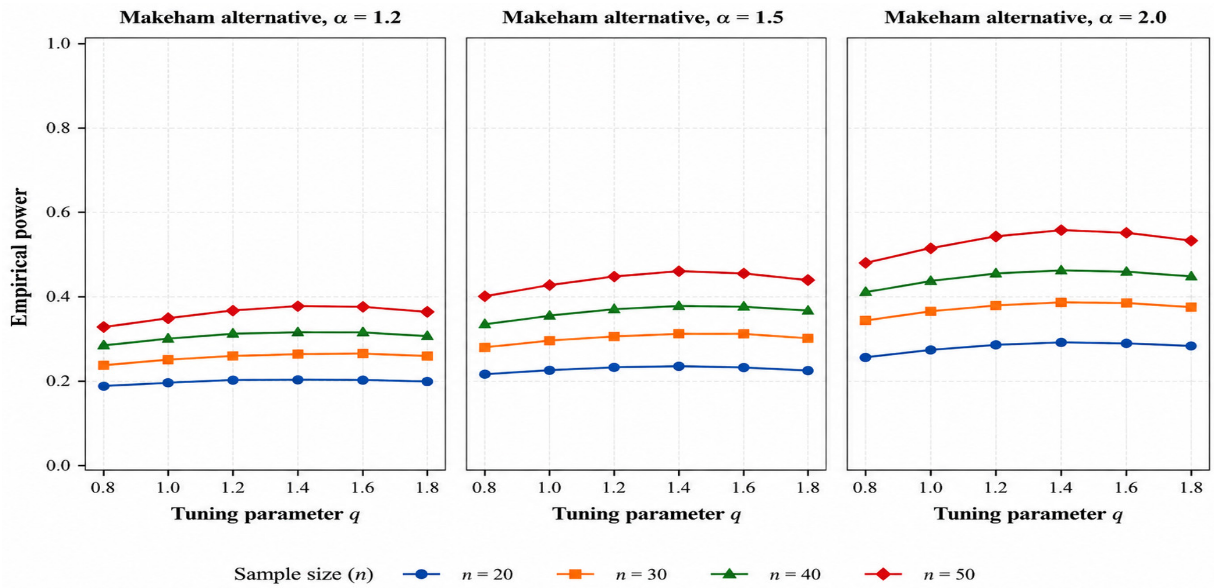

Figure 3. Empirical power over $q \in \{0.8, 1.0, \ldots, 1.8\}$ for $n = 20$, 30, 40, and 50 under the Makeham distribution, with $\alpha = 1.2, 1.5,$ and 2.0. Blue circles, orange squares, green triangles, and red diamonds denote n = 20, 30, 40, and 50, respectively; solid lines connect the empirical power values across q.

Table 2. Empirical power of the exactly standardized fixed-q statistic for the Makeham distribution at $\gamma = 0.05$.

Dist.	α	q = 0.8	q = 1.0	q = 1.2	q = 1.4	q = 1.6	q = 1.8
$n = 20$	1.2	0.1862	0.1917	0.1972	0.1983	0.1987	0.1961
	1.5	0.2190	0.2266	0.2306	0.2322	0.2299	0.2248
	2.0	0.2575	0.2676	0.2758	0.2782	0.2769	0.2726
	2.5	0.3026	0.3147	0.3234	0.3261	0.3245	0.3215
	3.0	0.3441	0.3578	0.3679	0.3719	0.3685	0.3646
$n = 30$	1.2	0.2364	0.2512	0.2594	0.2633	0.2658	0.2615
	1.5	0.2867	0.3026	0.3113	0.3170	0.3191	0.3079
	2.0	0.3555	0.3800	0.3912	0.3991	0.3978	0.3882
	2.5	0.4138	0.4395	0.4526	0.4603	0.4587	0.4464
	3.0	0.4670	0.4949	0.5110	0.5216	0.5219	0.5060
$n = 40$	1.2	0.2903	0.3075	0.3204	0.3215	0.3224	0.3142
	1.5	0.3443	0.3696	0.3856	0.3928	0.3907	0.3825
	2.0	0.4299	0.4587	0.4773	0.4833	0.4813	0.4709
	2.5	0.4999	0.5315	0.5524	0.5575	0.5582	0.5452
	3.0	0.5621	0.5979	0.6206	0.6259	0.6207	0.6069
$n = 50$	1.2	0.3306	0.3603	0.3826	0.3918	0.3890	0.3769
	1.5	0.4080	0.4432	0.4659	0.4749	0.4701	0.4552
	2.0	0.5050	0.5484	0.5772	0.5853	0.5794	0.5601
	2.5	0.5777	0.6237	0.6522	0.6630	0.6572	0.6361
	3.0	0.6518	0.6993	0.7236	0.7331	0.7214	0.7028

The Makeham pattern confirms that the tuning parameter changes the sensitivity profile. A value optimized for LFR need not be optimal for Makeham departures, but the useful region remains moderate and the gains from $q > 1.4$ are small. This observation motivates including $q = 1.4$ in the adaptive grid while excluding more extreme components from the maximum.

4.4. Fixed-q Power Under Weibull Alternatives

Table 3 and Figure 4 summarize the Weibull results. For mild departures, power tends to increase through $q = 1.6$ or $q = 1.8$; for $\alpha \geq 2$, all moderate and large q values rapidly approach unit power. Thus, larger q values are especially useful for mild Weibull-type wear-out, whereas differences among moderate and large q values become negligible under stronger departures.

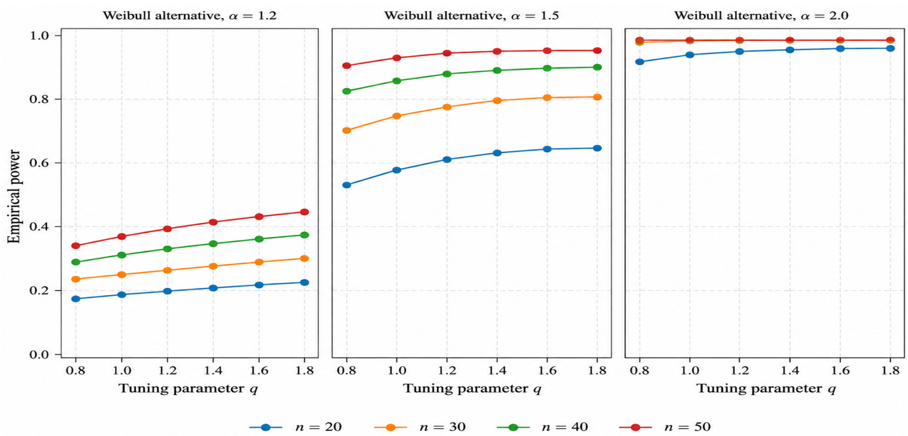

Figure 4. Empirical power over $q \in \{0.8, 1.0, \ldots, 1.8\}$ for $n = 20$, 30, 40, and 50 under the Weibull distribution, with $\alpha = 1.2, 1.5,$ and 2.0. Blue circles, orange squares, green triangles, and red diamonds denote n = 20, 30, 40, and 50, respectively; solid lines connect the empirical power values across q.

Table 3. Empirical power of the exactly standardized fixed-q statistic for the Weibull distribution at $\gamma = 0.05$.

Dist.	α	q = 0.8	q = 1.0	q = 1.2	q = 1.4	q = 1.6	q = 1.8
	1.2	0.1872	0.1973	0.2031	0.2097	0.2124	0.2140
	1.5	0.5636	0.5961	0.6192	0.6353	0.6411	0.6427
$n = 20$	2.0	0.9571	0.9696	0.9761	0.9790	0.9802	0.9793
	2.5	0.9993	0.9997	0.9998	0.9998	0.9998	0.9997
	3.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.2	0.2443	0.2648	0.2778	0.2917	0.2980	0.2970
	1.5	0.7300	0.7764	0.8028	0.8205	0.8294	0.8258
$n = 30$	2.0	0.9960	0.9984	0.9990	0.9991	0.9991	0.9991
	2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	3.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.2	0.2953	0.3195	0.3411	0.3527	0.3622	0.3654
	1.5	0.8415	0.8816	0.9062	0.9169	0.9220	0.9217
$n = 40$	2.0	0.9993	0.9996	0.9997	0.9998	0.9998	0.9999
	2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	3.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.2	0.3338	0.3729	0.4017	0.4232	0.4320	0.4335
	1.5	0.9079	0.9416	0.9589	0.9656	0.9670	0.9640
$n = 50$	2.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	3.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

The Weibull results show that $q = 1.8$ can be useful for mild wear-out, but it adds little once the departure is moderate. Its inclusion in a fixed-q sensitivity study is informative, whereas including it in the adaptive maximum produces an extra multiplicity cost with limited incremental benefit.

4.5. Summary of Tuning-Parameter Sensitivity

The fixed-q simulations lead to a clear structural conclusion. The LFR alternatives exhibit their strongest performance for tuning values around $q = 1.2$, the Makeham alternatives favor values near $q = 1.4$–1.6, and the milder Weibull alternatives attain their highest power for q around 1.6–1.8. Thus, no individual fixed-q statistic is uniformly optimal across the representative IFR models considered, although the most informative sensitivity region is consistently concentrated around moderate tuning values.

These empirical findings validate, rather than determine, the proposed adaptive design. The a priori tuning grid captures the principal moderate score geometries while excluding components that either exhibit excessive endpoint sensitivity or contribute little additional geometric diversity. Accordingly, the fixed-q power results should be interpreted as an independent post-design validation of the structural grid rather than as a data-driven tuning-selection procedure.

5. Performance of the Adaptive CRE-PH Test and Comparison with Existing Procedures

This section evaluates the finite-sample performance of the Adaptive CRE-PH Test and compares it with existing nonparametric procedures for testing exponentiality against IFR alternatives. The objective is not merely to determine whether the adaptive statistic achieves high power, but to assess whether it provides robust and reliable performance when the underlying form of positive aging is unknown.

5.1. Adaptive Statistic Used in the Simulation Study

The adaptive statistic is computed from the exact finite-sample standardized components in Theorem 2. The prespecified grid is the four-point set introduced in Section 3.4. Its members represent distinct moderate score geometries while avoiding the most endpoint-sensitive and highly redundant candidates. The fixed-q study in Section 4 is used only as a post-design diagnostic of this structural choice, not as a retrospective grid-selection rule $\mathcal{Q}_A = \{0.8, 1.0, 1.2, 1.4\}$.

For each $q \in \mathcal{Q}_A$, $Z_n(q)$ uses the exact null mean and variance in Theorem 2, and $T_n^{\max}$ is their maximum. Critical values are generated from the exact Dirichlet null representation, so dependence among the four components is retained.

The adaptive calibration is simulation-based only in its numerical implementation: the distribution being simulated is the exact joint finite-sample null law established in Theorem 2. Consequently, no independence approximation, Bonferroni adjustment, or asymptotic calibration is required. The mathematical development of the adaptive calibration is presented in Section 3, whereas Supplementary File S1 provides only illustrative R scripts for a representative fixed-q simulation together with the graphical diagnostics used in the real-data applications.

5.2. Empirical Size of the Adaptive CRE-PH Test

The first step of the simulation study is to verify the numerical implementation of the exact finite-sample null distribution developed in Section 3. The critical values reported in Table 4 are obtained from a prespecified principal calibration stream consisting of 100,000 independent standard exponential samples, whereas the empirical significance levels are evaluated using an independent set of 50,000 samples. To assess the numerical stability of the calibration, the section on Monte Carlo Calibration Stability further considers ten additional independent calibration streams generated with different random seeds.

Table 4. Empirical size of the Adaptive CRE-PH Test at $\gamma = 0.05$.

n	Critical Value	Empirical Size	MC SE	95% CI
25	1.6812	0.0481	0.0010	(0.0462, 0.0500)
50	1.7128	0.0493	0.0010	(0.0474, 0.0512)
100	1.7491	0.0494	0.0010	(0.0475, 0.0513)

Note. The adaptive grid is $\mathcal{Q}_A = \{0.8, 1.0, 1.2, 1.4\}$. Critical values were estimated from 100,000 null samples, and empirical sizes were evaluated using an independent set of 50,000 null samples. The approximate 95% confidence intervals are calculated as the empirical size plus or minus 1.96 times its Monte Carlo standard error. $\mathcal{Q}_A = \{0.8, 1.0, 1.2, 1.4\}$.

The empirical sizes are close to 0.05 for all sample sizes, and the nominal level lies within every approximate 95% Monte Carlo confidence interval reported in Table 4. These results demonstrate that exact joint centering and scaling, combined with Dirichlet Monte Carlo evaluation of the adaptive maximum, successfully controls the overall Type I error rate at the nominal significance level. The repeated-seed analysis presented below further confirms that this conclusion is robust and does not depend on any particular calibration stream.

Monte Carlo Calibration Stability

To assess numerical reproducibility, the complete adaptive calibration was repeated ten times for each sample size using independent random seeds. Each calibration run used 100,000 draws from the exact Dirichlet null representation. For every run, empirical size was then evaluated on a separate independent sample of 50,000 null draws. The results are summarized in Table 5.

Table 5. Stability of the adaptive critical values and empirical sizes over ten independent Monte Carlo calibration runs.

n	Critical Value: Mean (SD)	Critical-Value Range	Empirical Size: Mean (SD)	Empirical-Size Range
25	1.6695 (0.0065)	1.6587–1.6798	0.0505 (0.0010)	0.0489–0.0524
50	1.7068 (0.0051)	1.6971–1.7129	0.0495 (0.0011)	0.0482–0.0511
100	1.7431 (0.0048)	1.7322–1.7488	0.0490 (0.0010)	0.0472–0.0506

Each critical value is estimated as the 95th percentile of 100,000 independent draws from the exact Dirichlet null distribution. The empirical significance level is evaluated using a separate set of 50,000 independent null draws for each calibration run. The reported standard deviations measure the variability of the estimated critical values across independent calibration streams.

Across all sample sizes, the between-stream standard deviation of the estimated critical value never exceeds 0.0065, while the corresponding empirical significance levels remain consistently close to the nominal level of 0.05. These results indicate that repeated calibration introduces only negligible numerical variability and has no practically meaningful effect on the Type I error rate. Consequently, the Monte Carlo implementation provides a stable and reproducible numerical evaluation of the exact Dirichlet null distribution.

5.3. Power Comparison Between the Adaptive and Selected Fixed-q Tests

The next comparison diagnoses the multiplicity cost directly. For each alternative, the adaptive statistic is compared with all four members of $\mathcal{Q}_A$. The adaptive procedure is not expected to dominate the best fixed component, which is an oracle selected after the alternative is known. A meaningful stability criterion is that the adaptive power remains between the weakest and strongest grid components and close to the oracle maximum.

Tables 6–8 demonstrate the anticipated performance of the proposed adaptive procedure. Across all reported scenarios, the adaptive power consistently lies between the weakest and strongest fixed-q components and, in most cases, remains within only a few percentage points of the oracle maximum. These findings indicate that the adaptive procedure incurs only a modest multiplicity penalty while avoiding the poor performance associated with any single unfavorable tuning choice. Moreover, the restricted a priori grid mitigates the loss of power that would otherwise arise from including weak or highly redundant components.

Table 6. Power of the Adaptive CRE-PH Test and its four fixed-q components for $n = 25$ and $\gamma = 0.05$.

Dist.	α	$q = 0.8$	$q = 1.0$	$q = 1.2$	$q = 1.4$	Adaptive
LFR	1.2	0.3795	0.3910	0.3929	0.3885	0.3859
	1.5	0.4330	0.4479	0.4463	0.4427	0.4415
	2.0	0.5068	0.5203	0.5234	0.5187	0.5179
	2.5	0.5595	0.5770	0.5807	0.5762	0.5734
	3.0	0.5993	0.6160	0.6213	0.6186	0.6161
Makeham	1.2	0.2151	0.2254	0.2311	0.2331	0.2309
	1.5	0.2449	0.2592	0.2681	0.2740	0.2691
	2.0	0.3047	0.3236	0.3301	0.3357	0.3305
	2.5	0.3522	0.3715	0.3833	0.3906	0.3827
	3.0	0.3970	0.4203	0.4330	0.4411	0.4330
Weibull	1.2	0.2089	0.2248	0.2352	0.2439	0.2372
	1.5	0.6444	0.6887	0.7164	0.7374	0.7244
	2.0	0.9843	0.9908	0.9933	0.9947	0.9939
	2.5	1.0000	1.0000	1.0000	1.0000	1.0000
	3.0	1.0000	1.0000	1.0000	1.0000	1.0000

Table 7. Power of the Adaptive CRE-PH Test and its four fixed-q components for $n = 50$ and $\gamma = 0.05$.

Dist.	α	q = 0.8	q = 1.0	q = 1.2	q = 1.4	Adaptive
LFR	1.2	0.6495	0.6728	0.6731	0.6586	0.6627
	1.5	0.7259	0.7468	0.7480	0.7319	0.7386
	2.0	0.7997	0.8214	0.8239	0.8118	0.8175
	2.5	0.8503	0.8699	0.8710	0.8618	0.8653
	3.0	0.8858	0.9030	0.9074	0.8991	0.9020
Makeham	1.2	0.3350	0.3624	0.3762	0.3849	0.3790
	1.5	0.4035	0.4360	0.4526	0.4617	0.4543
	2.0	0.5011	0.5435	0.5645	0.5718	0.5638
	2.5	0.5872	0.6281	0.6493	0.6561	0.6492
	3.0	0.6478	0.6891	0.7111	0.7180	0.7108
Weibull	1.2	0.3379	0.3739	0.4014	0.4205	0.4079
	1.5	0.9035	0.9372	0.9545	0.9634	0.9576
	2.0	1.0000	1.0000	1.0000	1.0000	1.0000
	2.5	1.0000	1.0000	1.0000	1.0000	1.0000
	3.0	1.0000	1.0000	1.0000	1.0000	1.0000

Table 8. Power of the Adaptive CRE-PH Test and its four fixed-q components for $n = 100$ and $\gamma = 0.05$.

Dist.	α	q = 0.8	q = 1.0	q = 1.2	q = 1.4	Adaptive
LFR	1.2	0.9174	0.9324	0.9313	0.9161	0.9258
	1.5	0.9517	0.9637	0.9634	0.9549	0.9605
	2.0	0.9781	0.9843	0.9844	0.9801	0.9825
	2.5	0.9888	0.9927	0.9933	0.9908	0.9920
	3.0	0.9950	0.9970	0.9969	0.9957	0.9962
Makeham	1.2	0.5220	0.5842	0.6166	0.6272	0.6135
	1.5	0.6258	0.6865	0.7200	0.7266	0.7150
	2.0	0.7473	0.8075	0.8363	0.8439	0.8345
	2.5	0.8298	0.8814	0.9039	0.9081	0.9014
	3.0	0.8880	0.9286	0.9454	0.9489	0.9438
Weibull	1.2	0.5319	0.6034	0.6510	0.6814	0.6592
	1.5	0.9944	0.9989	0.9996	0.9997	0.9994
	2.0	1.0000	1.0000	1.0000	1.0000	1.0000
	2.5	1.0000	1.0000	1.0000	1.0000	1.0000
	3.0	1.0000	1.0000	1.0000	1.0000	1.0000

5.4. Comparison with Existing Nonparametric Procedures

For a comprehensive performance assessment, the proposed adaptive procedure is compared with several well-established nonparametric tests for exponentiality against IFR alternatives. These benchmark procedures were selected because they represent the principal methodological paradigms developed for this problem, including tests based on normalized spacings, U-statistics, L-statistics, moment inequalities, and entropy-based measures of departure from exponentiality. The resulting power comparisons are reported in Tables 9–11 for the three sample-size settings considered.

Klefsjö's [9] spacing-based statistic is denoted by V_1. The U-statistic of Ahmad [7] is V_2, the second-moment inequality test of Ahmad et al. [15] is V_3, the L-statistic of Mitra and Anis [11] is V_4, and the moment-inequality statistic of Ahmad [16] is V_5. The entropy test of Alqefari [13] is denoted by V_β and is evaluated at $\beta = 0.6, 0.8$, and 1.0. The restricted adaptive statistic $T_n^{\max}$ is included as the final competitor.

The proposed adaptive procedure remains highly competitive with the established benchmark tests. Although it is not uniformly the most powerful across all scenarios, its performance consistently remains close to that of the strongest competing procedure over the three representative IFR families while improving monotonically with both the sample size and the strength of departure from exponentiality. Its principal advantage lies in providing a single a priori decision rule supported by an exact finite-sample null distribution, thereby avoiding the retrospective selection of an optimal tuning parameter.

Table 9. Power comparison with existing tests at $n = 25$ and $\gamma = 0.05$.

Distribution	α	V_1	V_2	V_3	V_4	V_5	$V_{0.6}$	$V_{0.8}$	$V_{1.0}$	Adaptive
LFR	1.2	0.3200	0.1966	0.2560	0.3930	0.2690	0.3920	0.4338	0.3998	0.3859
	1.5	0.3586	0.2110	0.2944	0.4228	0.3098	0.4582	0.4846	0.4456	0.4415
	2.0	0.4168	0.2022	0.3504	0.5100	0.3356	0.5220	0.5610	0.5158	0.5179
	2.5	0.4414	0.2020	0.3758	0.5658	0.3922	0.5698	0.5992	0.5800	0.5734
	3.0	0.5052	0.1922	0.4106	0.6066	0.4292	0.6224	0.6496	0.6164	0.6161
Makeham	1.2	0.1788	0.1184	0.1508	0.2230	0.1472	0.2356	0.2478	0.2090	0.2309
	1.5	0.2158	0.1180	0.1884	0.2672	0.1796	0.2754	0.2776	0.2484	0.2691
	2.0	0.2508	0.1272	0.2230	0.3358	0.2218	0.3518	0.3594	0.2964	0.3305
	2.5	0.2974	0.1288	0.2634	0.3726	0.2494	0.4024	0.4146	0.3538	0.3827
	3.0	0.3254	0.1320	0.2842	0.4298	0.2864	0.4476	0.4638	0.3956	0.4330
Weibull	1.2	0.1664	0.2006	0.1366	0.2480	0.1434	0.2396	0.2548	0.2232	0.2372
	1.5	0.4380	0.5736	0.4172	0.7420	0.4086	0.7444	0.7544	0.6844	0.7244
	2.0	0.8168	0.9638	0.8768	0.9950	0.8952	0.9972	0.9978	0.9914	0.9939
	2.5	0.9504	0.9994	0.9958	1.0000	0.9958	1.0000	1.0000	1.0000	1.0000
	3.0	0.9874	1.0000	1.0000	1.0000	0.9998	1.0000	1.0000	1.0000	1.0000

Table 10. Power comparison with existing tests at $n = 50$ and $\gamma = 0.05$.

Distribution	α	V_1	V_2	V_3	V_4	V_5	$V_{0.6}$	$V_{0.8}$	$V_{1.0}$	Adaptive
LFR	1.2	0.5066	0.4112	0.5104	0.6520	0.5304	0.6368	0.6694	0.6664	0.6627
	1.5	0.5794	0.4444	0.5738	0.7122	0.5978	0.7138	0.7422	0.7324	0.7386
	2.0	0.6542	0.4838	0.6568	0.7994	0.6834	0.8090	0.8294	0.8096	0.8175
	2.5	0.7108	0.4912	0.7056	0.8564	0.7322	0.8576	0.8694	0.8584	0.8653
	3.0	0.7436	0.4996	0.7500	0.8838	0.7730	0.8934	0.9152	0.9018	0.9020
Makeham	1.2	0.2834	0.1924	0.2638	0.4028	0.2602	0.3732	0.3622	0.3564	0.3790
	1.5	0.3336	0.2214	0.3128	0.4770	0.3004	0.4590	0.4328	0.4568	0.4543
	2.0	0.4250	0.2562	0.3876	0.5780	0.3692	0.5648	0.5608	0.5472	0.5638
	2.5	0.4836	0.2682	0.4506	0.6752	0.4468	0.6372	0.6430	0.6334	0.6492
	3.0	0.5446	0.2746	0.5134	0.7230	0.4948	0.7034	0.7084	0.6956	0.7108
Weibull	1.2	0.2846	0.2856	0.2012	0.4224	0.2028	0.4122	0.4054	0.3656	0.4079
	1.5	0.7442	0.8348	0.6962	0.9658	0.6858	0.9686	0.9572	0.9418	0.9576
	2.0	0.9788	0.9994	0.9968	1.0000	0.9972	1.0000	1.0000	1.0000	1.0000
	2.5	0.9984	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	3.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 11. Power comparison with existing tests at $n = 100$ and $\gamma = 0.05$.

Distribution	α	V_1	V_2	V_3	V_4	V_5	$V_{0.6}$	$V_{0.8}$	$V_{1.0}$	Adaptive
LFR	1.2	0.7794	0.6592	0.8192	0.9306	0.8424	0.9202	0.9426	0.9452	0.9258
	1.5	0.8384	0.7166	0.8812	0.9568	0.8946	0.9610	0.9718	0.9694	0.9605
	2.0	0.8970	0.7766	0.9290	0.9818	0.9424	0.9830	0.9872	0.9880	0.9825
	2.5	0.9276	0.7970	0.9516	0.9916	0.9628	0.9926	0.9954	0.9938	0.9920
	3.0	0.9534	0.8020	0.9702	0.9960	0.9792	0.9952	0.9978	0.9986	0.9962
Makeham	1.2	0.4726	0.3588	0.4204	0.6308	0.4372	0.6360	0.6068	0.5822	0.6135
	1.5	0.5582	0.4132	0.5042	0.7240	0.5294	0.7356	0.7102	0.7110	0.7150
	2.0	0.6664	0.4944	0.6282	0.8360	0.6424	0.8628	0.8318	0.8216	0.8345
	2.5	0.7566	0.5226	0.7120	0.9100	0.7300	0.9126	0.9026	0.8918	0.9014
	3.0	0.8040	0.5756	0.7846	0.9440	0.7970	0.9520	0.9434	0.9304	0.9438
Weibull	1.2	0.4428	0.4804	0.3366	0.7080	0.3664	0.7062	0.6482	0.6052	0.6592
	1.5	0.9330	0.9878	0.9532	0.9998	0.9542	0.9998	0.9996	0.9994	0.9994
	2.0	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	3.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

5.5. Power Under an IFRA-but-Not-IFR Alternative

Example 2 shows that the population sign condition extends beyond strict IFR aging. To examine whether this broader theoretical domain also yields practical detection capability, we simulated the Adaptive CRE-PH Test under the lifetime of a two-component parallel system with independent exponential components having rates $\lambda_1 = 1$ and $\lambda_2 = 2$. The system lifetime is IFRA but not globally IFR: its hazard rate is non-monotone, whereas the average hazard remains non-decreasing.

The experiment follows the same calibration protocol adopted throughout Section 5. For each sample size, the adaptive critical value is obtained from 100,000 Monte Carlo draws from the exact Dirichlet null distribution, whereas power is estimated from 20,000 independently generated parallel-system lifetimes. The resulting rejection probabilities are reported together with the four exactly standardized fixed-q statistics corresponding to the prespecified adaptive grid. These results are summarized in Table 12.

Table 12. Empirical power under the IFRA-but-not-IFR parallel exponential system with $\lambda_1 = 1$ and $\lambda_2 = 2$ at $\gamma = 0.05$.

n	Adaptive Critical Value	Adaptive	$q = 0.8$	$q = 1.0$	$q = 1.2$	$q = 1.4$
25	1.6796	0.3934	0.2822	0.3257	0.3691	0.4075
50	1.7107	0.6399	0.4069	0.4990	0.5890	0.6650
100	1.7416	0.8854	0.5622	0.7061	0.8261	0.9051

The results show substantial and steadily increasing sensitivity to this IFRA-but-not-IFR alternative. Adaptive power rises from 0.3934 at $n = 25$ to 0.6399 at $n = 50$ and 0.8854 at $n = 100$. Among the fixed components, larger tuning values are more sensitive for this benchmark, with $q = 1.4$ attaining the highest componentwise power. The adaptive maximum remains close to the strongest fixed component while retaining a single prespecified rule. Thus, the broader IFRA domain is not merely formal; it is accompanied by meaningful finite-sample detection capability.

5.6. Specificity Under DFR and Bathtub-Shaped Alternatives

To assess directional specificity, the Adaptive CRE-PH Test was simulated under Weibull DFR models with shape $\alpha = 0.5, 0.7$, and 0.9, and under piecewise-exponential bathtub hazards. For the bathtub models, the hazard is λ_1 for $0 < t \leq 0.5$, λ_2 for $0.5 < t \leq 2$, and λ_3 for $t > 2$. The balanced, mild late-dominant, and strong late-dominant profiles use $(\lambda_1, \lambda_2, \lambda_3) = (1.5, 0.5, 1.2)$, $(1.2, 0.6, 1.8)$, and $(1.2, 0.25, 4.0)$, respectively. The corresponding rejection rates are reported in Table 13.

Table 13. Rejection rates of the Adaptive CRE-PH Test under off-direction alternatives.

Alternative	Parameter/Profile	$n = 25$	$n = 50$	$n = 100$
Weibull DFR	$\alpha = 0.5$	0.0000	0.0000	0.0000
	$\alpha = 0.7$	0.0006	0.0001	0.0000
	$\alpha = 0.9$	0.0161	0.0094	0.0034
Piecewise bathtub	Balanced	0.0061	0.0053	0.0034
	Mild late-dominant	0.0483	0.0943	0.1989
	Strong late-dominant	0.3773	0.7353	0.9740

The rejection rates under the DFR alternatives remain at or below the nominal significance level, providing empirical evidence of the directional specificity of the proposed procedure with respect to decreasing hazard functions. The results for the bathtub alternatives are more nuanced. Nearly symmetric bathtub profiles are rarely rejected, whereas the rejection probability increases progressively as the late-life increasing phase becomes more dominant. Consequently, rejection should be interpreted as evidence of departure from exponentiality in an IFR-sensitive direction rather than as definitive evidence that the hazard function is globally increasing.

5.7. Overall Simulation Conclusions

The simulation study leads to five principal conclusions. First, the exactly standardized fixed-q statistics exhibit increasing power as both the sample size and the strength of departure from exponentiality increase. Second, the most effective tuning value depends on the underlying IFR model but is consistently located within a moderate range of q. Third, the adaptive maximum based on the prespecified restricted grid consistently achieves power between the weakest and strongest fixed-q components while remaining close to the oracle maximum. Fourth, the parallel-system experiment demonstrates meaningful detection capability under an IFRA-but-not-IFR alternative, providing empirical support for the broader population sign condition established in Section 2. Finally, the DFR and bathtub experiments confirm that rejection should be interpreted as directional evidence of an IFR-oriented departure from exponentiality rather than as a definitive classification of the global hazard shape.

This final point is particularly important in reliability applications. The Adaptive CRE-PH Test is designed to detect departures from exponentiality in the direction of increasing failure rates; it is not intended to serve as a general hazard-shape classifier. Consequently, formal rejection should be interpreted together with complementary graphical diagnostics, such as the TTT plot, or other hazard-shape assessment tools before concluding that the underlying lifetime distribution is globally IFR.

6. Illustrative Real-Data Applications

The purpose of this section is to illustrate the practical application and interpretation of the Adaptive CRE-PH Test in reliability studies rather than to provide a formal classification of hazard-function shapes. The proposed procedure is a directional test of exponentiality against IFR alternatives. By contrast, the accompanying graphical tools—including the exponential Q–Q plot and the TTT transform—serve solely as exploratory diagnostic aids for visualizing the overall aging pattern of the data. These graphical displays complement the formal hypothesis test but do not constitute independent inferential procedures. Accordingly, the real-data analyses are intended to demonstrate the recommended inferential workflow rather than to establish formal hazard-shape classifications.

Three lifetime datasets are analyzed. For each dataset, we report the adaptive p-value together with an exponential Q–Q plot and a TTT transform. The formal null hypothesis is exponentiality, and rejection provides statistical evidence of departure in the IFR-sensitive direction. The graphical diagnostics are included only to aid interpretation of the observed aging pattern and should not be regarded as formal evidence for any particular hazard-function class. Definitive identification of IFR, DFR, bathtub, or other hazard structures requires dedicated shape-specific inferential procedures.

Under exponentiality, the TTT curve is expected to lie close to the diagonal. A predominantly concave curve is consistent with IFR-type behavior, whereas a predominantly convex curve is suggestive of DFR behavior. An S-shaped curve indicates a non-monotone hazard structure, such as a bathtub-shaped hazard. In these situations, the Adaptive CRE-PH Test should be interpreted as detecting an IFR-oriented departure from exponentiality rather than as establishing a globally increasing hazard function. As demonstrated by the off-direction simulation experiments in Section 5.6, this distinction is particularly important for bathtub distributions, where a pronounced late-life wear-out phase may generate substantial directional evidence despite the absence of globally monotone aging.

6.1. Glass Fiber Strengths

The first dataset consists of tensile strengths of glass fibers of length 1.5 cm, originally reported by Smith and Naylor [17]. These data are frequently used in lifetime and reliability modeling because they exhibit a clear departure from a simple exponential baseline and are commonly associated with Weibull-type behavior.

Table 14 provides descriptive evidence that the exponential model does not adequately represent the data. Specifically, the exponential fit yields a K–S statistic of 0.4180 with a p-value below 0.001, indicating substantial lack of fit. By comparison, the Weibull model achieves the smallest AIC and BIC values and is not rejected by the K–S test at the 5% significance level, suggesting that a non-exponential aging model provides a more appropriate description of the observed lifetimes.

Table 14. Parametric model selection criteria for the glass fiber strength data.

Distribution	Log-Likelihood	AIC	BIC	K-S	p-Value
Exponential	−88.8303	179.6606	181.8038	0.4180	0.0000
Weibull	−15.2068	34.4137	38.7000	0.1522	0.1079
Log-normal	−28.0049	60.0099	64.2961	0.2313	0.0024
Gamma	−23.9515	51.9031	56.1893	0.2164	0.0055

The exponential Q–Q plot in Figure 5 shows systematic curvature, and the TTT plot is concave. The fixed-q statistics and the adaptive statistic are all highly significant. The formal rejection is therefore compatible with, but not independently confirmed by, the IFR-like graphical pattern.

6.2. Device Failure Times

The second dataset consists of the 50 device lifetimes reported by Aarset [18], a benchmark dataset widely recognized for exhibiting a bathtub-shaped hazard function. It is included intentionally as an out-of-class diagnostic example to illustrate an important interpretational aspect of the proposed methodology. Specifically, it demonstrates that rejection of exponentiality by an IFR-oriented test should not be interpreted, in isolation, as evidence of a globally increasing hazard function.

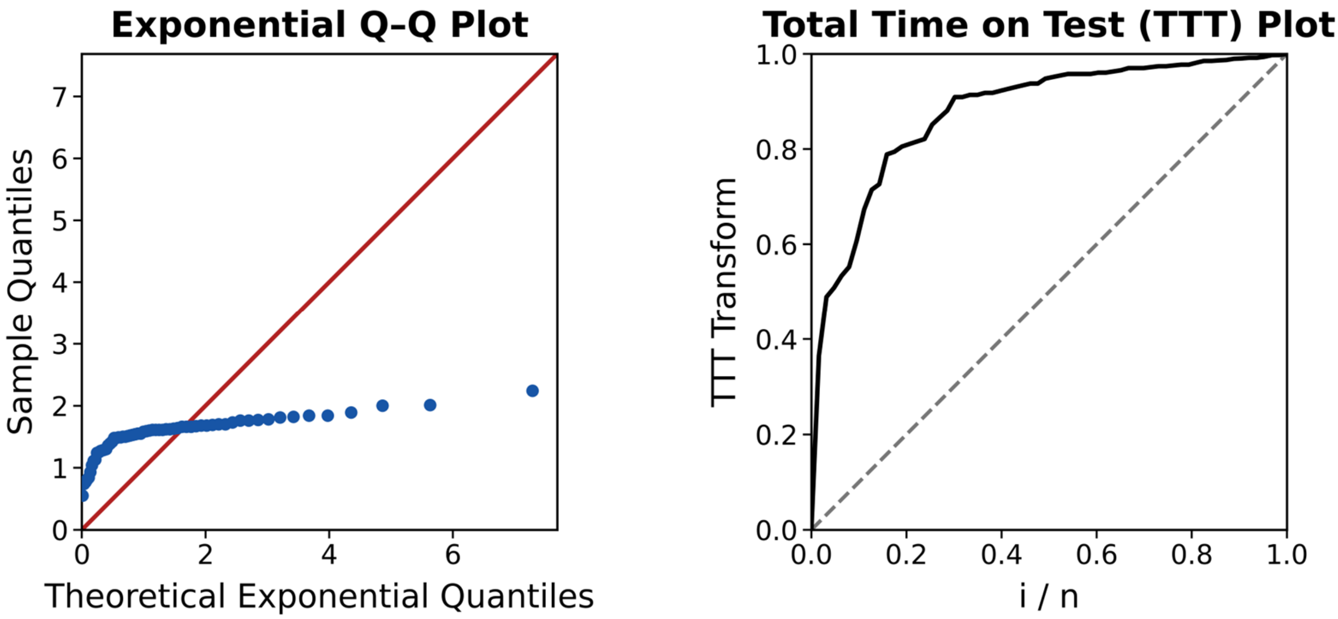

Figure 5. Exponential Q–Q plot and TTT plot for the glass fiber strength data. In the Q–Q panel, the straight reference line represents the fitted exponential model; in the TTT panel, the diagonal represents the exponential benchmark. The blue points are the observed sample quantiles and the red straight line is the fitted exponential reference; in the TTT panel, the black empirical TTT curve is compared with the gray diagonal exponential benchmark.

Table 15 indicates only borderline evidence against the exponential model according to the K–S test. For the present illustrative example, however, the S-shaped TTT curve provides descriptive evidence of a non-monotone aging pattern. Neither the parametric model comparison nor the graphical diagnostics should be interpreted as formal inferential procedures for classifying the data as IFR or for establishing a bathtub hazard. Such conclusions require dedicated shape-specific inferential methods.

Table 15. Parametric model selection criteria for the device failure-time data.

Distribution	Log-Likelihood	AIC	BIC	K-S	p-Value
Exponential	−241.0896	484.1792	486.0912	0.1911	0.0519
Weibull	−241.0018	486.0036	489.8277	0.1927	0.0487
LFR	−238.0636	480.1272	483.9513	0.1768	0.0877
Gamma	−240.1902	484.3804	488.2045	0.2022	0.0335

The K–S p-value for the exponential model ($p = 0.0519$) lies only marginally above the conventional 5% significance level and should therefore not be regarded as compelling evidence in favor of exponentiality. Moreover, although the LFR model achieves the smallest AIC among the fitted parametric models, AIC provides only a measure of relative model adequacy and does not determine the qualitative shape of the underlying hazard function. Consequently, it cannot, by itself, justify classifying these data as exhibiting IFR behavior, particularly in light of the S-shaped TTT curve, which is descriptively consistent with a non-monotone hazard pattern. The corresponding exponential Q–Q and TTT diagnostics are shown in Figure 6.

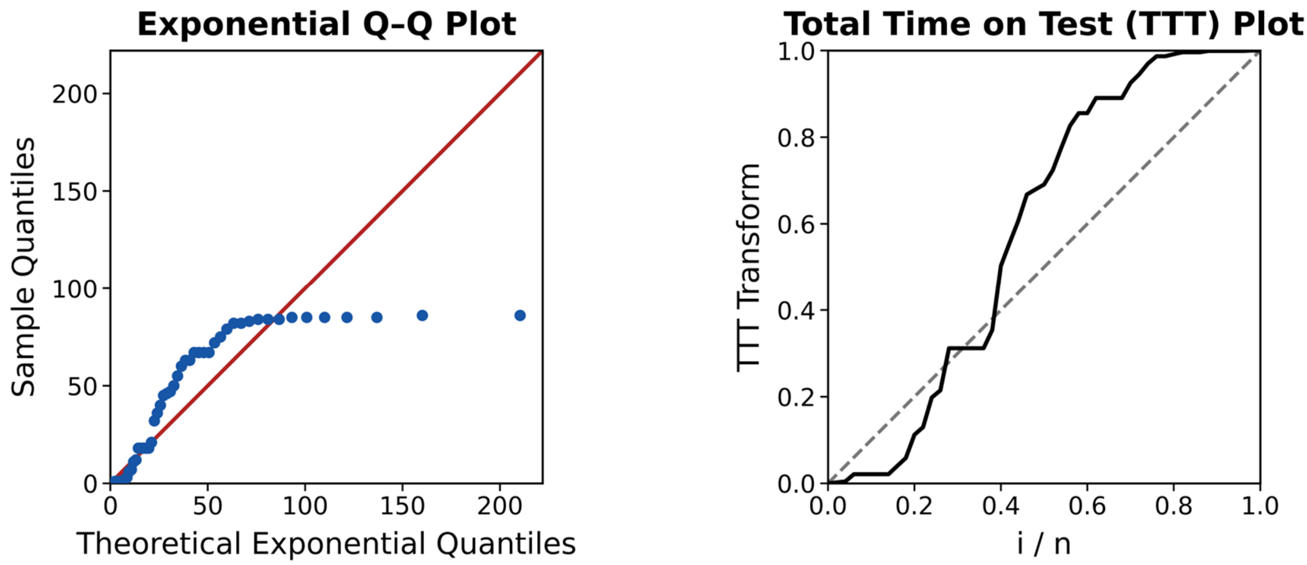

Figure 6. Exponential Q–Q plot and TTT plot for the device failure-time data. In the Q–Q panel, the straight reference line represents the fitted exponential model; in the TTT panel, the diagonal represents the exponential benchmark. The blue points are the observed sample quantiles and the red straight line is the fitted exponential reference; in the TTT panel, the black empirical TTT curve is compared with the gray diagonal exponential benchmark.

The highly significant adaptive p-value provides formal evidence of departure from exponentiality in the direction of increasing failure rates. However, this result should not be interpreted as establishing that the underlying lifetime distribution possesses a globally increasing hazard function. In the present dataset, the accompanying TTT transform exhibits an S-shaped pattern that is descriptively consistent with a bathtub-type hazard. This graphical display is included solely as an exploratory diagnostic to characterize the observed aging pattern, whereas the Adaptive CRE-PH Test provides formal directional inference against exponentiality. Consequently, the numerical test and the TTT diagnostic should be viewed as complementary rather than independent inferential tools, and their combined use does not constitute a formal statistical procedure for distinguishing IFR from bathtub hazards. Such discrimination requires dedicated hazard-shape inference specifically designed for that purpose.

The primary role of the Adaptive CRE-PH Test is to provide formal directional evidence against exponentiality in the direction of increasing failure rates rather than to classify the global shape of the hazard function. When the accompanying graphical diagnostics suggest more complex aging patterns, such as bathtub or upside-down bathtub hazards, any conclusion regarding the underlying hazard structure should be based on dedicated shape-specific inferential procedures rather than on the Adaptive CRE-PH Test alone.

6.3. Bladder Cancer Remission Times

The third dataset consists of remission times (in months) for 128 bladder cancer patients, as reported by Shanker et al. [19]. It is included as a contrasting example in which the exponential model receives no strong evidence against it from either the formal Adaptive CRE-PH Test or the accompanying graphical diagnostics.

Figure 7 exhibits only mild upper-tail deviations from the exponential Q–Q line, and the TTT curve remains close to the diagonal, both of which are descriptively consistent with exponential behavior. The Adaptive CRE-PH Test likewise produces a non-significant p-value ($p = 0.5385$), indicating that the exponential model is not rejected. Taken together, the formal test and the accompanying graphical diagnostics provide a coherent interpretation in which there is no convincing evidence against exponentiality for this dataset.

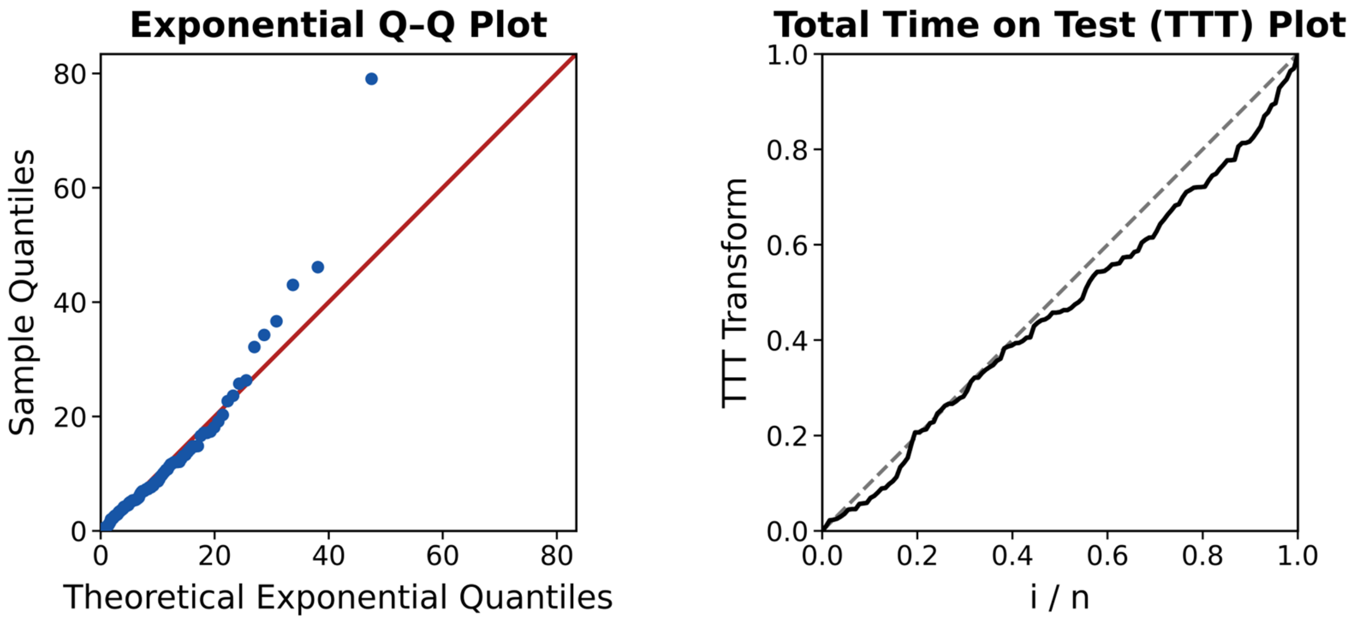

Figure 7. Exponential Q–Q plot and TTT plot for the bladder cancer remission-time data. In the Q–Q panel, the straight reference line represents the fitted exponential model; in the TTT panel, the diagonal represents the exponential benchmark. The blue points are the observed sample quantiles and the red straight line is the fitted exponential reference; in the TTT panel, the black empirical TTT curve is compared with the gray diagonal exponential benchmark.

6.4. Summary of Fixed-q and Adaptive Decisions

Table 16 summarizes the adaptive analyses for the three real-data applications. In addition to the four exactly standardized fixed-q components, it reports the adaptive statistic, the corresponding Monte Carlo p-value, and the resulting testing decision. The glass-fiber data provide overwhelming evidence against exponentiality, while the accompanying TTT curve is descriptively consistent with an IFR aging pattern. The Aarset device-lifetime data also yield a significant departure from exponentiality; however, their S-shaped TTT curve is more consistent with a bathtub-type pattern than with a globally increasing hazard. By contrast, the bladder cancer remission data show no statistically significant evidence against exponentiality, in agreement with the nearly linear exponential Q–Q plot and the TTT curve lying close to the diagonal.

Table 16. Complete summary of the adaptive analysis for the three real-data applications.

Data Set	n	T(0.8)	T(1.0)	T(1.2)	T(1.4)	Adaptive Statistic (Max)	Adaptive p-Value/Decision
Glass fiber strengths	63	6.1214	7.4628	8.7857	10.0554	10.0554	<0.0001; Reject H_0
Aarset device failures	50	3.1650	3.0602	2.7961	2.3691	3.1650	0.0006; Reject H_0
Bladder-remission times	128	−1.7881	−1.2036	−0.5285	0.1690	0.1690	0.5385; Do not reject H_0

Note. The adaptive statistic is defined as the maximum of the four exactly standardized fixed-q statistics over the prespecified grid $\mathcal{Q}_A = \{0.8, 1.0, 1.2, 1.4\}$. The adaptive p-value is computed by Monte Carlo evaluation of the exact finite-sample Dirichlet null distribution established in Theorem 2. All testing decisions are reported at the nominal 5% significance level.

Taken together, these applications clarify the inferential scope of the Adaptive CRE-PH Test. The procedure is highly sensitive to IFR-oriented departures, remains appropriately non-significant for data that are approximately exponential, and may also reject certain non-monotone hazard structures when a sufficiently pronounced late-life increasing phase dominates the directional score. Accordingly, rejection should be interpreted as evidence of departure from exponentiality in an IFR-sensitive direction rather than as confirmation that the underlying hazard function is globally increasing.

The real-data examples therefore illustrate the recommended practical workflow: the Adaptive CRE-PH Test provides formal directional inference, whereas the accompanying Q–Q and TTT plots provide complementary descriptive information about the overall aging pattern. Neither the graphical diagnostics alone nor their combined use with the adaptive test constitutes a formal hazard-shape classification procedure. Definitive discrimination among IFR, DFR, bathtub, and other complex hazard structures requires dedicated shape-specific inferential methods.

7. Conclusions and Future Research

This paper developed the Adaptive Cumulative Residual Entropy Test under Proportional Hazards (Adaptive CRE-PH Test) for testing exponentiality against IFR alternatives. The proposed procedure combines a prespecified finite stability-oriented grid with exact finite-sample centering and scaling, thereby separating structural adaptation from retrospective tuning. The tuning parameter q serves as a common transformation and Tsallis-entropy index that determines the proportional-hazards representation, the entropy weighting, and the induced score geometry, and is therefore not interpreted as an estimated regression hazard ratio.

A principal theoretical contribution is the extension of the underlying population departure condition from the IFR class to the broader IFRA class. Although the reliability interpretation of the proposed procedure remains IFR-oriented, the population functional is non-negative precisely under IFRA aging. The additional parallel-system experiment involving heterogeneous exponential components demonstrates that this broader theoretical scope is accompanied by meaningful finite-sample detection capability for an IFRA-but-not-IFR benchmark.

The adaptive statistic itself is developed within a complete theoretical framework. Under exponentiality, the vector of fixed-q statistics admits an exact joint finite-sample representation through a linear transformation of a Dirichlet normalized-spacing vector, yielding exact means, covariances, and the joint null distribution of the adaptive maximum. Joint asymptotic normality follows from multivariate L-statistic theory and the delta method, while consistency is established under fixed alternatives whenever at least one grid component has a strictly positive population departure. The present asymptotic development is intentionally confined to the joint null distribution and consistency under fixed alternatives. A rigorous local asymptotic efficiency analysis for the adaptive maximum would require a separate joint derivation under contiguous alternatives and remains an important direction for future research.

The simulation study confirms accurate finite-sample calibration of the proposed procedure and demonstrates competitive adaptive performance across representative IFR mechanisms. The restricted adaptive grid $\{0.8, 1.0, 1.2, 1.4\}$ consistently controls the nominal significance level, achieves power between the weakest and strongest fixed-q components, and remains close to the oracle fixed-q procedure across the reported settings. The additional IFRA-but-not-IFR experiment further demonstrates substantial adaptive power beyond the strict IFR class. Complementary DFR and bathtub experiments clarify the inferential scope of the procedure: rejection rates remain negligible under decreasing-failure-rate alternatives, whereas pronounced late-life wear-out phases in bathtub hazards may also produce rejection. These findings confirm that the Adaptive CRE-PH Test provides directional evidence against exponentiality rather than formal classification of the global hazard shape.

The real-data applications reinforce this interpretation. The glass-fiber data exhibit clear evidence of IFR-oriented departure from exponentiality, whereas the bladder-remission data remain consistent with the exponential model. The Aarset device-lifetime data illustrate an important limitation of directional inference: although the Adaptive CRE-PH Test rejects exponentiality, the accompanying TTT curve displays the classical bathtub pattern. Consequently, the proposed procedure should be viewed as a formal directional test against exponentiality, while graphical diagnostics such as the exponential Q–Q plot and the TTT transform provide complementary descriptive information regarding the overall aging pattern. Neither separately nor jointly do these analyses constitute formal procedures for hazard-shape classification.

Overall, the proposed framework is scale-invariant, possesses an exact joint finite-sample null characterization, admits a rigorous multivariate asymptotic theory, and avoids retrospective tuning selection. No claim of universal or power-optimal grid selection is made; instead, the adopted grid represents a principled compromise among numerical stability, representative score geometry, and multiplicity economy. Any alternative prespecified finite admissible grid can be calibrated within the same exact Dirichlet framework. From a practical perspective, the Adaptive CRE-PH Test is best regarded as a reliable screening procedure for detecting departures from exponentiality in the IFR direction while deliberately avoiding stronger hazard-shape claims than those supported by the underlying inferential theory.

Several important extensions remain for future research. These include right-censored and progressively censored data, adaptive data-driven grid construction with appropriate multiplicity adjustment, local asymptotic theory for the adaptive score vector under contiguous alternatives, and formal asymptotic relative-efficiency comparisons with existing competitors. Another important direction is the development of procedures capable of directly discriminating among IFR, DFR, bathtub, upside-down bathtub, and other complex hazard structures through dedicated shape-specific statistical inference.

Supplementary File S1 contains illustrative R scripts for a representative fixed-q simulation under a linear-failure-rate alternative together with the code used to generate the exponential Q–Q plots and TTT transforms presented in the real-data applications. These scripts are intended solely as illustrative examples rather than as a complete software implementation of the Adaptive CRE-PH Test. The fixed-q construction, exact finite-sample centering and scaling, adaptive maximum, and Dirichlet calibration are fully specified mathematically in Sections 2, 3 and 5, thereby supporting independent implementation of the proposed methodology.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/math14162902/s1, Supplementary File S1, entitled "Illustrative R Scripts," contains code for a representative fixed-q linear-failure-rate simulation and for the exponential Q–Q and total-time-on-test graphical diagnostics.

Funding: This research supported by the Ongoing Research Funding program (ORF-2026-1913), King Saud University, Riyadh, Saudi Arabia.

Data Availability Statement: The original contributions presented in this study are included in the article and Supplementary Materials. Further inquiries can be directed to the corresponding author.

Acknowledgments: The author sincerely thanks the Academic Editor and the three anonymous reviewers for their careful evaluation, constructive comments, and valuable suggestions. Their insightful feedback has substantially improved the quality, clarity, and presentation of this manuscript and is gratefully acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CRE	Cumulative residual entropy
DFR	Decreasing failure rate
IFR	Increasing failure rate
LFR	Linear failure rate
PH	Proportional hazards
Q-Q	Quantile–quantile
SE	Standard error
TTT	Total time on test

References

1. Barlow, R.E.; Proschan, F. *Statistical Theory of Reliability and Life Testing: Probability Models*; Holt, Rinehart and Winston: New York, NY, USA, 1975.
2. Proschan, F.; Pyke, R. Tests for monotone failure rate. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 21 June–18 July 1965 and 27 December 1965–7 January 1966*; University of California Press: Berkeley, CA, USA, 1967; Volume 3, pp. 293–313.
3. Barlow, R.E.; Proschan, F. A note on tests for monotone failure rate based on incomplete data. *Ann. Math. Stat.* **1969**, *40*, 595–600. [CrossRef]
4. Bickel, P.J.; Doksum, K.A. Tests for monotone failure rate based on normalized spacings. *Ann. Math. Stat.* **1969**, *40*, 1216–1235. [CrossRef]
5. Bickel, P.J. Tests for monotone failure rate II. *Ann. Math. Stat.* **1969**, *40*, 1250–1260. [CrossRef]
6. Barlow, R.E.; Doksum, K.A. Isotonic tests for convex orderings. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 9 April–12 April 1970, 16 June–21 July 1970, and 26 December 1970–3 January 1971*; University of California Press: Berkeley, CA, USA, 1972; Volume 1, pp. 293–323.
7. Ahmad, I.A. A nonparametric test for the monotonicity of a failure rate function. *Commun. Stat. Theory Methods* **1975**, *4*, 967–974. [CrossRef]
8. Deshpande, J.V.; Kochar, S.C. A test for exponentiality against IFR alternatives. *Commun. Stat. Theory Methods* **1983**, *12*, 1229–1236.
9. Klefsjö, B. Some tests against aging based on the total time on test transform. *Commun. Stat. Theory Methods* **1983**, *12*, 907–927. [CrossRef]
10. Aly, E.-E. On some tests for exponentiality against IFR alternatives. *Statistics* **1990**, *21*, 217–226. [CrossRef]
11. Mitra, M.; Anis, M.Z. An L-statistic approach to a test of exponentiality against IFR alternatives. *J. Stat. Plan. Inference* **2008**, *138*, 3144–3148. [CrossRef]
12. Anis, M.Z. A family of tests for exponentiality against IFR alternatives. *J. Stat. Plan. Inference* **2013**, *143*, 1409–1415. [CrossRef]
13. Alqefari, A.A. Exact and entropy-based nonparametric tests for exponentiality against increasing failure rate alternatives. *AIMS Math.* **2026**, *11*, 12621–12649. [CrossRef]
14. Serfling, R.J. *Approximation Theorems of Mathematical Statistics*; Wiley: New York, NY, USA, 1980; Chapter 8. [CrossRef]
15. Ahmad, I.A.; Al-Wasel, I.A.; El-Bassiouny, A.H.; Kayid, M. A new approach to moment inequalities of some aging notions with hypothesis-testing applications. *J. Appl. Probab. Stat.* **2007**, *2*, 71–88.
16. Ahmad, I.A. Moment inequalities of aging families of distributions with hypothesis-testing applications. *J. Stat. Plan. Inference* **2001**, *92*, 121–132. [CrossRef]
17. Smith, R.L.; Naylor, J.C. A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1987**, *36*, 358–369. [CrossRef]
18. Aarset, M.V. How to identify a bathtub hazard rate. *IEEE Trans. Reliab.* **1987**, *36*, 106–108. [CrossRef]
19. Shanker, R.; Hagos, F.; Sujatha, S. On modeling lifetime data using one-parameter Akash, Lindley, and exponential distributions. *Biom. Biostat. Int. J.* **2016**, *3*, 54–62. [CrossRef]