Received: 4 January 2018 | Revised: 30 September 2019 | Accepted: 2 October 2019

DOI: 10.1002/nav.21875

RESEARCH ARTICLE

WILEY

Pairwise model discrimination with applications in lifetime distributions and degradation processes

Piao Chen[1] | **Zhi-Sheng Ye**[2] | **Xun Xiao**[3]

[1]Delft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands
[2]Department of Industrial Systems Engineering & Management, National University of Singapore, Singapore
[3]School of Fundamental Sciences, Massey University, Palmerston North, New Zealand

Correspondence
Zhi-Sheng Ye, Department of Industrial Systems Engineering & Management, National University of Singapore, Singapore.
Email: yez@nus.edu.sg

Funding information
Singapore MOE AcRF Tier 2, R-266-000-125-112. National Science Foundation of Jiangsu Province, BK20180232.

Abstract

Reliability data obtained from life tests and degradation tests have been extensively used for purposes such as estimating product reliability and predicting warranty costs. When there is more than one candidate model, an important task is to discriminate between the models. In the literature, the model discrimination was often treated as a hypothesis test and a pairwise model discrimination procedure was carried out. Because the null distribution of the test statistic is unavailable in most cases, the large sample approximation and the bootstrap were frequently used to find the acceptance region of the test. Although these two methods are asymptotically accurate, their performance in terms of size and power is not satisfactory in small sample size. To enhance the small-sample performance, we propose a new method to approximate the null distribution, which builds on the idea of generalized pivots. Conventionally, the generalized pivots were often used for interval estimation of a certain parameter or function of parameters in presence of nuisance parameters. In this study, we further extend the idea of generalized pivots to find the acceptance region of the model discrimination test. Through extensive simulations, we show that the proposed method performs better than the existing methods in discriminating between two lifetime distributions or two degradation models over a wide range of sample sizes. Two real examples are used to illustrate the proposed methods.

KEYWORDS

Bayesian, bootstrap, GPQ, hypothesis test, model selection, reliability

1 | INTRODUCTION

Reliability data such as lifetime data and degradation data from laboratory tests or designed experiments for products are widely available. The reliability data are valuable because they can be used to, for example, predict product reliability, predict product warranty costs, and assess the effect of a proposed design change (Su & Wang, 2016; Wang, Li, & Xie, in press). When analyzing the reliability data, a common approach is to assume a parametric model for the reliability data, especially when the data size is limited (Meeker & Escobar, 1998, chap. 1). Generally, there are often more than one candidate model for a certain reliability dataset. To make precise estimation and prediction, it is important to discriminate between the candidate models. The objective of this study is to propose a general framework for model discrimination between two reliability models.

One important source of the reliability information is from lifetime data (Zhao & Xie, 2017). Lifetimes are commonly modeled by continuous distributions. Some popular lifetime distributions include distributions in the log-location-scale family (eg, Weibull and log-normal), the gamma distribution and the inverse Gaussian (IG) distribution (Chen, Ye, & Zhai, in press). A bulk of the literature have developed model discrimination techniques among these distributions, and a pairwise model discrimination procedure is often carried out (Schennach & Wilhelm, 2017; Vuong, 1989). In most of the existing studies, discrimination between two distributions is treated as a hypothesis test problem, and the test statistic is constructed as the logarithm of the ratio of maximized likelihoods (RML). For example, Kundu and Manglick (2004) considered model discrimination between the Weibull distribution and the log-normal distribution based on the asymptotic distribution of the RML. Although this asymptotic

distribution has a closed form, its performance in terms of the size and the power is not satisfactory with small sample sizes. Bain and Engelhardt (1980) considered model discrimination between the Weibull distribution and the gamma distribution, and the critical values for the RML under different parameter values and sample sizes are provided by Monte Carlo simulations. This method is not easy to implement as it requires looking up values in a table. In addition, because the parameters need to be estimated in practice, this method is not accurate when the sample size is small. On the other hand, Shao and Tu (2012, chap. 4) proposed to use the bootstrap to approximate the distribution of the test statistic. As we will see later, there is still room for improvement of the bootstrap in small sample sizes.

In summary, the existing methods to obtain the critical values of the RML are often not satisfactory when the sample sizes are limited. In practice, a small size of the lifetime data is not uncommon; see the lifetime datasets reported in Lawless (2003, chap. 1) for a few examples. In view of this fact, a new model discrimination procedure that suffices for a wider range of data sizes is developed in this study. The key idea of the proposed methods is to use the generalized pivotal quantities (GPQs). The GPQ method was proposed by Weerahandi (1993) and it was used for interval estimation of a parameter in the presence of other nuisance parameters, when the conventional pivotal quantities are difficult to construct. Confidence intervals of the parameter of interest can be constructed based on the distribution of its GPQ. Recent applications of the GPQ method can be found in Hannig, Iyer, Lai, and Lee (2016), Yu, Mathew, and Zhu (2017), Wang and Wu (2017), to name a few. An appealing feature of the GPQ method is that GPQ of a function of parameters can be constructed from GPQs of the parameters by using substitution. In the context of model discrimination, the test statistic is a function of the model parameters and the data. Due to the intimate relationship between the hypothesis test and the interval estimation, once the GPQs of the parameters are constructed, the GPQ method can be used to find the acceptance region of the hypothesis test. In the literature, the GPQ method has been validated by asymptotic theory and by simulation in numerous small sample problems (eg, Hannig et al., 2016). In this study, extensive simulations are conducted to compare the GPQ method with the large-sample approximation method and the bootstrap.

On the other hand, our proposed methods can be naturally extended to degradation data. As modern products become more and more reliable, it is sometimes difficult to collect product failure times in a traditional life test, even after a long period of time. In such cases, degradation tests are widely adopted. In a degradation experiment, degradation levels of each test unit are recorded over time, and the resulting degradation data are useful for reliability evaluation. In terms of degradation modeling, a continuous-state stochastic process is often assumed for the degradation data. In the literature, the Wiener process, the gamma process, and the IG process have been widely used for degradation modeling (eg, Ye, Xie, Tang, & Chen, 2014; Zhai & Ye, 2017; Zhao, He, Kuo, & Xie, in press). Given a degradation dataset, the first task is to discriminate between these processes and choose an appropriate model for the data. To the best of our knowledge, no formal studies have been done on this discrimination problem. In this study, discrimination between the degradation processes is treated as a hypothesis test, and the RML test statistics is constructed. The GPQ method is then used to find the acceptance region of the hypothesis test. The performance of the proposed method will be assessed by simulations using a wide range of sample sizes.

The remainder of the paper is organized as follows. In Section 2, we introduce common lifetime models and degradation models. In Section 3 and 4, model discrimination procedures between lifetime distributions and degradation models are respectively proposed. In each section, simulations are conducted to assess the performance of the proposed methods and a real example is used for illustration. At last, Section 5 concludes the paper and provides some further comments.

2 | PRELIMINARIES AND MODELS

2.1 | Lifetime models

The log-location-scale family of distributions provides a good fit to many lifetime datasets. A random variable X belongs to the location-scale family of distributions if its cumulative distribution function (CDF) $F(x)$ can be expressed as

$$F(x; a, b) = G\left(\frac{x-a}{b}\right), \quad x \in \mathbb{R},$$

where $a \in \mathbb{R}$ is a location parameter, $b > 0$ is a scale parameter, and G does not depend on any unknown parameters. On the other hand, X belongs to the log-location-scale family of distributions if $\log(X)$ is a location-scale distribution. Two important log-location-scale parametric distributions are the Weibull distribution and the log-normal distribution, and they are extensively used to model product lifetime. The probability density function (PDF) of a Weibull distribution is

$$f_{WB}(x; k, \lambda) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left[-\left[\frac{x}{\lambda}\right]^{k}\right], \quad x > 0,$$

where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. On the other hand, a log-normal distribution $LN(\mu, \sigma^2)$ with location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$ has a PDF

$$f_{LN}(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right], \quad x > 0.$$

Two popular alternative lifetime models to the log-location-scale distributions are the gamma distribution and the IG distribution. The gamma distribution $GA(\vartheta, \beta)$

has a PDF

$$f_{GA}(x;\vartheta,\beta)=\frac{\beta^{\vartheta}}{\Gamma(\vartheta)}x^{\vartheta-1}e^{-\beta x},\quad x>0,$$

where $\vartheta>0$ is the shape parameter and $\beta>0$ is the rate parameter. On the other hand, the IG distribution $\mathcal{IG}(\mu,\lambda)$ has the following PDF

$$f_{\mathcal{IG}}(x;\mu,\lambda)=\left(\frac{\lambda}{2\pi x^3}\right)^{1/2}\exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right],\quad x>0,$$

where $\mu>0$ is the mean parameter and $\lambda>0$ is the shape parameter.

2.2 | Degradation models

Consider a stochastic degradation process $\{Y(t);\ t>0\}$. A product is defined to have failed when the degradation process first reaches a degradation failure threshold D. Three popular models for degradation data are the Wiener process, the gamma process and the IG process. The Wiener process assumes that $Y(t)\sim N(\mu t,\sigma^2 t)$, where $N(\mu,\sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. Under the gamma process model, $Y(t)$ follows a gamma distribution with shape parameter ϑt and rate parameter β, that is, $Y(t)\sim GA(\vartheta t,\beta)$. On the other hand, the IG process model assumes that $Y(t)$ follows an IG distribution with mean parameter μt and shape parameter λt^2, that is, $Y(t)\sim\mathcal{IG}(\mu t,\lambda t^2)$.

3 | DISCRIMINATION BETWEEN LIFETIME MODELS

Consider two distribution families with respective densities $f_1(x;\boldsymbol{\theta})$ and $f_2(x;\boldsymbol{\xi})$, where $\boldsymbol{\theta}\in\Theta$ and $\boldsymbol{\xi}\in\Xi$ are the model parameters, and Θ and Ξ are the corresponding parameter spaces. It is common to formulate the discrimination problem as the following hypothesis test (Schennach & Wilhelm, 2017)

$$\begin{aligned}&H_0:\ X\sim f_1(x;\boldsymbol{\theta}_0)\ \text{ for some }\ \boldsymbol{\theta}_0\in\Theta,\\ &\text{and}\quad H_1:\ X\sim f_2(x;\boldsymbol{\xi}_0)\ \text{ for some }\ \boldsymbol{\xi}_0\in\Xi.\end{aligned}\tag{1}$$

Based on the observed lifetime data $\mathbf{D}=\{X_1,\ldots,X_n\}$, the RML is then defined as

$$\text{RML}=\frac{\max\prod_{i=1}^n f_1(X_i;\boldsymbol{\theta})}{\max\prod_{i=1}^n f_2(X_i;\boldsymbol{\xi})}=\frac{\prod_{i=1}^n f_1(X_i;\widehat{\boldsymbol{\theta}}_n)}{\prod_{i=1}^n f_2(X_i;\widehat{\boldsymbol{\xi}}_n)},\tag{2}$$

where $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\boldsymbol{\xi}}_n$ are the ML estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ based on $\mathbf{D}$, respectively. ML estimators of commonly used distributions, including the log-location-scale distribution family, and the gamma and IG distributions, can be found in Meeker and Escobar (1998). The test statistic $\mathcal{T}$ is constructed as the logarithm of RML, that is,

$$\begin{aligned}\mathcal{T}(\mathbf{D})\equiv\mathcal{T}&=\log(\text{RML})=\sum_{i=1}^n[\log f_1(X_i;\widehat{\boldsymbol{\theta}}_n)-\log f_2(X_i;\widehat{\boldsymbol{\xi}}_n)]\\&=\ell_1(\widehat{\boldsymbol{\theta}}_n;\mathbf{D})-\ell_2(\widehat{\boldsymbol{\xi}}_n;\mathbf{D}),\end{aligned}\tag{3}$$

where $\ell_i(\cdot;\mathbf{D})$ denotes log-likelihood function under the ith model, $i=1,2$.

As seen from (3), the test statistic $\mathcal{T}$ is a function of $\mathbf{D}$, and the distribution of $\mathcal{T}$ depends on the underlying model that generates $\mathbf{D}$. Under the null hypothesis, we only know that $\boldsymbol{\theta}_0\in\Theta$, and the true value of $\boldsymbol{\theta}_0$ is unknown. This means that the distribution of $\mathcal{T}$ under the null is generally unknown. As a result, it is difficult to determine accurate critical values for the test. In the literature, the large-sample approximation (eg, Kundu & Manglick, 2004) and the bootstrap (eg, Shao & Tu, 2012, chap. 4) are frequently used to approximate the distribution of $\mathcal{T}$. Due to its asymptotic nature, the large-sample approximation often performs unsatisfactorily in small samples. On the other hand, the bootstrap tries to replace $\boldsymbol{\theta}_0$ by $\widehat{\boldsymbol{\theta}}_n$ and then generate samples from $f_1(x;\widehat{\boldsymbol{\theta}}_n)$. However, when the sample size is small, the estimator $\widehat{\boldsymbol{\theta}}_n$ may have large mean square errors (MSEs) and the bootstrap may not perform well, as evidenced by the simulation results in Section 3.1.

Our idea to approximate the distribution of $\mathcal{T}$ is based on the GPQ method. As reported by many studies (eg, Wang & Wu, 2017; Yu et al., 2017), the GPQ related methods often show small sample performance superior to other competing methods. Similar to the bootstrap, our proposed procedure is also a resampling-based method. Consider the distribution $f_1(\cdot;\boldsymbol{\theta}_0)$ under the null hypothesis. Let $\mathcal{G}_{\boldsymbol{\theta}}$ be the GPQ of $\boldsymbol{\theta}_0$ based on the observed lifetime data $\mathbf{D}$. Detailed expressions of $\mathcal{G}_{\boldsymbol{\theta}}$ depend on the distribution f_1. Appendices A and B respectively provide GPQs for the log-location-scale distribution families and the gamma distribution. GPQs for IG distribution could be found in Krishnamoorthy and Tian (2008).

Unlike the bootstrap that resamples from $f_1(\cdot,\widehat{\boldsymbol{\theta}}_n)$, our proposed method consists of two steps. In the first step, a realization $\boldsymbol{\theta}$ is imputed from $\mathcal{G}_{\boldsymbol{\theta}}$. Then, a pseudo lifetime dataset $\mathbb{D}$ of size n is sampled from $f_1(\cdot,\boldsymbol{\theta})$. Based on $\mathbb{D}$, a realization of the test statistic $\mathcal{T}(\mathbb{D})$ can be obtained. An advantage of the proposed method is that the estimation errors in $\widehat{\boldsymbol{\theta}}_n$ are taken into account in the resampling procedure. Many existing studies have revealed the good performance of the GPQs in approximating the distribution of $\widehat{\boldsymbol{\theta}}_n$. By using these GPQs to account for the uncertainties in the resampling procedure, it can be reasonably believed that the performance can be greatly enhanced, especially when the sample sizes are small and the variation in $\widehat{\boldsymbol{\theta}}_n$ is significant.

When applying to model discrimination, the above resampling procedure can be repeated for a number of times, after which a sufficient number of realizations of $\mathcal{T}$ are obtained. Percentiles of these realizations of $\mathcal{T}$ are then used to obtain

a rejection region for the hypothesis test (1). The detailed procedure is summarized in Algorithm 1.

Algorithm 1: Decision procedure for the model-discrimination between lifetime models

Step 1. Based on the observed data $\mathbf{D} = \{X_1, \ldots, X_n\}$, compute $\mathcal{T}$ by (3), and denote it as $\mathcal{T}_0$.
Step 2. Generate B realizations of $\boldsymbol{\theta}$ from the distributions of the GPQs $\mathcal{G}_{\boldsymbol{\theta}}$, and denote them as $(\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(B)})$.
Step 3. For each $\boldsymbol{\theta}^{(b)}$, $b = 1, \ldots, B$, generate a sample $\mathbb{D}^{(b)} = \{X_1^{(b)}, \ldots, X_n^{(b)}\}$. Based on $\mathbb{D}^{(b)}$, compute the ML estimates $\widehat{\boldsymbol{\theta}}_n^{(b)}$ and $\widehat{\boldsymbol{\xi}}_n^{(b)}$ for the two models. Then, compute $\mathcal{T}^{(b)}$ by (3).
Step 4. Given a fixed level α, the critical value $\mathcal{T}_c$ is set as the α empirical percentile of $\{\mathcal{T}^{(1)}, \ldots, \mathcal{T}^{(B)}\}$.
Step 5. Accept the null hypothesis if $\mathcal{T}_0 > \mathcal{T}_c$; otherwise accept the alternative hypothesis.

In case that both models f_1 and f_2 belong to the log-location-scale family (eg, the Weibull distribution and the log-normal distribution), we can prove that the test statistic (3) is independent of the unknown parameters, as shown in Theorem 1. The proof is given in Appendix E.

Theorem 1 *If both f_1 and f_2 belong to the log-location-scale family, the test statistic in* (3) *is independent of the unknown parameters.*

Based on Theorem 1, we can further show that the test (3) based on Algorithm 1 is Monte Carlo exact when f_1 and f_2 are both from the log-location-scale family. That is, the test (3) has a Type I error equal to the significance level when B in Algorithm 1 goes to infinity. The proof of Theorem 2 is also given in Appendix E.

Theorem 2 *If both f_1 and f_2 belong to the log-location-scale family, the hypothesis test* (1) *based on Algorithm* 1 *is Monte Carlo exact.*

3.1 | Simulation study

Simulation is conducted to assess the performance of Algorithm 1. Throughout this section, we set $\alpha = 0.05$ and $B = 10{,}000$ in Algorithm 1. We first consider the model selection between the log-normal distribution and the Weibull distribution, and the hypothesis test is

$$H_0 : X \sim LN(\mu, \sigma^2) \text{ and } H_1 : X \sim WB(k, \lambda). \quad (4)$$

Denote by $\mathcal{T}_{LW}$ the test statistic for this hypothesis. A good property of $\mathcal{T}_{LW}$ is that its distribution is free of parameters μ and σ under the null hypothesis (see Theorem 1). Therefore, the exact distribution of $\mathcal{T}_{LW}$ under the null hypothesis can be obtained by the Monte Carlo simulation. On the other hand, Kundu and Manglick (2004) showed that $n^{-1/2}(\mathcal{T}_{LW} - 0.0810614n)$ asymptotically follows $N(0, 0.2182818)$ under the null hypothesis. Figure 1 shows that our proposed GPQ method yields the almost exact distribution of $\mathcal{T}_{LW}$ under the null hypothesis, which validates Theorem 2. On the other hand, the asymptotic approximation does not work well when the sample size n is small.

Next we consider the power of the hypothesis test (4) based on the GPQ method and the asymptotic approach. Assume the data are from $WB(k, \lambda)$. Given a fixed significance level α of the hypothesis test, the rejection criterion for the GPQ method is stated in Algorithm 1. For the asymptotic method, we reject the null hypothesis when $\mathcal{T}_{LW}$ based on the observed data is less than $0.0810614n - z_{1-\alpha}\sqrt{0.2182818n}$, where $z_{1-\alpha}$ is the $1-\alpha$ percentile of a standard normal distribution. We fix $\lambda = 1$ and consider $k = 1, 3, 5$. For each parameter setting, $n = 10, 20, \ldots, 100$ are considered. The statistical power based on 10,000 replications is shown in Figure 2. As can

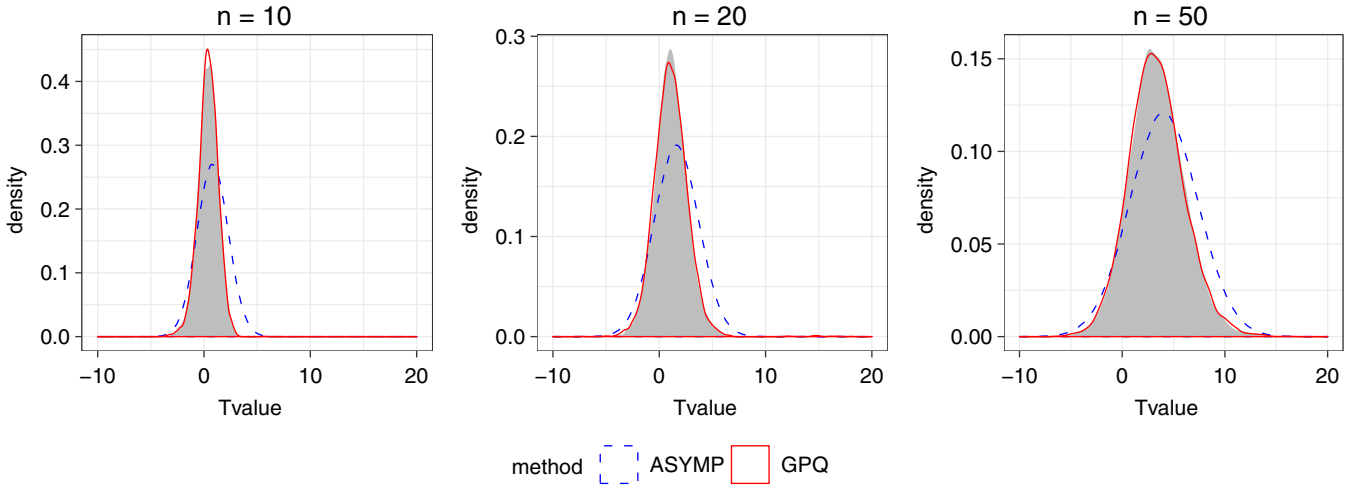

FIGURE 1 The distribution of $\mathcal{T}_{LW}$ based on the Monte Carlo simulation (gray area), the proposed generalized pivotal quantities method (GPQ) and the asymptotic approximation (ASYMP) when $n = 10, 20, 50$ [Colour figure can be viewed at wileyonlinelibrary.com]

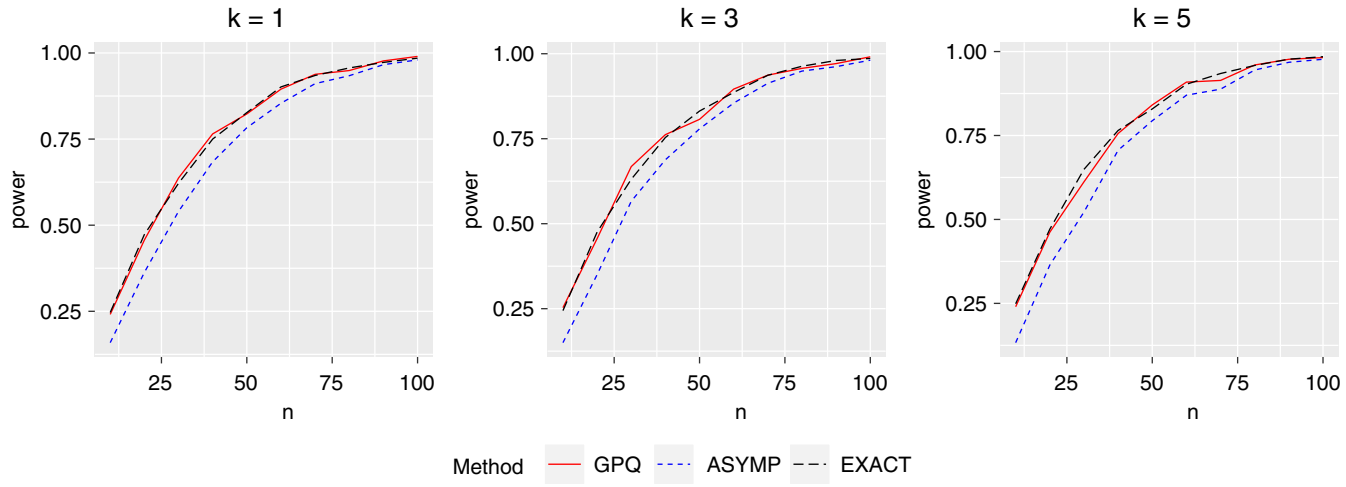

FIGURE 2 Statistical power for model discrimination between the log-normal distribution and the Weibull distribution based on the proposed generalized pivotal quantities method (GPQ) and the asymptotic approximation (ASYMP) when $k = 1, 3, 5$ and $n = 10, 20, \ldots, 100$ [Colour figure can be viewed at wileyonlinelibrary.com]

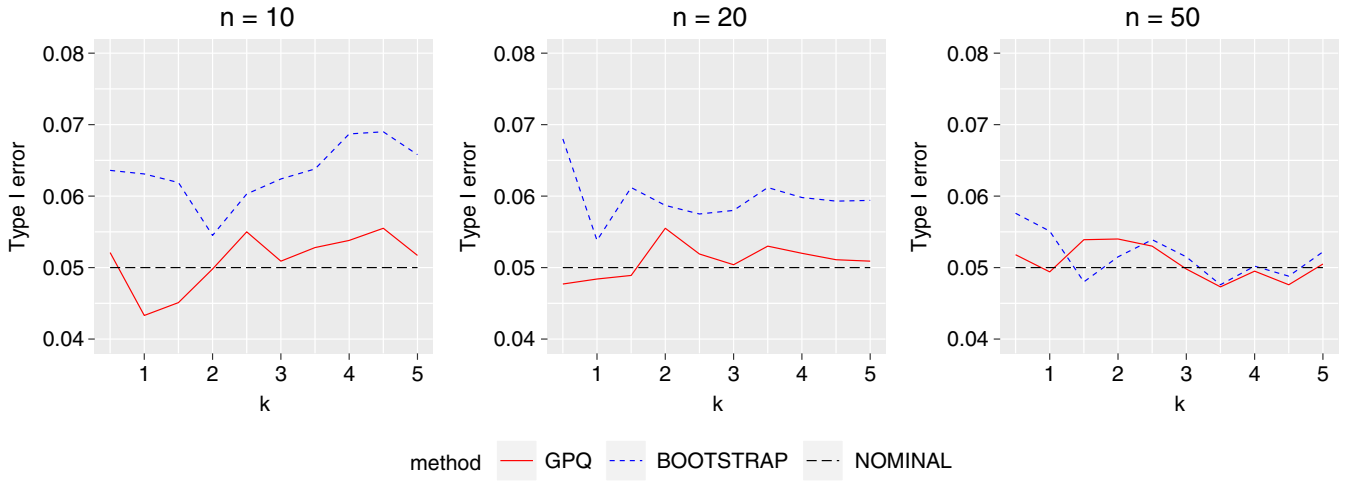

FIGURE 3 Type I error for model discrimination between the Weibull distribution and the gamma distribution based on the proposed generalized pivotal quantities method (GPQ) and the percentile bootstrap (BOOTSTRAP) when $n = 10, 20, 50$ and $k = 0.5, 1, \ldots, 4.5, 5$ [Colour figure can be viewed at wileyonlinelibrary.com]

be seen, the GPQ method is Monte Carlo exact, and it is uniformly more powerful than the asymptotic method.

We then consider model discrimination between the Weibull distribution and the gamma distribution. The hypothesis test is formulated as

$$H_0 : X \sim WB(k, \lambda) \quad \text{and} \quad H_1 : X \sim GA(\vartheta, \beta). \tag{5}$$

Denote by $\mathcal{T}_{WG}$ the test statistic for this test. We compare the size and the power of the hypothesis test (5) based on Algorithm 1 and the percentile bootstrap method. We do not consider the asymptotic method because the asymptotic distribution of $\mathcal{T}_{WG}$ is not in closed form. In the percentile bootstrap method, we generate B bootstrap samples from $WB(\widehat{k}, \widehat{\lambda})$, based on which $\widehat{\mathcal{T}}_{WG,b}$, $b = 1, 2, \ldots, B$ are obtained. Given the level of significance α, we select the Weibull distribution when $\mathcal{T}_{WG}$ based on the observed data is larger than the α percentile of $\{\widehat{\mathcal{T}}_{WG,1}, \ldots, \widehat{\mathcal{T}}_{WG,B}\}$, and otherwise select the gamma distribution. During this simulation, we set $B = 10{,}000$ for both Algorithm 1 and the bootstrap method. According to our preliminary simulation (not shown), $B = 10{,}000$ often suffices to achieve a reasonable approximation, and a larger value of B (eg, $B = 100{,}000$) often yields almost identical results.

We first generate data from $WB(k, 1)$ with $k = 0.5, 1, \ldots, 4.5, 5$, and consider $n = 10, 20, 50$. The Type I errors based on 10,000 replications are shown in Figure 3. As seen, Algorithm 1 accurately controls the Type I error. The poor small-sample performance of the bootstrap is due to the relative large MSE of $\widehat{k}$ when n is small. On the contrary, the proposed method inherently takes the variation in $\widehat{k}$ into account and hence achieves a satisfactory performance even in small sample size. According to our simulation results (not shown), another advantage of our proposed method over the bootstrap is that much less computation time is required. Because both methods perform similarly in terms of the statistical powers, the related figures are not reported to save space.

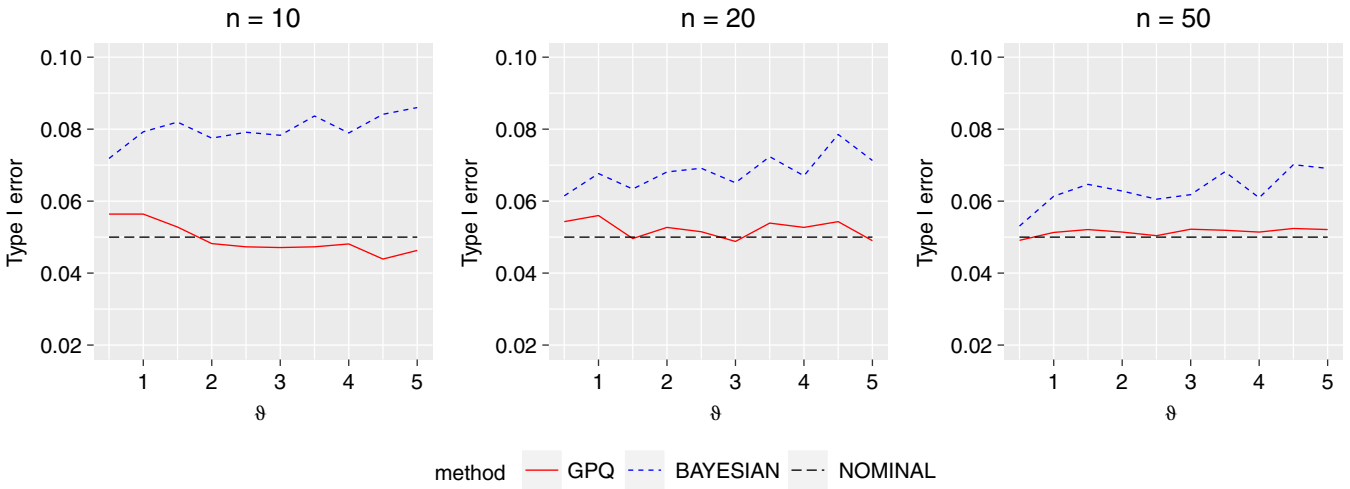

FIGURE 4 Type I error for model discrimination between the gamma distribution and the log-normal distribution based on the proposed generalized pivotal quantities method (GPQ) and the Bayesian method (BAYESIAN) when $n = 10, 20, 50$ and $\vartheta = 0.5, 1, \ldots, 4.5, 5$ [Colour figure can be viewed at wileyonlinelibrary.com]

Our last simulation considers model discrimination between the gamma distribution $GA(\vartheta, \beta)$ and the log-normal distribution $LN(\mu, \sigma^2)$. The hypothesis test is

$$H_0 : X \sim GA(\vartheta, \beta) \quad \text{and} \quad H_1 : X \sim LN(\mu, \sigma^2), \tag{6}$$

and the test statistic is denoted as $\mathcal{T}_{GL}$. In this scenario, we compare the proposed method with the Bayesian method in terms of the Type I error and the statistical power. Since no past information is available, the inverse gamma distribution, which is one of the popular noninformative priors, is assumed for the gamma parameters ϑ and β. A random variable $X \sim IGA(\vartheta, \beta)$ if $1/X \sim GA(\vartheta, \beta)$. The inverse gamma distribution is commonly used as a prior in Bayes analysis; see Huang, Wand, et al. (2013) and Rhodes, Turner, and Higgins (2015) for a few examples. We set $\vartheta = \beta = 0.1$ so that the prior is flat and the parameter can explore the whole positive real line. Based on Bayes' rule, the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{D})$ is proportional to $\pi(\boldsymbol{\theta})L(\boldsymbol{\theta}|\mathbf{D})$, where $\pi(\boldsymbol{\theta})$ is the prior distribution and $L(\boldsymbol{\theta}|\mathbf{D})$ is the likelihood function based on the gamma distribution. The random-walk Metropolis algorithm is then used for sampling from the posterior distribution. In the algorithm, the multivariate normal distribution can be used as the proposal distribution as well as the distribution of the random increment (Tierney, 1994). The detailed procedure of the random-walk Metropolis algorithm can be found in Givens and Hoeting (2012, sec 7.1) and it can be implemented by using the metrop function in the R package "mcmc". Based on B posterior samples of $\boldsymbol{\theta}$, B realizations of the test statistic are obtained and the critical values is set as α percentile of these realizations.

We first consider the Type I error by generating data from $GA(\vartheta, 1)$ with $\vartheta = 0.5, 1, \ldots, 5$. In addition, sample sizes $n = 10, 20, 50$ are considered. The Type I errors in Figure 4 are estimated based on 10,000 replications. As we can see, the Type I errors based on our proposed method are close to the nominal values in all cases, while the Bayesian method does not perform well in small sample sizes. In terms of the statistical power, both methods have the similar performance and hence the figures are not reported. We do note that the performance of the Bayesian method may be improved by carefully selecting the prior distributions, which is a topic beyond the scope of this study. The proposed method, on the other hand, requires no prior distribution and hence is more applicable in practice.

3.2 | Illustrative example

Drill lifetime data from a manufacturing factory in Singapore, as shown in Table 1, are used to illustrate the proposed methods (Chen & Ye, 2017). To conduct reliability analysis of the drill, it is desirable to find an appropriate model for the drill lifetime. Two common lifetime models, that is, the Weibull distribution and the gamma distribution, are considered in this case. Figure 5 shows the probability plots of the drill lifetime based on the fitted gamma and Weibull distributions. As seen, it is difficult to discriminate between these two distributions based on the probability plots only. Moreover, the Kolmogorov-Smirnov (KS) test gives respective P-values of .897 and .672 in testing the goodness-of-fit of the Weibull distribution and the gamma distribution, meaning that neither model could be rejected by the KS test. The same conclusion applies if other classical goodness-of-fit tests such as the Anderson-Darling test and the Cramér-von Mises test are conducted. In order to select a better model, the proposed model discrimination procedure is then adopted. Because of the extensive use of the Weibull distribution in the industry, the Weibull distribution is preferred unless there is convincing evidence against it. Consequently, the Weibull distribution is first assigned to the null hypothesis, and the gamma distribution is considered as the alternative. Based on Algorithm 1 with $B = 10{,}000$, the P-value of the hypothesis is .048, indicating an inadequate fit of the Weibull distribution to the

TABLE 1 Drill lifetimes data (in minutes)

105	105	95	87	112	80	95	97	77	103	78	87	107	96	79
91	108	97	80	76	92	85	76	96	77	80	100	94	82	104
91	95	93	99	99	94	84	99	91	85	86	79	89	89	100

TABLE 2 *P*-values for discrimination between the Weibull and the gamma distributions based on the GPQ method, the bootstrap method and the Bayesian method

H_0	GPQ	Bootstrap	Bayesian
Weibull	0.048	0.046	0.043
Gamma	0.300	0.460	0.312

data. If the gamma distribution is assigned to the null hypothesis and the Weibull distribution the alternative, the *P*-value is .300. Based on the above results, the gamma distribution is recommended for the drill lifetime.

For comparison purposes, we also consider the bootstrap method and the Bayesian method for this discrimination problem. Analogous to Section 3.1, the percentile bootstrap is used and the *P*-values are estimated based on 10,000 replications. In terms of the Bayesian method, the inverse gamma distribution $IGA(0.1, 0.1)$ is used as the prior for the gamma and Weibull parameters, and the *P*-values are obtained based on 10,000 posterior samples. Table 2 summarizes the *P*-values from the proposed GPQ method, the bootstrap method, and the Bayesian method. As seen, these methods lead to consistent conclusions and they all favor the gamma distribution for the drill lifetimes.

4 | DISCRIMINATION BETWEEN DEGRADATION MODELS

In this section, we consider model discrimination between the degradation models. The hypothesis test is formulated as

$$H_0 : Y(t) \sim P_1(t, \boldsymbol{\theta}_0) \text{ for some } \boldsymbol{\theta}_0 \in \Theta, \quad \text{and} \quad H_1 : Y(t) \sim P_2(t, \boldsymbol{\xi}_0) \text{ for some } \boldsymbol{\xi}_0 \in \Xi, \tag{7}$$

where P_1 and P_2 are two different processes, $\boldsymbol{\theta}_0$ and $\boldsymbol{\xi}_0$ are the model parameters, and Θ and Ξ are the corresponding parameter spaces.

Degradation data are used for the above discrimination. Suppose degradation data from n independent and identical units are observed. Let $\mathbf{t}_i = \{t_{i,j}; j = 0, 1, \ldots, m_i\}$ be the ordered observation times for the ith unit and $\mathbf{Y}_i = \{Y_i(t_{i,j}); j = 0, 1, \ldots, m_i\}$ be the corresponding degradation levels, where $t_{i,0} = 0$ and $Y_i(t_{i,0}) = 0$ by convention. The sample data collected from the n units are denoted as $\mathbf{D} = \{(\mathbf{t}_i, \mathbf{Y}_i); i = 1, 2, \ldots, n\}$. Similar to the discrimination between two parametric distributions, the test statistic for (7) is constructed as

$$\mathcal{T}(\mathbf{D}) \equiv \mathcal{T} = \ell_1(\hat{\boldsymbol{\theta}}_n; \mathbf{D}) - \ell_2(\hat{\boldsymbol{\xi}}_n; \mathbf{D}), \tag{8}$$

where $\ell_i(\cdot; \mathbf{D})$ is the log-likelihood function under model P_i, $i = 1, 2$. ML estimation of the Wiener process, the gamma process and the IG process was well studied in the literature, for example, see Kahle and Lehmann (2010), Lawless and Crowder (2004), and Wang and Xu (2010). Because the distribution of $\mathcal{T}$ under the null hypothesis depends on the unknown parameters $\boldsymbol{\theta}_0$, it is difficult to find an accurate acceptance region for (7). On the other hand, the GPQ method can be again invoked. In specific, realizations of $\boldsymbol{\theta}$ are first imputed from $\mathcal{G}_{\boldsymbol{\theta}}$, that is, the GPQ of $\boldsymbol{\theta}_0$ based on the observed degradation data $\mathbf{D}$. GPQs for the Wiener process and the IG process are respectively provided in Appendices C and D, and GPQs for the gamma process and can be found in Chen and Ye (2019). For each imputation of $\boldsymbol{\theta}$ from $\mathcal{G}_{\boldsymbol{\theta}}$, a pseudo degradation dataset is sampled from $P_1(\cdot, \boldsymbol{\theta})$, and then a realization of the test statistic $\mathcal{T}$ can be obtained. The percentiles of the realizations of $\mathcal{T}$ can be used to determine critical values of the hypothesis (7). The detailed procedure for discrimination between two degradation models is summarized in Algorithm 2.

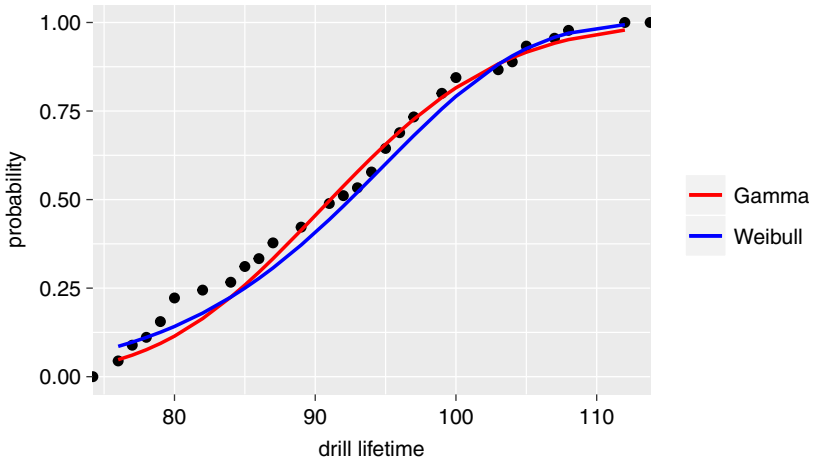

FIGURE 5 Probability plots of the drill lifetime based on the fitted gamma and Weibull distributions. The circles denote the estimated empirical probability [Colour figure can be viewed at wileyonlinelibrary.com]

Algorithm 2: Decision procedure for the model-discrimination between degradation models

Step 1. Based on the observed data $\mathbf{D} = \{(\mathbf{t}_i, \mathbf{Y}_i); i = 1, 2, \ldots, n\}$, compute $\mathcal{T}$ by (8), and denote it as $\mathcal{T}_0$.
Step 2. Generate B realizations of $\boldsymbol{\theta}$ from the distributions of their GPQs $\mathcal{G}_{\boldsymbol{\theta}}$, and denote them as $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(B)}$.
Step 3. For each $\boldsymbol{\theta}^{(b)}$, $b = 1, \ldots, B$, generate a sample $\mathbf{Y}_i^{(b)}$ based on the observation times $\mathbf{t}_i$, $i = 1, 2, \ldots, n$. Based on $\mathbb{D}^{(b)} = \{(\mathbf{t}_i, \mathbf{Y}_i^{(b)}); i = 1, 2, \ldots, n\}$, compute the ML estimates $\widehat{\boldsymbol{\theta}}_n^{(b)}$ and $\widehat{\boldsymbol{\xi}}_n^{(b)}$ and the corresponding $\mathcal{T}^{(b)}$ by (8).
Step 4. Given a fixed level α, the critical value $\mathcal{T}_c$ is set as the α percentile of $\{\mathcal{T}^{(1)}, \ldots, \mathcal{T}^{(B)}\}$.
Step 5. Accept the null hypothesis if $\mathcal{T}_0 > \mathcal{T}_c$; otherwise accept the alternative hypothesis.

4.1 | Simulation study

In this subsection, we conduct simulations to assess the performance of Algorithm 2. Throughout this section, we set $\alpha = 0.05$ and $B = 10{,}000$ in Algorithm 2. We first consider the hypothesis

$$H_0 : Y(t) \sim GA(\vartheta t, \beta) \text{ and } H_1 : Y(t) \sim \mathcal{IG}(\mu t, \lambda t^2), \quad (9)$$

where $GA(\vartheta, \beta)$ denotes the gamma distribution with shape parameter ϑ and rate parameter β, and $\mathcal{IG}(\mu, \lambda)$ denotes the IG distribution with mean parameter μ and shape parameter λ. In the Bayesian method, the inverse gamma distribution $IGA(0.1, 0.1)$ is selected as the priors for the model parameters. 10,000 posterior samples of the parameters are first generated by using the metrop function in the R package "mcmc" and these samples are then used to obtain 10, 000 realizations of the test statistics, whose α percentile is treated as the critical value by the Bayesian method. Without loss of generality, we use common observation times $\mathbf{t} = \{1, 3, 6, 10, 15\}$ for all the n units. We first generate degradation data from the gamma process with $\vartheta = 0.5, 1, \ldots, 4.5, 5$ and $\beta = 1$. In addition, $n = 5, 10, 20$ are considered. Based on 10,000 replications, the Type I errors are shown in Figure 6. As seen, the proposed method has an overall satisfactory performance in controlling the Type I error. On the other hand, although the Bayesian method seems to be inaccurate when $\vartheta < 2.5$, it has a similar performance as the proposed method when $\vartheta \geq 2.5$. We suspect this is due to the prior settings. To verify this, we consider another prior $U(0, 10)$ and observe that the proposed method performs uniformly better than the Bayesian method. This again implies that the proposed method is more reliable than the Bayesian method in practical applications as determining an appropriate prior distribution can be a quite difficult task. In terms of the power, both methods have similar performance and hence the results are omitted.

Our last simulation in this section considers

$$H_0 : Y(t) \sim \mathcal{IG}(\mu t, \lambda t^2) \text{ and } H_1 : Y(t) \sim GA(\vartheta t, \beta). \quad (10)$$

Similar to the previous simulation, we assess the performance of Algorithm 2 in terms of size, which is estimated by generating data from the IG process with $\mu = 0.5, 1, \ldots, 4.5, 5$, $\lambda = 1$ and $n = 5, 10, 20$. Figure 7 shows the Type I errors by using the proposed method and the bootstrap. As expected, the proposed method maintains a close Type I error to the nominal value while the bootstrap seems a bit conservative. Based on our comprehensive simulation study (not shown here), this is because the distribution of the test statistic $\mathcal{T}$ approximated by the bootstrap method is left skewed compared to the GPQ method. As a result, the critical value $\mathcal{T}_c$ based on the bootstrap tends to be smaller than that based on the GPQ method, and hence the bootstrap exhibits a conservative performance in controlling the Type I error. In terms of the power, both methods perform similarly and hence the results are not reported.

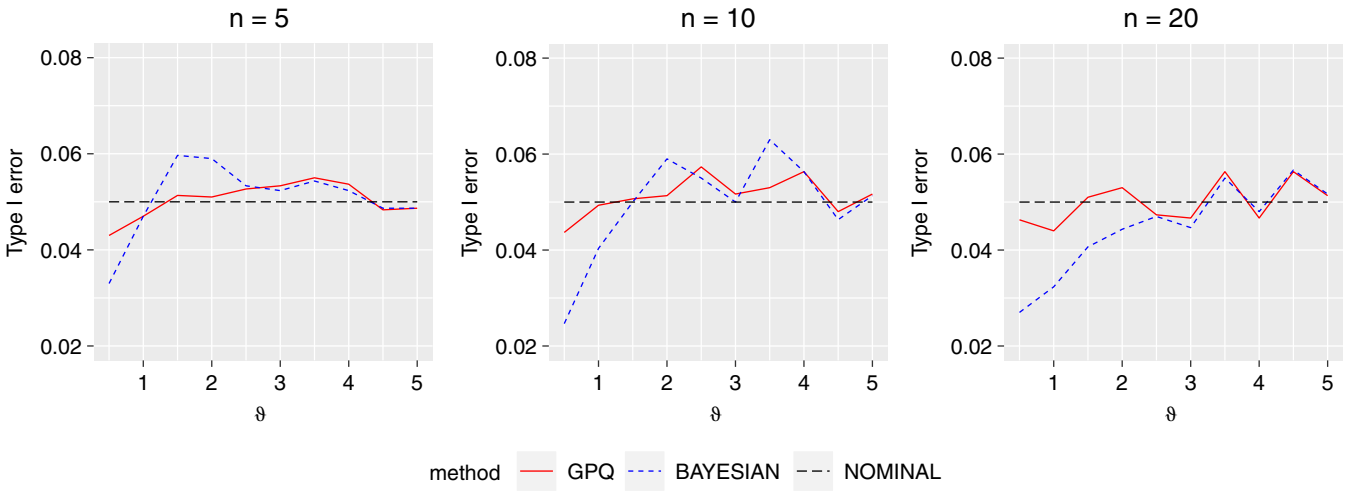

FIGURE 6 Type I error for model discrimination between the gamma process and the inverse Gaussian process based on the proposed generalized pivotal quantities method (GPQ) and the Bayesian method (BAYESIAN) when $n = 5, 10, 20$ and $\vartheta = 0.5, 1, \ldots, 4.5, 5$ [Colour figure can be viewed at wileyonlinelibrary.com]

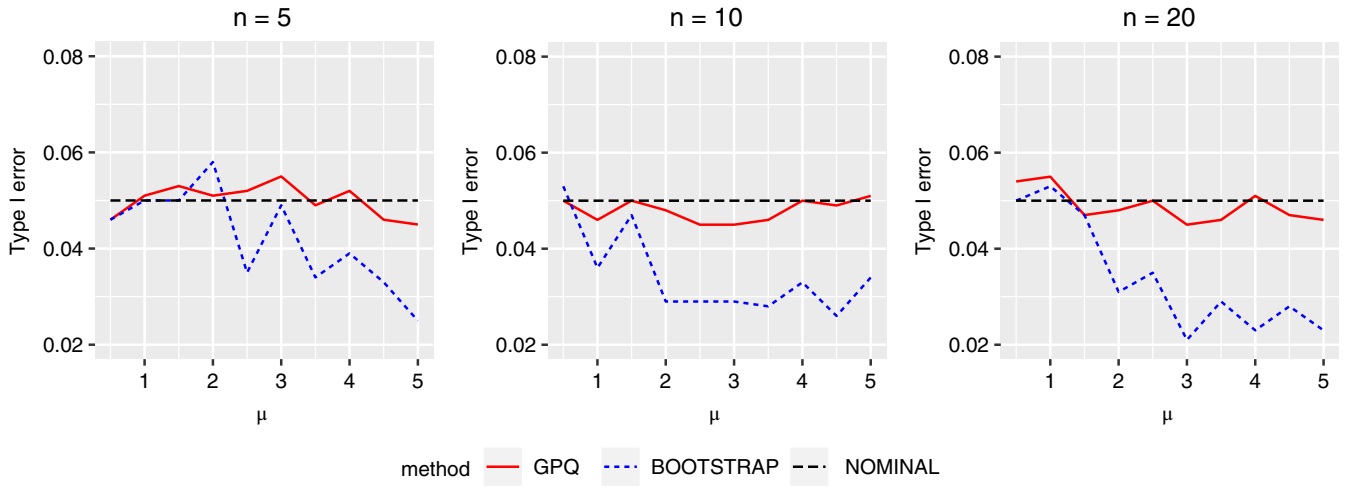

FIGURE 7 Type I error for model discrimination between the inverse Gaussian process and the gamma process based on the proposed generalized pivotal quantities method (GPQ) and the percentile bootstrap (BOOTSTRAP) when $n = 5, 10, 20$ and $\mu = 0.5, 1, \ldots, 4.5, 5$ [Colour figure can be viewed at wileyonlinelibrary.com]

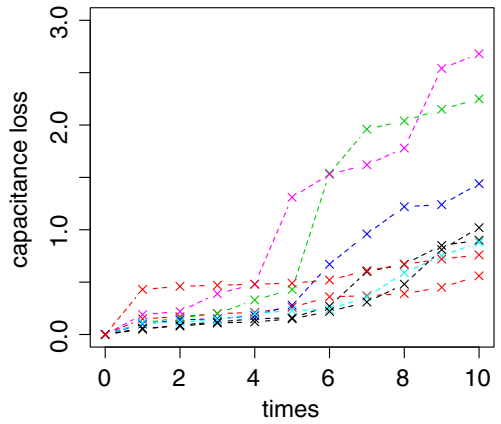

FIGURE 8 Capacitance loss data of eight capacitors [Colour figure can be viewed at wileyonlinelibrary.com]

4.2 | Illustrative example

In this subsection, degradation data of capacitors are used for illustration. Capacitors are used as energy storage units in many systems, and their performance is characterized by the capacitance. Because the capacitors are often designed to be highly reliable, degradation tests are often used in the design phase. Figure 8 shows the capacitance losses of eight capacitors in a degradation test (Sun, Tang, Feng, & Jin, 2013). The capacitance of each capacitor is measured every 1000 shots, where each shot represents a charge/discharge cycle. Because the capacitance decreases with time, the gamma process and the IG process are more appropriate than the Wiener process. The proposed Algorithm 2 with $B = 10{,}000$ is applied to discriminate between the two processes. In addition, the bootstrap and the Bayesian methods are also considered. The number of replications and posterior samples is set as 10,000 for the bootstrap method and the Bayesian method, respectively. The P-values from all these three methods are shown in Table 3. Obviously, the IG process is more appropriate than the gamma process for this degradation dataset.

TABLE 3 P-values for discrimination between the gamma and the IG processes based on the GPQ method, the bootstrap method and the Bayesian method

H_0	GPQ	Bootstrap	Bayesian
Gamma	6×10^{-4}	2×10^{-4}	2×10^{-4}
IG	0.315	0.334	0.315

5 | CONCLUDING REMARKS

This study has proposed a general model discrimination framework based on generalized pivots for reliability data. Our extensive simulations showed that the proposed method performs better than the large sample approximation, the bootstrap and the Bayesian method in controlling the Type I errors, especially when the sample sizes are limited. In terms of the power, the proposed method is uniformly more powerful than the large sample approximation, while it exhibits a similar performance as the bootstrap and the Bayesian method. There are a number of issues that are not thoroughly addressed here and could be topics of future research. First, it is interesting to investigate model selection when the number of candidates is more than two. When the number of candidates is not too large, the proposed model discrimination procedure may be applied to each pair of the models. However, such a treatment is not efficient when the number of candidates is large. A simple remedy is to first select two models from all the candidates by some criteria such as the Akaike information criterion and the Bayesian information criterion, and then apply the proposed model discrimination procedure. However, a more rigorous analysis is definitely needed. Second, it is possible that neither model is rejected at a certain significance level based on our proposed procedure. For example, if we randomly select $n = 30$ of drill lifetimes from the dataset in Section 3.2, then it is common neither distribution is rejected at the significance level $\alpha = 0.05$. In such cases, more data are needed for further discrimination between the two models (Lawless, 2003, chap. 1). Hence, one future direction could be determination of the minimum

sample size required for successful model discrimination at a certain significance level. Finally, the proposed discrimination procedure is based on a complete dataset. In practice, it is possible that the reliability data are censored, for example, the lifetime data from a life test. Based on Type-II censored lifetime data, GPQs for distributions in the log-location-scale distribution family can be readily obtained, and the procedure is similar to that in Appendix A. Nevertheless, GPQs for the gamma distribution and the IG distribution still remains unexplored. Therefore, more effort is needed to extend our proposed method to censored data.

ACKNOWLEDGMENTS

We are grateful to the editor, the associate editor, and two referees for their insightful comments that have lead to a substantial improvement to an earlier version of the paper. This work was supported by the National Science Foundation of Jiangsu Province under Grant BK20180232, and Singapore MOE AcRF Tier 2 under Grant R-266-000-125-112.

ORCID

Piao Chen https://orcid.org/0000-0001-9714-450X

REFERENCES

Bain, L. J., & Engelhardt, M. (1980). Probability of correct selection of weibull versus gamma based on likelihood ratio. *Communications in Statistics – Theory and Methods*, *9*(4), 375–381.

Chen, P., & Ye, Z.-S. (2017). Approximate statistical limits for a gamma distribution. *Journal of Quality Technology*, *49*(1), 64–77.

Chen, P., & Ye, Z.-S. (2019). Uncertainty quantification for monotone stochastic degradation models. *Journal of Quality Technology*, *50*(2), 207–219.

Chen, P., Ye, Z.-S., & Zhai, Q. (in press). Parametric analysis of time-censored aggregate lifetime data. *IISE Transactions*. https://www.tandfonline.com/doi/abs/10.1080/24725854.2019.1628374

Dumonceaux, R., Antle, C. E., & Haas, G. (1973). Likelihood ratio test for discrimination between two models with unknown location and scale parameters. *Technometrics*, *15*(1), 19–27.

Givens, G. H., & Hoeting, J. A. (2012). Computational statistics. New Jersey: John Wiley & Sons.

Hannig, J., Iyer, H., Lai, R. C., & Lee, T. C. (2016). Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, *111*(515), 1346–1361.

Hannig, J., Iyer, H., & Patterson, P. (2006). Fiducial generalized confidence intervals. *Journal of the American Statistical Association*, *101*(473), 254–269.

Huang, A., & Wand, M. P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, *8*(2), 439–452.

Kahle, W., & Lehmann, A. (2010). *The wiener process as a degradation model: Modeling and parameter estimation*. In Advances in degradation modeling (pp. 127–146). Springer.

Krishnamoorthy, K., & Tian, L. (2008). Inferences on the difference and ratio of the means of two inverse gaussian distributions. *Journal of Statistical Planning and Inference*, *138*(7), 2082–2089.

Kundu, D., & Manglick, A. (2004). Discriminating between the weibull and log-normal distributions. *Naval Research Logistics*, *51*(6), 893–905.

Lawless, J., & Crowder, M. (2004). Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime Data Analysis*, *10*(3), 213–227.

Lawless, J. F. (2003). Statistical models and methods for lifetime data. New Jersey: John Wiley & Sons.

Meeker, W. Q., & Escobar, L. A. (1998). Statistical methods for reliability data. New Jersey: John Wiley & Sons.

Rhodes, K. M., Turner, R. M., & Higgins, J. P. (2015). Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *Journal of Clinical Epidemiology*, *68*(1), 52–60.

Schennach, S. M., & Wilhelm, D. (2017). A simple parametric model selection test. *Journal of the American Statistical Association*, *112*, 1663–1674.

Shao, J., & Tu, D. (2012). The jackknife and bootstrap. New York: Springer Science & Business Media.

Su, C., & Wang, X. (2016). A two-stage preventive maintenance optimization model incorporating two-dimensional extended warranty. *Reliability Engineering & System Safety*, *155*, 169–178.

Sun, Q., Tang, Y., Feng, J., & Jin, T. (2013). Reliability assessment of metallized film capacitors using reduced degradation test sample. *Quality and Reliability Engineering International*, *29*(2), 259–265.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, *22*(4), 1701–1728.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, *57*(2), 307–333.

Wang, B. X., & Wu, F. (2017). Inference on the gamma distribution. *Technometrics*, *60*(2), 235–244.

Wang, X., Li, L., & Xie, M. (in press). An unpunctual preventive maintenance policy under two-dimensional warranty. *European Journal of Operational Research*. https://www.sciencedirect.com/science/article/abs/pii/S0377221719307817

Wang, X., & Xu, D. (2010). An inverse gaussian process model for degradation data. *Technometrics*, *52*(2), 188–197.

Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, *88*(423), 899–905.

Ye, Z.-S., Xie, M., Tang, L.-C., & Chen, N. (2014). Semiparametric estimation of gamma processes for deteriorating products. *Technometrics*, *56*(4), 504–513.

Yu, P. L., Mathew, T., & Zhu, Y. (2017). A generalized pivotal quantity approach to portfolio selection. *Journal of Applied Statistics*, *44*(8), 1402–1420.

Zhai, Q., & Ye, Z. S. (2017). Rul prediction of deteriorating products using an adaptive wiener process model. *IEEE Transactions on Industrial Informatics*, *13*(6), 2911–2921.

Zhao, X., He, K., Kuo, W., & Xie, M. (in press). Planning accelerated reliability tests for mission-oriented systems subject to degradation and shocks. *IISE Transactions*, *52*(1), 91–103.

Zhao, X., & Xie, M. (2017). Using accelerated life tests data to predict warranty cost under imperfect repair. *Computers & Industrial Engineering*, *107*, 223–234.

How to cite this article: Chen P, Ye Z-S, Xiao X. Pairwise model discrimination with applications in lifetime distributions and degradation processes. *Naval Research Logistics* 2019;66:675–686. https://doi.org/10.1002/nav.21875

APPENDIX A: GPQS FOR THE LOG-LOCATION-SCALE PARAMETRIC DISTRIBUTIONS

Based on the structural method for GPQ construction (Hannig, Iyer, & Patterson, 2006), GPQs for the two parameters require two pivotal quantities. For a distribution in the log-location-scale family, it can be shown that $U_a = (\widehat{a}-a)/\widehat{b}$ and $U_b = \widehat{b}/b$ are two pivotal quantities, where $\widehat{a}$ and $\widehat{b}$ are the ML estimators (Lawless, 2003, Appendix E). Therefore, based on the observed data, a GPQ for a is

$$\mathcal{G}_a = \widehat{a} - \widehat{b}U_a,$$

and a GPQ for b is

$$\mathcal{G}_b = \widehat{b}/U_b.$$

The distributions of $\mathcal{G}_a$ and $\mathcal{G}_b$ depend on the distributions of U_a and U_b, respectively. Because U_a and U_b are two pivotal quantities, their distributions are independent of the parameters and they can be estimated by the Monte Carlo simulation. For example, assume that $X \sim WB(k,\ \lambda)$. By letting $b = 1/k$ and $a = \log(\lambda)$, it can be shown that the Weibull distribution belongs to the log-location-scale family. Therefore, $U_a = \widehat{k}\log(\widehat{\lambda}/\lambda)$ and $U_b = \widehat{k}/k$ for the Weibull distribution. Let $\widehat{k}^*$ and $\widehat{\lambda}^*$ be the ML estimators based on a sample of n observations from $WB(1, 1)$. Based on the property of pivotal quantities, U_b has the same distribution with $\widehat{k}^*$ and U_a has the same distribution with $\widehat{k}^*\log(\widehat{\lambda}^*)$. Therefore, the distributions of U_b and U_a can be estimated by a sufficient number of realizations of $\widehat{k}^*$ and $\widehat{\lambda}^*$, and they can be used to estimate the distributions of $\mathcal{G}_b$ and $\mathcal{G}_a$ thereafter.

APPENDIX B: GPQS FOR THE GAMMA DISTRIBUTION

GPQs for the gamma distribution were developed by Wang and Wu (2017). A GPQ $\mathcal{G}_\vartheta$ for ϑ is the solution to

$$\log(\widetilde{X}/\overline{X}) = c_1(\mathcal{G}_\vartheta) + [c_2(\mathcal{G}_\vartheta)]^{1/2}Z(\mathcal{G}_\vartheta, U_\vartheta),$$

where c_i is the ith cumulant of $\log(\widetilde{X}/\overline{X})$, U_ϑ follows the standard uniform distribution and

$$\begin{aligned} Z(\vartheta,p) =& z_p + \frac{1}{6}\widetilde{c}_3(z_p^2-1) + \frac{1}{24}\widetilde{c}_4(z_p^3-3z_p) \\ &- \frac{1}{36}(\widetilde{c}_3)^2(2z_p^3-5z_p) + \frac{1}{120}\widetilde{c}_5(z_p^4-6z_p^2+3) \\ &- \frac{1}{24}\widetilde{c}_3\widetilde{c}_4(z_p^4-5z_p^2+2) + \frac{1}{324}(\widetilde{c}_3)^3 \\ &\times(12z_p^4-53z_p^2+17), \end{aligned}$$

where $\widetilde{c}_i = c_i/(c_2)^{i/2}$ and z_p is the pth quantile of a standard normal distribution. On the other hand, a GPQ for β can then be constructed as

$$\mathcal{G}_\beta = \frac{U_\beta}{2n\bar{x}},$$

where $U_\beta \mid U_\vartheta \sim \chi^2(2n\mathcal{G}_\vartheta)$.

APPENDIX C: GPQS FOR THE WIENER PROCESS

Let $K = \sum_i\sum_j t_{i,j}$ and $M = \sum_i m_i$. Observe that $U_\mu = (\widehat{\mu}-\mu)/\sqrt{S^2/K} \sim \chi^2_{M-1}$ and $U_\sigma = (M-1)S^2/\sigma^2 \sim t_{M-1}$, where χ^2_{M-1} denotes a χ^2 distribution with degrees of freedom $M-1$, t_{M-1} denotes a t distribution with degrees of freedom $M-1$, and S^2 is an unbiased estimator of σ^2 (eg, Kahle & Lehmann, 2010). In addition, U_μ and U_σ are independent pivotal quantities. By making use of the above relations, we can express μ and σ^2 in terms of U_μ and U_σ to obtain the pivotal quantities for μ and σ^2 as follows:

$$\mathcal{G}_\mu = \widehat{\mu} - \sqrt{S^2/K}U_\mu, \qquad \mathcal{G}_{\sigma^2} = (M-1)S^2/U_\sigma,$$

where U_μ and U_σ are treated as independent χ^2_{M-1} and t_{M-1} random variables, while the other quantities, for example, $\widehat{\mu}, S^2$, are treated as fixed numbers computed from the observed data.

APPENDIX D: GPQS FOR THE IG PROCESS

According to Chen and Ye (2019), a GPQ for the shape parameter λ can be $\mathcal{G}_\lambda = \widehat{\lambda}U_\lambda/M$, where $M = \sum_i m_i$ and U_λ is a $\chi^2(M-1)$ random variable. On the other hand, an approximate GPQ for μ is $\mathcal{G}_\mu = \widehat{\mu}\left|\frac{\sqrt{M\widehat{\mu}/\widehat{\lambda}}}{\sqrt{|M-1|K}}U_\mu + 1\right|^{-1}$, where $K = \sum_i\sum_j t_{i,j}$ and U_μ is a Student-t random variable with $M-1$ degrees of freedom.

APPENDIX E: TECHNICAL PROOFS

Proof of Theorem 1 Suppose X is from a log-location-scale distribution with CDF $F(x) = F_0((x/a)^b)$, where $a>0$ and $b>0$ are the unknown parameters, and $F_0(\cdot)$ is a completely known distribution function (Lawless, 2003, section 1.3.6). According to the definition of location-scale distributions, $Y = \log(X)$ follows a location-scale distribution with location parameter $\log(a)$ and scale parameter $1/b$. Based on the results from Dumonceaux, Antle, and Haas (1973), we note that $\frac{\max\prod_{i=1}^n g_1(Y_i;\log(a),1/b)}{\max\prod_{i=1}^n g_2(Y_i;\log(a),1/b)}$ is independent of the unknown parameters a and b, where $g(\cdot)$ is the density function for Y. Thus, it suffices to show that

$$\mathrm{RML} = \frac{\max\prod_{i=1}^n f_1(X_i;a,b)}{\max\prod_{i=1}^n f_2(X_i;a,b)} = \frac{\max\prod_{i=1}^n g_1(Y_i;\log(a),1/b)}{\max\prod_{i=1}^n g_2(Y_i;\log(a),1/b)},$$

which can be easily verified by noting that $g(y;\log(a),1/b) = xf(x;a,b)$. ∎

Proof of Theorem 2 Recall that T_c is the α percentile of $\{\mathcal{T}^{(1)}, \ldots, \mathcal{T}^{(B)}\}$ and let $\mathcal{T}_\alpha$ be the α-quantile of the distribution of the test statistic $\mathcal{T}$. Based on Theorem 1, we have $\mathcal{T}$ is independent of the unknown parameters when both f_1 and f_2 are both log-location-scale distributions. Therefore, it is easy to show that $\lim_{B\to\infty} P(\mathcal{T} \leq \mathcal{T}_c) = P(\mathcal{T} \leq \mathcal{T}_\alpha) = \alpha$. This completes the proof. ■