Received: 30 October 2023 | Accepted: 8 January 2024

DOI: 10.1002/asmb.2848

RESEARCH ARTICLE

WILEY

New tests for trend in time censored recurrent event data

Bo Henry Lindqvist[1] | **Jan Terje Kvaløy**[2]

[1]Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway

[2]Department of Mathematics and Physics, University of Stavanger, Stavanger, Norway

Correspondence
Bo Henry Lindqvist, Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.
Email: bo.lindqvist@ntnu.no

Abstract

We consider testing for trend in recurrent event data. More precisely, for such data we consider testing of the null hypothesis of data coming from a renewal process. The new tests are essentially obtained by considering appropriate integrated versions of classical trend tests. Moreover, adaptive versions of earlier considered tests versus non-monotonic alternatives, like bathtub trend, are suggested. A simulation study shows that the new tests have favorable properties and sometimes outperform classical tests. Examples with real data are also considered.

KEYWORDS

Brownian bridge, integrated Brownian bridge, non-monotonic trend, renewal process, time censoring, trend-renewal process, trend testing

1 | INTRODUCTION

In reliability engineering and a number of other fields, understanding the patterns and occurrences of recurrent events is important. Recurrent events refer to phenomena that happen repeatedly over time, such as equipment failures, system breakdowns, or product defects. Analyzing these events and identifying underlying trends are important for ensuring the reliability and efficiency of systems and processes. There is a vast literature on testing for trend in recurrent events data. For treatments and reviews of some of the relevant literature we refer to the books by Ascher and Feingold[1] and Cook and Lawless.[2] A nice review of the literature can also be found in Lawless, Çiğşar and Cook.[3]

A trend in recurrent event data means intuitively that the pattern of events shows some kind of systematic alterations. In reliability engineering, a system may for example show an increasing trend of failure events. Alternatively a system may show a so called bathtub behavior, where there is first a period of decreasing trend ("infant diseases"), then a period of relatively constant trend and then an increasing trend of failures ("ageing"). Trend tests seek to reveal such features.

Intuitively, the null hypothesis of a trend test should state that the process is stationary in some sense. An analytically tractable choice of null hypothesis is to consider the null hypothesis of event times forming a renewal process (RP). Such a null hypothesis was the basis of the celebrated trend test by Lewis and Robinson,[4] who modified the so called Laplace test going back to the 18th century. While the latter test is most appropriate for detecting monotonic deviations from a homogeneous Poisson process, the Lewis-Robinson test turns out to be a more robust test under the "no trend" null hypothesis.

Kvaløy and Lindqvist[5] studied trend tests for Poisson process models, where the null hypothesis of a homogeneous Poisson process was tested versus various alternatives with time-varying intensities. The above mentioned Laplace test then served as a kind of "standard" for comparison to the tests under consideration. More recently, Kvaløy and Lindqvist[6] studied trend testing with the null hypothesis of RP, considering the case of time censored recurrent events. Here, time censored means that the recurrent event process is censored after a predetermined observation period (or, more generally, that the censoring time is independent of the recurrent event process). It should be noted that much of the classical literature on trend testing in reliability is based on the, perhaps less intuitive, concept of event censoring. Here, formally, censoring is performed at the occurrence of a predetermined number of events, see for example, Kvaløy and Lindqvist[6] for a brief discussion. Some practical issues connected to event censoring were considered by Caroni.[7]

Kvaløy and Lindqvist[6] constructed a class of trend tests for the time censored case by means of adapting a functional central limit theorem for renewal processes.[8] The appropriate limits then involve Brownian bridges. In this setting, the authors showed how to derive a variety of test statistics, together with their asymptotic distributions, under the null hypothesis of RP. As special cases falling out of their approach were the above mentioned test by Lewis and Robinson, as well as tests versus both monotonic and non-monotonic trends of the classical types of Kolmogorov–Smirnov, Cramér-von Mises and Anderson-Darling. Of particular interest was, moreover, an extended Lewis-Robinson type test for alternatives of bathtub type.

It turns out that trend testing, as described above, to a certain extent shares ideas and methodology with goodness-of-fit testing. Here the null hypothesis is that a certain sample of observations comes from a given distribution or distribution class, and testing is based on the empirical distribution function. Classical tests are again the Kolmogorov–Smirnov test, the Cramér–von Mises test and the Anderson-Darling test. Here, functional central limit theorems for the empirical processes are essential tools. In this setting, Henze and Nikitin[9] considered the extension of the classical goodness-of-fit tests by basing the test statistics on the integrated empirical process, leading to integrated Brownian bridges as the key limiting processes.

In the present paper, we borrow ideas from the cited paper by Henze and Nikitin in order to derive and study new trend tests obtained by appropriate integral or optimizing operations on the basic processes considered by Kvaløy and Lindqvist.[6]

The article is organized as follows. In Section 2, we review some basic results from Kvaløy and Lindqvist,[6] including the key limit result Theorem 1. The new tests for monotonic trend based on integrated statistics are presented and studied in Section 3. Section 4 is devoted to the study of tests versus non-monotonic trend, for example bathtub trend. The basis is here a test from Kvaløy and Lindqvist[6] for the case of known time of change in direction of trend. Adaptive tests are presented for the case of unknown change point. A simulation study is presented in Section 5, while two examples with real data are presented in Section 6. Some concluding remarks are given in Section 7.

2 | NOTATION AND SOME BASIC RESULTS AND TESTS

In this paper, we consider a single recurrent event process observed from time $t = 0$. Let the event times be modeled by a counting process $\{N(t), t \geq 0\}$, where $N(t)$ is the number of events in the time interval $(0, t]$ for $t > 0$. Let $T_1, T_2, \ldots$ be the event times, while $X_1, X_2, \ldots$ are the interevent times (gap times). Thus $X_i = T_i - T_{i-1}$ for $i = 1, 2, \ldots,$ where $T_0 = 0$.

Suppose now that the above counting process is a renewal process. This means that the X_i are independent and identically distributed. Let now $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$, and assume that $\sigma^2 < \infty$. Let, furthermore, the coefficient of variation of the interevent times X_i be denoted $\gamma = \sigma/\mu$.

Now, for $t > 0$ define

$$V_{t,\gamma}(s) = \frac{1}{\gamma} \frac{N(st) - sN(t)}{\sqrt{N(t)}} \text{ for } 0 \leq s \leq 1. \tag{1}$$

By using a functional central limit theorem in Billingsley,[8] Kvaløy and Lindqvist[6] showed that

Theorem 1. *Under the above assumptions,*

$$V_{t,\gamma} \Rightarrow W^0 \text{ as } t \to \infty,$$

where W^0 is the Brownian bridge and $\Rightarrow$ means weak convergence on $[0, 1]$.

Note that the result of Theorem 1 still holds (by Slutsky's theorem) if we replace γ by a consistent estimator $\hat{\gamma}$, that is, such that $\hat{\gamma} \to_p \gamma$ as $t \to \infty$ (convergence in probability).

Kvaløy and Lindqvist[6] used the above result to suggest test statistics for testing the null hypothesis that a counting process $N(t)$, observed on a given time interval $[0, \tau]$, is a renewal process. The idea was to consider transformations of Brownian bridges which can be used as measures of deviance from a Brownian bridge, and then use Theorem 1 to derive the corresponding test statistics and their asymptotic null distributions as $\tau \to \infty$.

As should be clear from the above, the use of Theorem 1 to derive tests, involves the estimation of the coefficient of variation, $\gamma = \sigma/\mu$. Kvaløy and Lindqvist[6] discuss this estimation at some length. In the present paper, we will stick to the simple choice of estimating μ and σ by the sample estimators $\hat{\mu}$ and $\hat{\sigma}$ based on the fully observed inter-event times, $X_1, X_2, \dots, X_{N(\tau)}$, giving the consistent estimator $\hat{\gamma} = \hat{\mu}/\hat{\sigma}$ under the null hypothesis. As noted by Kvaløy and Lindqvist,[6] the disadvantage of this method of estimation is that the censored time $\tau - T_{N(\tau)}$ is ignored.

Kvaløy and Lindqvist[6] noted the usefulness of the so called trend-renewal process (TRP)[10] for modeling of alternatives to the RP. In the simulation section and for some asymptotic power calculations we will therefore use TRP processes for forming alternatives to the null hypothesis of RP. A short definition is as follows. Let $\lambda(t)$ be a non-negative function defined for $t \geq 0$ and let $\Lambda(t) = \int_0^t \lambda(u)du$. Then the process $T_1, T_2, \dots$ is a TRP with trend function $\lambda(t)$ and renewal distribution F if $\Lambda(T_1), \Lambda(T_2), \dots$ is an RP with interevent times with distribution F. It is now readily seen that the RP is a special case of a TRP when $\lambda(t)$ is constant in t. Moreover, it is seen that the TRP is a nonhomogeneous Poisson process if F is an exponential distribution.

For the new tests considered in this paper, we find it useful to start by having a closer look at two of the tests derived in Kvaløy and Lindqvist.[6]

2.1 | The Lewis-Robinson test

A simple way to detect a deviation from a Brownian bridge is to consider the signed area under the path of the process. For the process $N(t)$ observed on $(0, \tau]$, Theorem 1 thus suggests the test statistic $\int_0^1 V_{\tau,\hat{\gamma}}(s)ds$. This statistic will, under the null hypothesis, as $\tau \to \infty$, converge in distribution to $\int_0^1 W^0(s)ds$, known to be normally distributed with expectation 0 and variance 1/12.

By scaling, we obtain the test statistic

$$\mathrm{LR} = -\sqrt{12}\int_0^1 V_{\tau,\hat{\gamma}}(s)ds = \frac{1}{\hat{\gamma}} \cdot \frac{\sqrt{12}}{\sqrt{N(\tau)}}\left[\frac{1}{\tau}\sum_{i=1}^{N(\tau)} T_i - \frac{1}{2}N(\tau)\right], \tag{2}$$

which is asymptotically standard normally distributed under the null hypothesis. If the factor $1/\hat{\gamma}$ is ignored, this is the well known Laplace test statistic for the null hypothesis of homogeneous Poisson process (HPP). The division by $\hat{\gamma}$ corresponds to the correction suggested by Lewis and Robinson.[4] The resulting test is of most interest for alternatives of RP involving monotonic trends.

2.2 | The extended Lewis-Robinson test

In order to test the null hypothesis of RP versus alternatives with non-monotonic trend (e.g., bathtub trend), Kvaløy and Lindqvist [6] considered the expression

$$\int_0^a V_{\tau,\hat{\gamma}}(s)ds - \int_a^1 V_{\tau,\hat{\gamma}}(s)ds, \tag{3}$$

where $0 \leq a \leq 1$. The idea was that a test based on (3) would have the ability to detect non-monotonic trends when the trend in $[0, a\tau]$ and $[a\tau, \tau]$ are in opposite directions. Now, under the null hypothesis, (3) converges in distribution to $\int_0^a W^0(s)ds - \int_a^1 W^0(s)ds$, which is normally distributed with expectation 0 and variance $1/12 - a^2(1-a)^2$.[6] As shown by Kvaløy and Lindqvist,[6] the test statistic (3), scaled to having an asymptotically standard normal distribution under the null hypothesis, can be written

$$\mathrm{ELR}(a) = \frac{1}{\hat{\gamma}} \cdot \frac{1}{\sqrt{N(\tau)}} \cdot \frac{1}{\sqrt{(1/12) - a^2(1-a)^2}}\left\{\frac{1}{\tau}\sum_{i=1}^{N(\tau)} |T_i - a\tau| - \left(\frac{1}{2} - a(1-a)\right)N(\tau)\right\}. \tag{4}$$

It is seen that $a = 0$ in fact gives the Lewis-Robinson test statistic (2), while $a = 1$ gives its negative.

It should be noted that the statistic ELR(a) tends to be positive for data corresponding to a bathtub type trend, where the trend is first decreasing and then increasing. A test based on ELR(a) versus bathtub trend should therefore be considered as one-sided, with rejection only for large values of the statistic. Conversely, for alternatives of first increasing and then decreasing trends, the corresponding test should reject the null hypothesis for low values (negative) of ELR(a).

As discussed by Kvaløy and Lindqvist,[6] an obvious disadvantage of the above test is that the value of a has to be decided prior to looking at the data. This apparent drawback has motivated the approach of Section 4 of the present paper, considering certain adaptive tests which avoid the specification of a.

3 NEW TESTS FOR MONOTONIC TREND

3.1 Integrated Lewis-Robinson type statistics

Henze and Nikitin[9] considered the integrated Brownian bridge defined by

$$\overline{W}(s) = \int_0^s W^0(u)\, du, \;\; 0 \le s \le 1. \tag{5}$$

This is clearly a Gaussian process with mean 0, and for which the covariance function is,[9]

$$K(s,t) = \mathrm{Cov}(\overline{W}(s), \overline{W}(t)) = \frac{st(s \wedge t)}{2} - \frac{(s \wedge t)^3}{6} - \frac{s^2t^2}{4}, \;\; 0 \le s,t \le 1. \tag{6}$$

This implies in particular that

$$\int_0^1 \overline{W}(s) ds \sim N(0, 1/45)$$

where we use that

$$\int_0^1 \int_0^1 K(s,t)\, ds\, dt = 1/45.$$

A statistic with this limiting distribution would hence be given by

$$\int_0^1 \int_0^a V_{\tau,\hat{\gamma}}(s)\, ds\, da \tag{7}$$

which using eq. (B.1) in app. B of Kvaløy and Lindqvist,[6] can be shown to equal

$$\frac{1}{\hat{\gamma}} \cdot \frac{1}{\sqrt{N(\tau)}} \left\{ \frac{1}{3} N(\tau) - \frac{1}{\tau} \sum_{i=1}^{N(\tau)} T_i + \frac{1}{2\tau^2} \sum_{i=1}^{N(\tau)} T_i^2 \right\}.$$

A test statistic with an asymptotic standard normal distribution under the null hypothesis may then be given by

$$\mathrm{ILR1} = \frac{1}{\hat{\gamma}} \cdot \frac{\sqrt{45}}{\sqrt{N(\tau)}} \left\{ \frac{1}{\tau} \sum_{i=1}^{N(\tau)} T_i - \frac{1}{2\tau^2} \sum_{i=1}^{N(\tau)} T_i^2 - \frac{1}{3} N(\tau) \right\}. \tag{8}$$

The corresponding test would assumingly be appropriate for alternatives with monotonic trend.

Similarly, one might consider the statistic

$$\int_0^1 \int_a^1 V_{\tau,\hat{\gamma}}(s)\, ds\, da \tag{9}$$

which can be calculated using eq. (B.2) in app. B of Kvaløy and Lindqvist[6] and which would again have the limiting distribution $N(0, 1/45)$. By appropriate scaling, we arrive at the following test statistic with an asymptotic standard normal distribution under the null hypothesis,

$$\mathrm{ILR2} = \frac{1}{\hat{\gamma}} \cdot \frac{\sqrt{45}}{\sqrt{N(\tau)}} \left\{ \frac{1}{2\tau^2} \sum_{i=1}^{N(\tau)} T_i^2 - \frac{1}{6} N(\tau) \right\}. \tag{10}$$

Because of the integration from a to 1 in (9), an apparent difference between ILR1 and ILR2 is that the latter puts more weight to events at later times. An indication of this will be noted in the example of Section 6.1.

It is seen that by summing the two statistics (8) and (10), we obtain a statistic proportional to the LR test statistic (2). In order to compare the two new tests, and compare them to the ordinary LR-test, we will in the next subsection consider their asymptotic powers when testing RP versus certain power law alternatives, following the asymptotic study of Kvaløy and Lindqvist.[6] The tests will also be compared in a simulation study in Section 5.

3.1.1 | Asymptotic power of the integrated LR-tests

Suppose now that the alternative to RP is a TRP with power law trend,

$$\lambda(t) = bt^{b-1}; \quad b > 0.$$

It was shown by Kvaløy and Lindqvist[6] that if $b \to 1$ as $\tau \to \infty$ in the way

$$b = 1 + \frac{\theta}{\sqrt{\tau}},$$

then

$$V_{\tau,\hat{\gamma}}(s) \Rightarrow W^0(s) + \rho\, s \ln s, \tag{11}$$

where $\rho = \frac{\theta}{\gamma}\sqrt{1/\mu}$. Using the above fact, and noting that $\int_0^1 s \ln s\, ds = -1/4$, it follows that[6] LR in (2) converges in distribution to

$$-\sqrt{12}\int_0^1 W^0(s)ds + \frac{\sqrt{3}}{2}\rho \sim N\left(\frac{\sqrt{3}}{2}\rho\,,1\right).$$

Using (11) and noting that

$$\int_0^1\int_0^a s \ln s\, ds\, da = -\frac{5}{36}$$
$$\int_0^1\int_a^1 s \ln s\, ds\, da = -\frac{1}{9}$$

it follows that, under the alternative of power law TRP as considered above,

$$\text{ILR1} \ \to_d \ N\left(\frac{5\sqrt{5}}{12}\rho\,,1\right)$$
$$\text{ILR2} \ \to_d \ N\left(\frac{\sqrt{5}}{3}\rho\,,1\right),$$

where $\to_d$ means convergence in distribution.

Now, since

$$\frac{5\sqrt{5}}{12} > \frac{\sqrt{3}}{2} > \frac{\sqrt{5}}{3}$$

it is seen that ILR1 has a higher local power than LR, which on the other hand has a higher local power than ILR2.

3.2 | Integrated Cramér-von Mises type statistics

Recall from Kvaløy and Lindqvist [6] that a Cramér-von Mises type test has the test statistic

$$\int_0^1 (V_{\tau,\hat{\gamma}}(s))^2\, ds$$

which under the null hypothesis converges to $\int_0^1 (W^0(s))^2 ds$, having a known distribution.

Consider now instead the following statistic,

$$\mathrm{ICvM} = \int_0^1 \left(\int_0^a V_{\tau,\hat{\gamma}}(s)\, ds \right)^2 da,$$

converging under the null hypothesis to

$$\int_0^1 \left(\int_0^a W^0(s)\, ds \right)^2 da = \int_0^1 (\overline{W}(a))^2\, da.$$

The distribution of this limit has been considered by Henze and Nikitin.[9] In particular they present a table of values from its cumulative distribution, as well as an asymptotic expansion in terms of i.i.d. standard normal random variables.

The calculation of the statistic ICvM can be done using eq. (B.1) in app. B of Kvaløy and Lindqvist.[6]

Similar to the ILR2-statistic of Section 3.1, we may of course also consider the statistic

$$\int_0^1 \left(\int_a^1 V_{\tau,\hat{\gamma}}(s)\, ds \right)^2 da,$$

having the same limiting distribution as the previous one.

3.2.1 | Asymptotic power of the integrated CvM-test

By (11) we should consider

$$\begin{aligned}
&\int_0^1 \left[\int_0^a (W^0(s) + \rho\, s \ln s) ds \right]^2 da = \int_0^1 \left[\overline{W}(a) + (\rho/4)a^2(2\ln a - 1) \right]^2 da \\
&= \int_0^1 (\overline{W}(a))^2 da + (\rho/2) \int_0^1 \overline{W}(a) a^2 (2\ln a - 1)\, da + (\rho^2/16) \int_0^1 a^4 (2\ln a - 1)^2\, da \\
&= \int_0^1 (\overline{W}(a))^2 da + (\rho/2) \int_0^1 \overline{W}(a) a^2 (2\ln a - 1)\, da + (\rho^2/16) \frac{53}{125}
\end{aligned}$$

which is now the limiting distribution of the test statistic. This should be compared to the corresponding expression for the ordinary Cramér-von Mises statistic obtained in sect. 7.3 of Kvaløy and Lindqvist.[6]

3.3 | Integrated Kolmogorov-Smirnov type statistics

Recall now from Kvaløy and Lindqvist [6] that the Kolmogorov-Smirnov type has the test statistic

$$\mathrm{KS} = \sup_{s \in [0,1]} |V_{\tau,\hat{\gamma}}(s)|$$

which in the limit as $\tau \to \infty$ has the Kolmogorov distribution.[11]

Consider now instead the following statistic,

$$\mathrm{IKS} = \sup_{a \in [0,1]} \left| \int_0^a V_{\tau,\hat{\gamma}}(s)\, ds \right|$$

converging under the null hypothesis to

$$\sup_{a \in [0,1]} \left| \int_0^a W^0(s)\, ds \right| = \sup_{a \in [0,1]} |\overline{W}(a)|.$$

The distribution of this limit is difficult to obtain. The problem is discussed, for example, by Henze and Nikitin,[9] who refer to the papers Lachal [12] and Lachal.[13] In the simulation study in Section 5 we use instead numerical calculations and simulations in order to derive the value of test statistics and critical values of the tests. The calculation of the statistic IKS can be done using eq. (B.1) in app. B of Kvaløy and Lindqvist.[6]

Alternatively we may consider the statistic

$$\sup_{a\in[0,1]} \left| \int_a^1 V_{\tau,\hat{\gamma}}(s)\, ds \right|$$

having the same limiting distribution as the previous one.

3.3.1 | Asymptotic power of the integrated KS-statistic

By (11), putting $\rho = \frac{\theta}{\gamma}\sqrt{1/\mu}$, we should consider

$$\sup_{a\in[0,1]} \left| \int_0^a (W^0(s) + \rho s \ln s)\, ds \right| = \sup_{a\in[0,1]} \left| \overline{W}(a) + \rho a^2 (2\ln a - 1) \right|$$

which is now the limiting distribution of the test statistic. This should be compared to the expression for the KS statistic obtained in sect. 7.4 of Kvaløy and Lindqvist.[6]

4 | NEW TESTS FOR NON-MONOTONIC TREND

Theorem 1 implies that (3) converges as $\tau \to \infty$ to

$$\begin{aligned} X(a) &= \int_0^a W^0(s)ds - \int_a^1 W^0(s)ds \\ &= 2\int_0^a W^0(s)ds - \int_0^1 W^0(s)ds \\ &= 2\,\overline{W}(a) - \overline{W}(1), \end{aligned} \tag{12}$$

where $\overline{W}(s)$ is integrated Brownian motion as defined in (5). Now $X(a)$ is a Gaussian process with mean 0 and covariance function which can be obtained from the covariance function of $\overline{W}$ given in (6). It hence follows from (12) that

$$\begin{aligned} K_X(s,t) &= \mathrm{Cov}(X(s), X(t)) = 4K(s,t) - 2K(s,1) - 2K(1,t) + K(1,1) \\ &= 2\,st(s\wedge t) - \frac{2}{3}(s\wedge t)^3 - s^2t^2 - \frac{1}{2}(s^2+t^2) + \frac{1}{3}(s^3+t^3) + \frac{1}{12} \end{aligned}$$

for $0 \le s, t \le 1$.

4.1 | Integrated extended LR tests

Consider now $\int_0^1 X(a)da$. This is a normally distributed random variable with expectation 0 and variance

$$\int_0^1 \int_0^1 K_X(s,t)\, ds\, dt = 1/180.$$

By integrating (3) with respect to a from 0 to 1, one hence gets a statistic converging in distribution to $N(0, 1/180)$. Now the integral of (3) equals

$$\frac{1}{\hat{\gamma}} \cdot \frac{1}{\tau\sqrt{N(\tau)}} \int_0^1 \left\{ \sum_{i=1}^{N(\tau)} |T_i - a\tau| - \left(\frac{1}{2} - a(1-a)\right)\tau N(\tau) \right\} da$$

which by noting that $\int_0^1 |T - a\tau|\, da = \frac{\tau}{2} - \frac{T(\tau-T)}{\tau}$ gives the statistic

$$\text{IELR0} = \frac{1}{\hat{\gamma}} \cdot \frac{1}{\sqrt{N(\tau)}} \left\{ \frac{1}{6} N(\tau) - \frac{1}{\tau^2} \sum_{i=1}^{N(\tau)} T_i(\tau - T_i) \right\}.$$

Alternatively one may instead integrate the ELR(a) in (4), that is, consider

$$\text{IELR1} = \int_0^1 \text{ELR}(a)\, da,$$

This statistic takes into account the normalizing factor

$$\phi(a) \equiv 1/\sqrt{1/12 - a^2(1-a)^2}$$

from the definition of ELR(a). Recall that this factor makes ELR(a) asymptotically standard normal under the null hypothesis of RP. A calculation shows that ϕ has a maximum value of 6.928 at $a = 0.5$ and minimum value 3.464 at $a = 0$ or 1.

It follows from the above that IELR1 converges in distribution to

$$\int_0^1 \phi(a) X(a) da$$

which is normally distributed with expected value 0 and variance

$$\int_0^1 \int_0^1 \phi(s)\phi(t) K_X(s,t)\, ds\, dt = 0.174943,$$

found by numerical integration. The estimator IELR1 can as well be calculated by numerical integration. Simulations have shown that the properties of IELR0 and IELR1 are rather similar.

Similarly as discussed for the ELR(a) test in Section 2, when used as a test versus bathtub trend, or other V-shaped trends, this test should be used as a one-sided test with rejection only for large values of the statistic.

4.2 | Adaptive extended LR tests

As indicated in Section 2.2, when the change point a of the trend is unknown, an adaptive choice of a in (4) should be tried. Thus consider the test statistic

$$\text{SELR1} = \sup_a \text{ELR}(a),$$

where the maximizing a can be considered as the estimated trend change point a.

We note that SELR1 is a positive random variable. To see this, suppose that ELR(a) ≤ 0 for all a. Then by (3),

$$\int_0^a V_{\tau,\hat{\gamma}}(s) ds \leq \int_a^1 V_{\tau,\hat{\gamma}}(s) ds$$

and adding $\int_0^a V_{\tau,\hat{\gamma}}(s) ds$ to both sides of the inequality, we get

$$2\int_0^a V_{\tau,\hat{\gamma}}(s) ds \leq \int_0^1 V_{\tau,\hat{\gamma}}(s) ds$$

From this we conclude, by letting first $a = 0$ and then $a = 1$, that

$$\int_0^1 V_{\tau,\hat{\gamma}}(s) ds = 0$$

with probability 1, which is impossible since the left hand side above converges to $\int_0^1 W^0(s)ds$ which is $N(0, 1/12)$. It hence follows that we must have SELR1 positive.

Similar to the consideration of the statistics IELR0 and IELR1 in the previous subsection, one might consider a test statistic SELR0 by excluding the factor $\phi(a)$ in the expression for ELR(a).

While we were able to find the limiting distributions under the null hypothesis of the test statistics IELR0 and IELR1, this is not straightforward for the SELR0 and SELR1 statistics which, in the same manner as the IKS statistic of Section 3.3, are based on suprema of Gaussian processes. Thus again we need to use numerical calculations and simulations in order to derive the values of test statistics and critical values of the tests.

The way we have constructed the statistics SELR0 and SELR1 defined above we expect tests based on these to have power against bathtub and other V-shaped trends as well as monotonic trends. The latter because the ELR(a) statistic equals the LR statistic when $a = 0$ and minus the LR statistic when $a = 1$ as discussed in Section 2. If we want to have a test with power for alternatives of first increasing and then decreasing trends, the test statistics to be used should rather be $\inf_a$ ELR(a) (and its counterpart without the $\phi(a)$ factor).

5 | SIMULATIONS

In order to study the properties of the new tests, as well as for comparison with previous tests, we have done a simulation study much in the same manner as in Kvaløy and Lindqvist.[6] In the study, we have estimated rejection probabilities by simulating 100,000 data sets for each choice of model and parameter values, recording the relative number of rejections for each test. The standard errors of the simulated rejection probabilities are then ≤ 0.0016. The nominal significance level was set to $\alpha = 0.05$. All simulations are done in R.

In addition to the tests already discussed in the present paper, we have included the Anderson–Darling test[6,14] in the study. This is because this test is considered to be a test with very good overall properties, and is hence of interest for comparisons.

The following abbreviations are used in the plots reporting the simulation results: ILR1, integrated Lewis-Robions version 1; ICvM, integrated Cramér-von Mises; IKS, integrated Kolmogorov-Smirnov; LR, Lewis-Robinson; CvM, Cramér-von Mises; KS, Kolmogorov-Smirnov; IELR1, integrated extended Lewis-Robinson version 1; SELR1, supremum of extended Lewis-Robinson version 1; ELR($a = 0.5$), extended Lewis-Robinson with $a = 0.5$; AD, Anderson-Darling.

When available, we have used the analytically derived asymptotic distributions for finding critical values for the test statistics. For the ICvM we used the critical values reported in Henze and Nikitin.[9] For the IKS and SELR1 tests we approximated the critical values by simulating the asymptotic distribution.

5.1 | Level properties

The level properties of the tests were studied by generating datasets from Weibull RPs with shape parameters 0.75 and 1.5, corresponding, respectively, to a process which is overdispersed and a process which is underdispersed relative to a homogeneous Poisson process. In Figure 1, the simulated level of the tests for systems with the expected number of events ranging from 10 to 60 is reported. The tests mostly have adequate level properties, but being based on asymptotic distributions the achieved levels tend to deviate a bit from the nominal level for small sample sizes. In these cases most of the tests are a bit non-conservative, except the KS test which is too conservative in the overdispersed case, and to some degree the ELR-test and its extensions, the IELR1 and SELR1 tests.

5.2 | Power properties: Monotonic alternatives

Datasets with a monotonic trend were generated by simulating data from TRPs with the renewal distribution being Weibull and the trend function $\lambda(t)$ being of the power law form $\lambda(t) = bt^{b-1}$ for $b > 0$. The rejection probability as a function of b was simulated, where $b < 1$ corresponds to a decreasing trend, $b = 1$ corresponds to no trend, and $b > 1$ corresponds to an increasing trend. Two different values of the shape parameter β of the Weibull renewal distribution were considered, $\beta = 0.75$ and $\beta = 1.5$, corresponding, respectively, to a process which is overdispersed and a process which is underdispersed compared to a nonhomogeneous Poisson process. The censoring times were adjusted such that the

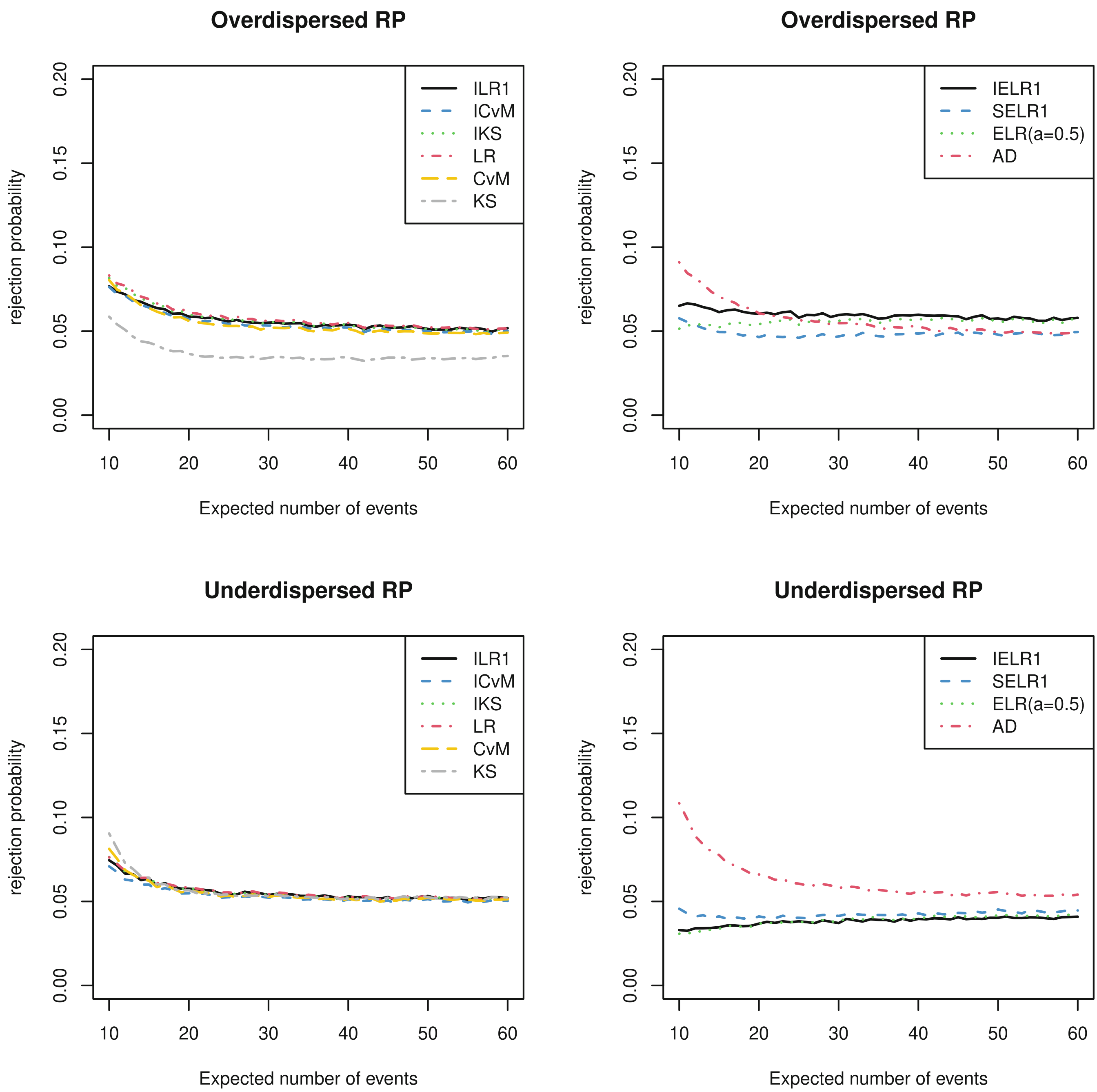

FIGURE 1 Level of the tests simulated as a function of the expected number of events. Data simulated from Weibull RPs with shape parameter 0.75 (overdispersed RP, upper plots) and 1.5 (underdispersed RP, lower plots), respectively. A nominal level of 0.05 was used.

expected number of failures was 30 in all simulations. The results are displayed in Figure 2. The two left panels show the results for the tests appropriate for monotonic trends. It might then be of interest to compare each of LR, CvM and KS with their integrated counterparts (with an 'I' in front of the notation). For the overdispersed case, it is noted that ILR1 is slightly better than LR for $b < 1$, but that LR is better for $b > 1$ (with some strange behavior of ILR1 for $b > 2.5$, say). The most surprising result is, however, that the IKS improves the KS significantly for all b. ICvM and CvM look rather close, although the integrated version is slightly better for $b < 2.5$, say. Much of the same is the case for the underdispersed case, but here ILR1 is seemingly uniformly better than LR, although they are rather close. We have not included ILR2 in the plots, but it appears in most cases to have a lower power than ILR1 (see, however, comments in Sections 3.1 and 6.1). The two right hand plots of Figure 2 are less interesting, since the tests are here tailored for non-monotonic alternatives. However, the Anderson-Darling and the SELR1 are expected to also have some power against monotonic alternatives,

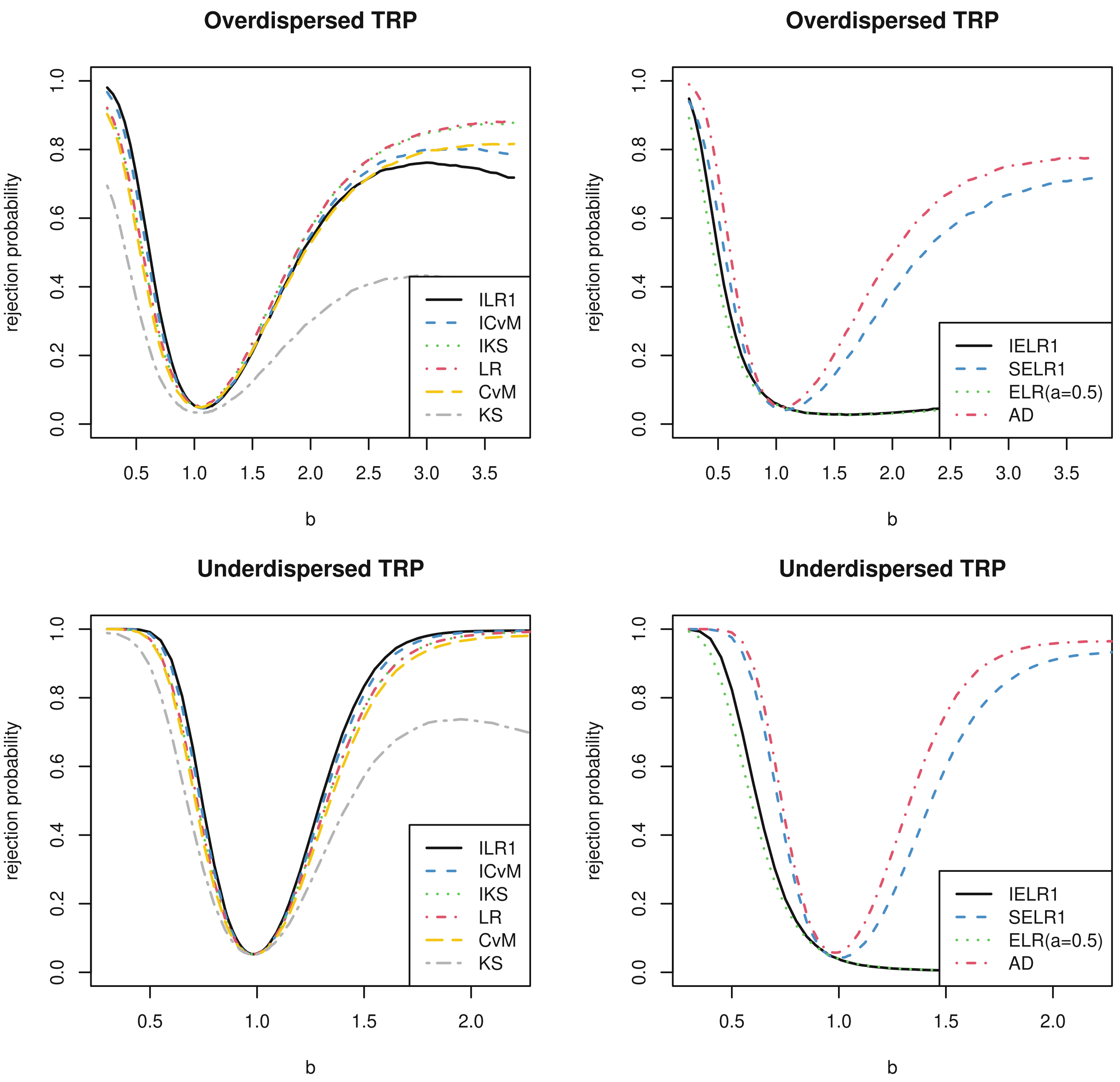

FIGURE 2 Power of the tests simulated as a function of the trend parameter b in Weibull TRPs with trend function bt^{b-1} and shape parameter $\beta = 0.75$ (overdispersed TRP, upper plots) and $\beta = 1.5$ (underdispersed TRP, lower plots), respectively. The expected number of failures in all simulations where 30.

and we see that they are having reasonable power against these monotonic alternatives. The more purely tailored ELR and IELR1 tests are not good at picking up monontonic trends, in particular not increasing trends.

5.3 | Power properties: Non-monotonic alternatives

Data sets with non-monotonic trends were generated by simulating data from TRPs with trend functions $\lambda(t)$ on the forms displayed in Figure 3.

We first consider the bathtub shaped trend generated with the trend function in the left plot in Figure 3. Here d is the average of $\lambda(t)$ over $[0, \tau]$. The degree of bathtub shape can be expressed by the parameter c, with $c = 0$ corresponding to a

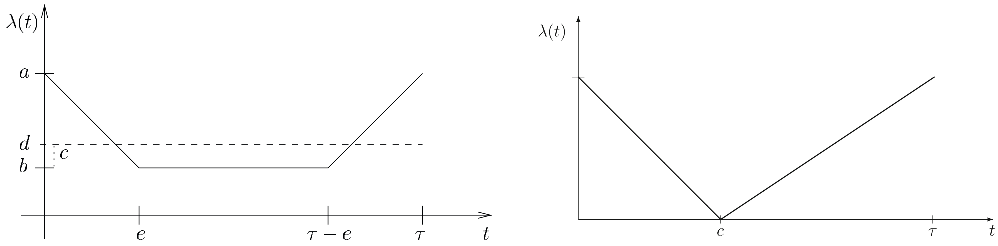

FIGURE 3 Shape of the trend functions $\lambda(t)$ used for simulating non-monotonic trends.

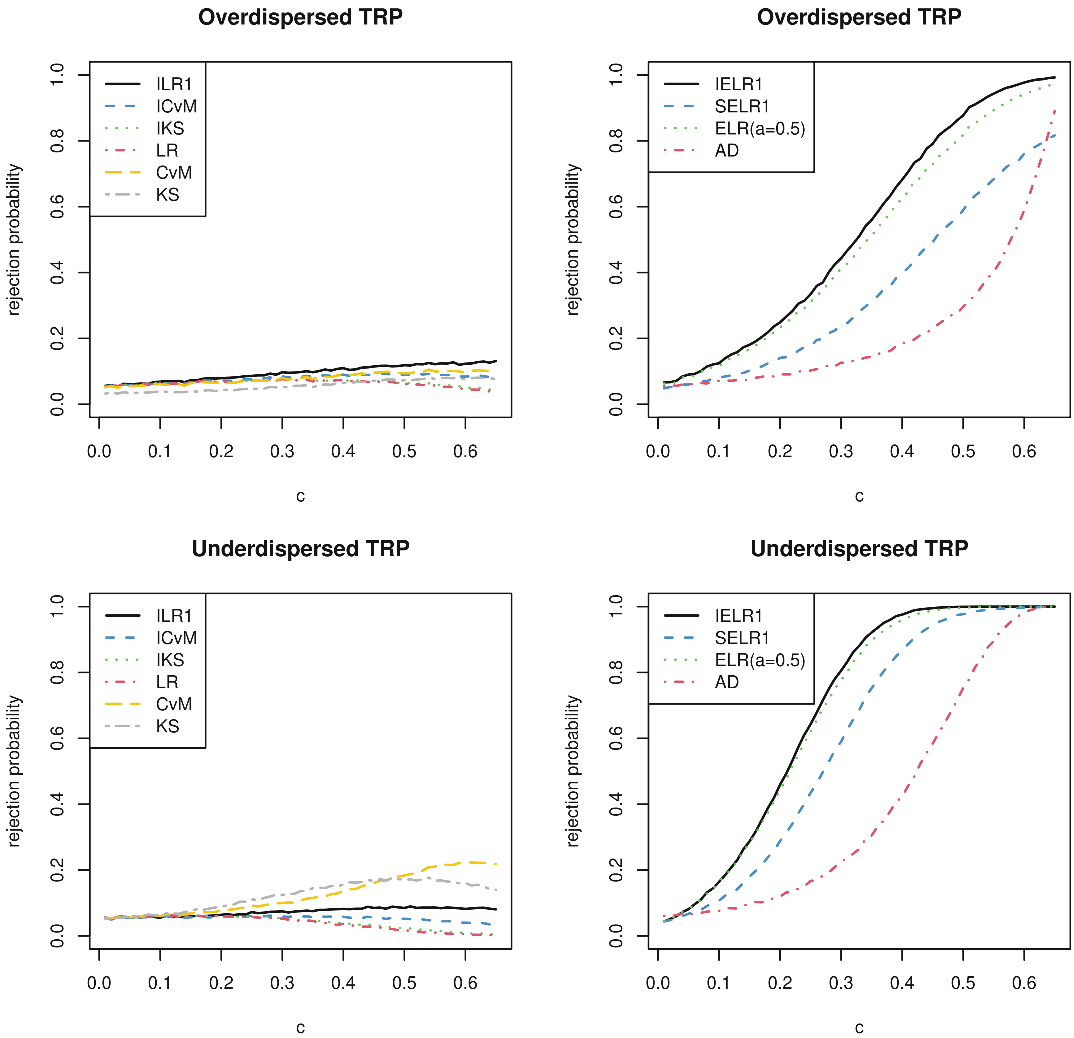

FIGURE 4 Power of the tests simulated as a function of the trend parameter c in Weibull TRPs with bathtub trend function (see left plot in Figure 3) and shape parameter $\beta = 0.75$ (overdispersed TRP, upper plots) and $\beta = 1.5$ (underdispersed TRP, lower plots), respectively. The expected number of failures in each phase where 20.

horizontal line (no trend). The rejection probability as a function of c was simulated with e and τ in each case set to values such that the expected number of failures in each phase (decreasing, no, increasing trend) were equal to 20. The shape parameter of the Weibull renewal distribution was set to, respectively, $\beta = 0.75$ and $\beta = 1.5$. The results are displayed in Figure 4. It is clear from the left hand curves that the tests versus monotonic trend have a very low power against the bathtub alternatives. Thus we concentrate on the right hand plots. Note first that the ELR-test here has a predefined $a = 0.5$, which means that it is centered at the very center of the bathtub curves under consideration (see Figure 3). It is therefore interesting to see that the test based on IELR1, which essentially integrates over all values of a, has a slightly higher power for all considered c.

The Anderson–Darling test (AD) is commonly recommended for trend testing with non-monotonic trend,[6] but is in the simulation seemingly outperformed by the more specialized tests. It can, on the other hand, be shown that the performance of the ELR-test drops when a less optimal choice of a is made. Simulations reported in Kvaløy and Lindqvist[6] have shown that the AD-test in many cases is better than the ELR-test if a somewhat "wrong" value of a is chosen in the ELR test. Consider finally the SELR1 test, which at least for $c < 0.6$ shows a, often substantially, higher power than the AD-test. Both the SELR1 and the AD tests seem to be tests with good overall performance, with the AD being best of the two for monotonic trends and SELR1 best for bathtub trend.

Finally we consider data generated with the V-shaped trend functions on the form given in the right plot in Figure 3. Here we consider power of the tests as a function of $a = c/\tau \in [0, 1]$. With $a = 0$ and $a = 1$ this corresponds to, respectively, monotonic decreasing and monotonic increasing trend. For other values of a we get a non-monotonic trend, which presumably is most pronounced for values of a close to 0.5 (c close to $\tau/2$).

The results are given in Figure 5. For a close to 0 and 1 the test for monotonic trend in the left plots in Figure 5 have high power, but we also notice that the SELR1 and AD tests have similar power in these cases. For non-monotonic trends

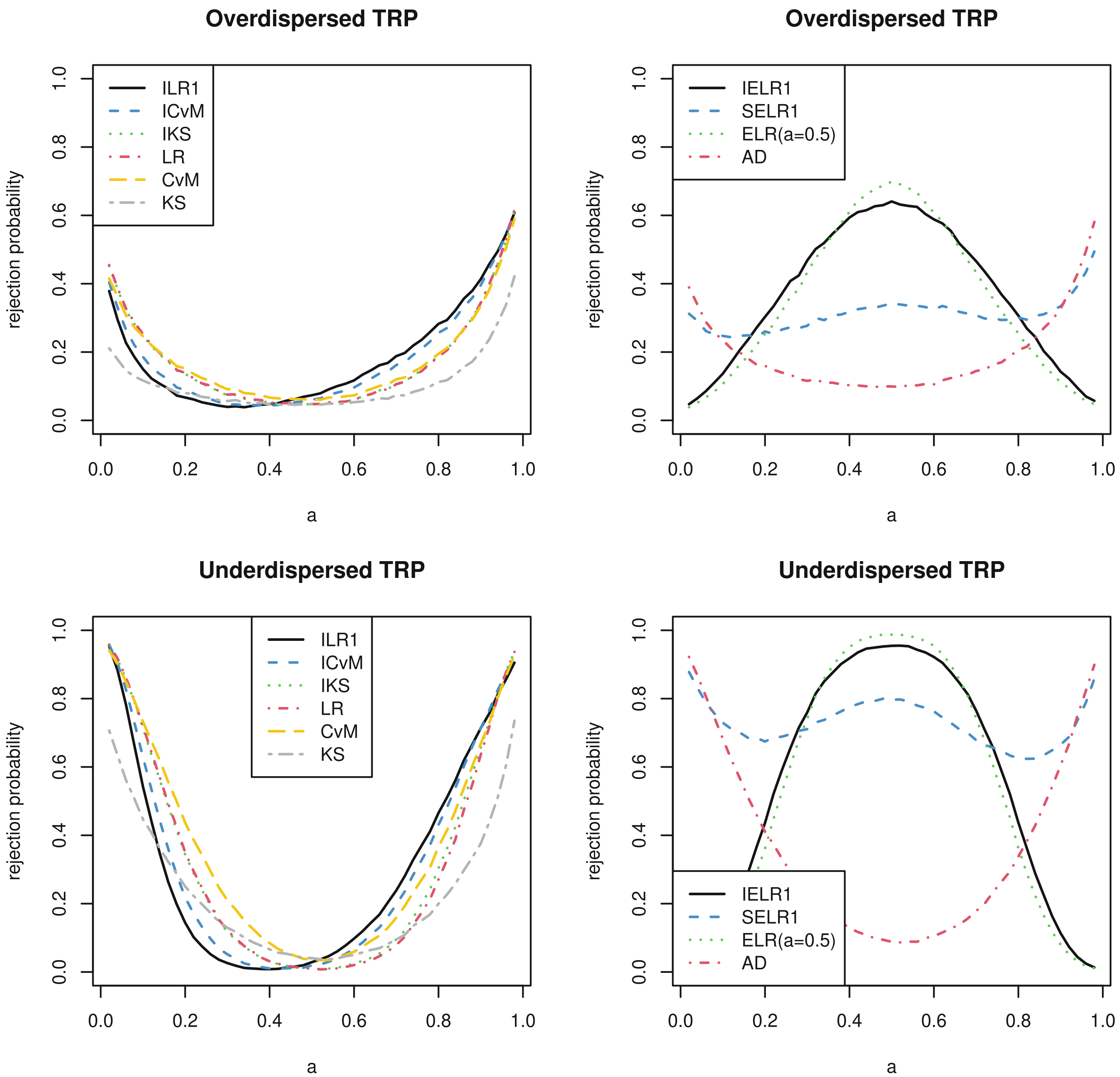

FIGURE 5 Power of the tests simulated as a function of the turning point $a = c/\tau$ in Weibull TRPs with V-shaped trend function (see right plot in Figure 3) and shape parameter $\beta = 0.75$ (overdispersed TRP, upper plots) and $\beta = 1.5$ (underdispersed TRP, lower plots), respectively. The expected number of failures where 30.

with a around 0.5 the IELR1 and ELR($a = 0.5$), which are particularly constructed for such trends, are as expected the best tests. However, these test are unable to detect the monotonic trends. We also notice that the AD test is not so good at detecting the type quickly turning non-monotonic trend studied here. The SELR1 test, however, shows good overall properties also here, in particular in the underdispersed case.

6 | DATA EXAMPLES

6.1 | U.S.S. Halfbeak data

Meeker and Escobar[15] [tab. 16.4] display 71 times of unscheduled maintenance actions for the U.S.S. Halfbeak number 4 main propulsion diesel engine. As the intensity of events seemingly makes an abrupt change around time 20,000 h, we have chosen to time censor the data at 20,000 h of operation, thereby having $n = 24$ failure times in our data set. The left plot in Figure 6 shows the plot of cumulative failure number against time. The plot indicates an increasing trend of events. We have calculated p-values for most of the tests considered in this paper, see the middle row of Table 1. The p-values are based on the asymptotic distributions of the test statistics. For the ICvM, IKS and SELR1 tests the asymptotic distribution was approximated by simulations. The general impression is that most tests manage to pick up this trend. For, say, a significance level of 5%, only the ELR($a = 0.5$) would not reject the null hypothesis. It should be noted that in the table, the tests LR, ILR1, ILR2 are treated as two-sided tests. As a final note, the low p-value for ILR2 is interesting in view of the comment in Section 3.1 that ILR2 puts more weight on the late event times. In the present data, the intensity of events seems to be highest close to the end of observation at time 20,000 h.

6.2 | Load-haul-dump machine data

Kumar, Klefsjö and Granholm[16] reported failure data for a load-haul-dump machine operating in a Swedish mine. The example was also used by Kvaløy and Lindqvist,[6] who considered the data to be time censored at 2000 h, as will also be

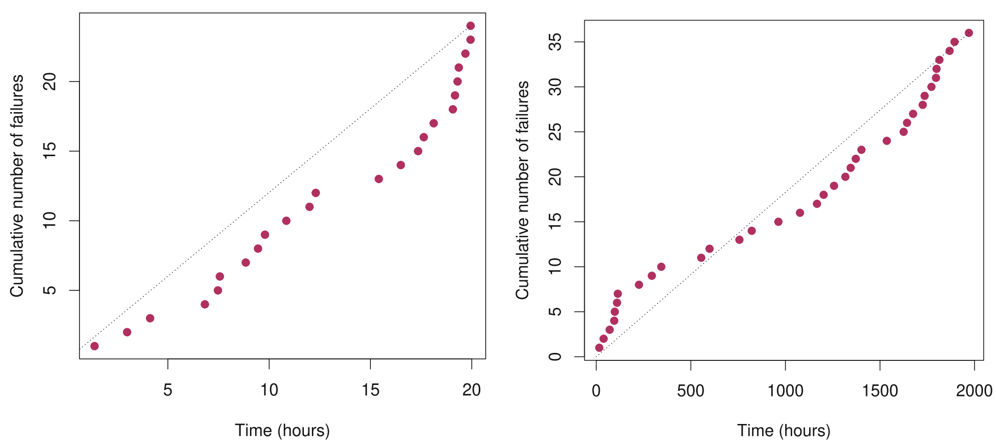

FIGURE 6 Plot of cumulative number of failures for USS Halfbeak (left) and load-haul-dump machine (right).

TABLE 1 The table reports p-values for a selection of trend tests applied to the USS Halfbeak failure time data (middle row) and the load-haul-dump data (lower row).

	ILR1	ILR2	ICvM	IKS	LR	CvM	KS	IELR1	SELR1	ELR	AD
USSH	0.028	0.002	0.023	0.005	0.006	0.009	0.029	0.032	0.013	0.090	0.001
LHD	0.99	0.18	0.55	0.54	0.50	0.13	0.29	0.004	0.013	0.006	0.086

Note: See the beginning of Section 5 for explanation of abbreviations. In addition, here the test ILR2 = Integrated Lewis-Robinson test version 2 is included. The ELR test is used with $a = 0.5$.

done in the present example. A plot of cumulative failure number against time is given in the right plot in Figure 6. The plot indicates a non-monotonic trend, apparently of bathtub type.

The last row of Table 1 gives calculated p-values for the same tests as in the previous example. It is seen that only the three tests that are constructed for the alternative of a bathtub type trend obtain a low p-value.

7 | CONCLUDING REMARKS

The paper extends the class of trend tests studied in Kvaløy and Lindqvist[6] by considering suitably integrated versions of the previous tests. As in Kvaløy and Lindqvist,[6] the observations are assumed to be time censored.

In an older paper, Kvaløy and Lindqvist [17] considered the event censored case. Recall that this means that observation is stopped at the nth event, for a fixed number $n \geq 1$. In such a case the data are hence given by the event times $T_1, T_2, \dots, T_n$, and the corresponding inter-event times $X_1, X_2, \dots, X_n$. In a similar way to (1), the basis of the derivations of Kvaløy and Lindqvist[17] was the process

$$V_{n,\gamma}(s) = \frac{\sqrt{n}}{\gamma}\left(\frac{T_{[ns]}}{T_n} + (ns - [ns])\frac{X_{[ns]+1}}{T_n} - s\right) \text{ for } 0 \leq s \leq 1. \tag{13}$$

Here $[t]$ means the integer part of t. Using Donsker's theorem[8] it was shown that the process $V_{n,\gamma}$ converges weakly to the Brownian bridge W^0 as $n \to \infty$. Kvaløy and Lindqvist[17] used this result to derive and study certain versions of classical trend tests. However, it is now clear that tests corresponding to the new tests of the present paper can be derived also for the event censored case.

It should be remarked that all the tests derived in the present paper, as well as in Kvaløy and Lindqvist[6] and Kvaløy and Lindqvist,[17] can be used as tests also for the null hypothesis of events coming from a homogeneous Poisson process. Indeed, in this case, the coefficient of variation, γ, of the inter-event times is known and equals 1. Thus, $\hat{\gamma}$ should then be replaced by 1 in the test statistics, whereas the limiting distributions and hence critical values will be the same as for the case of estimated γ. This follows from Theorem 1. Recall in this connection the fact, noted in Section 2, that putting $\hat{\gamma} = 1$ in the LR statistic (2) gives the Laplace test statistic.

Both papers Kvaløy and Lindqvist[6] and Kvaløy and Lindqvist[17] consider also the case when several similar processes are observed, and the interest is in testing the null hypothesis that they all have no trend. These papers then suggest various types of null hypotheses, and various types of tests extending the single process tests. In the same manner, one might consider multiple process versions of the new tests of the present paper. Such a study is, however, beyond the scope of the current paper.

DATA AVAILABILITY STATEMENT

The data sets and R code for the tests are available at: https://github.com/jtkgithub/trendtests.

ORCID

Bo Henry Lindqvist https://orcid.org/0000-0001-8952-9311
Jan Terje Kvaløy https://orcid.org/0000-0002-8829-6250

REFERENCES

1. Ascher H, Feingold H. *Repairable Systems Reliability. Modeling, Inference, Misconceptions and their Causes*. Marcel Dekker, Inc; 1984.
2. Cook R, Lawless J. *The Statistical Analysis of Recurrent Events*. Springer; 2007.
3. Lawless J, Çiğşar C, Cook R. Testing for monotone trend in recurrent event processes. *Technometrics*. 2012;54:147-158.
4. Lewis PAW, Robinson DW. Testing for a monotone trend in a modulated renewal process. In: Proschan F, Serfling RJ, eds. *Reliability and Biometry*. SIAM; 1974:163-182.
5. Kvaløy JT, Lindqvist BH. TTT-based tests for trend in repairable systems data. *Reliab Eng Syst Safe*. 1998;60:13-28.
6. Kvaløy JT, Lindqvist BH. A class of tests for trend in time censored recurrent event data. *Technometrics*. 2020;62(1):101-115.
7. Caroni C. "Failure limited" data and TTT-based trend tests in multiple repairable systems. *Reliab Eng Syst Safe*. 2010;95(6):704-706.
8. Billingsley P. *Convergence of Probability Measures*. Wiley; 1999.
9. Henze N, Nikitin YY. A new approach to goodness-of-fit testing based on the integrated empirical process. *Int J Comput Math*. 2000;12(3):391-416.
10. Lindqvist BH, Elvebakk G, Heggland K. The trend-renewal process for statistical analysis of repairable systems. *Technometrics*. 2003;45:31-44.

11. Kolmogorov A. Sulla determinazione empirica di una legge di distribuzione. *Giornale Dell' Istituto Italiano Degli Attuari*. 1933;4:83-91.
12. Lachal A. Sur l'intégrale du mouvement brownien. *Compt Rendus l'Acad Des Sci Série 1, Math*. 1990;311(7):461-464.
13. Lachal A. Sur la distribution de certaines fonctionnelles de l'intégrale du mouvement brownien avec dérives parabolique et cubique. *Commun Pure Appl Math*. 1996;49(12):1299-1338.
14. Anderson TW, Darling DA. Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Ann Math Stat*. 1952;23:193-212.
15. Meeker WQ, Escobar LA. *Statistical Methods for Reliability Data*. Wiley; 1998.
16. Kumar U, Klefsjö B, Granholm S. Reliability investigation for a fleet of load haul dump machines in a Swedish mine. *Reliab Eng Syst Safe*. 1989;24:341-361.
17. Kvaløy JT, Lindqvist BH. A class of tests for renewal process versus monontonic and nonmonotonic trend in repairable systems data. In: Lindqvist BH, Doksum KA, eds. *Mathematical and Statistical Methods in Reliability, Series on Quality, Reliability and Engineering Statistics*. Vol 7. World Scientific Publishing; 2003:401-414.

How to cite this article: Lindqvist BH, Kvaløy JT. New tests for trend in time censored recurrent event data. *Appl Stochastic Models Bus Ind*. 2024;1-16. doi: 10.1002/asmb.2848