Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

Predicting a Future Lifetime through Box-Cox Transformation

ZHENLIN YANG
Department of Statistics and Applied Probability, c/o Department of Economics
National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260

Abstract. In predicting a future lifetime based on a sample of past lifetimes, the Box-Cox transformation method provides a simple and unified procedure that is shown in this article to meet or often outperform the corresponding frequentist solution in terms of coverage probability and average length of prediction intervals. Kullback-Leibler information and second-order asymptotic expansion are used to justify the Box-Cox procedure. Extensive Monte Carlo simulations are also performed to evaluate the small sample behavior of the procedure. Certain popular lifetime distributions, such as Weibull, inverse Gaussian and Birnbaum-Saunders are served as illustrative examples. One important advantage of the Box-Cox procedure lies in its easy extension to linear model predictions where the exact frequentist solutions are often not available.

Keywords: Box-Cox transformation; coverage probability; Kullback-Leibler information; lifetime distributions; prediction interval.

1. Introduction

Predictions for lifetime distributions have played, over the years, an important role in the reliability, life testing, and quality control. These include predicting the safe life of a certain electronic component, predicting the fatigue life of a metal specimen in an accelerated life testing, and providing warranty limits for future performance of a specified system. There are many types of prediction problems that concern a future observation, a future sample, future order statistics, etc. The simplest and most important one is the prediction of a single future observation based on an independent past sample. There are several methods for studying the prediction problems, namely, frequentist, likelihood type, Bayesian and Fiducial methods, etc. The one that is most commonly used in practice is the frequentist approach. Common lifetime distributions include exponential, gamma, Weibull, lognormal, inverse Gaussian and Birnbaum-Saunders. Hahn and Nelson (1972), Patel (1989) and Hahn and Meeker (1991) give reviews on predictive inferences for all the common

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

distributions, except the Birnbaum-Saunders that was considered by Desmond and Yang (1995), among others.

In this article, we revisit the problem of predicting a future lifetime in a general way, using the frequentist approach incorporated with the Box-Cox transformation. There are many reasons why this topic is worthy for further study: (a) the exact frequentist prediction intervals are not available for certain distributions such as gamma and Birnbaum-Saunders unless a certain parameter is known; (b) the exact intervals may not be satisfactory such as the prediction intervals for the inverse Gaussian that can be very wide and are two-sided only; (c) it is desirable to have a unified approach that covers all the lifetime distributions to eliminate the problem of model selection one faces in practice, etc. The Box-Cox power transformation (Box and Cox, 1964) aims to transform a nonnegative random variable to normal, and it is a one-to-one transformation. Hence it provides a natural approach to predict a nonnegative future lifetime through inverse transformation. Hahn and Meeker (1991, p72) mentioned this approach for univariate distributions and Collins (1991) discussed it in the framework of a linear model. However, neither theoretical nor empirical considerations have been given for the properties of the prediction interval obtained through a Box-Cox transformation. Specifically, two issues need to be addressed: (i) lifetime observations are not transformable to exact normal using power transformation unless they come from a lognormal population, hence the amount of discrepancy from normality and its effect on the performance of the resulted prediction interval needs to be assessed; (ii) the transformation parameter is usually estimated from the same set of data, so the effect of this estimation on the prediction interval should also be investigated.

Section 2 introduces the Box-Cox method and presents some theoretical results concerning the two issues introduced above. Section 3 gives Monte Carlo results, to evaluate the small sample behavior of the Box-Cox predictive pivot when data come from various lifetime distributions and to compare the Box-Cox prediction interval with the corresponding frequentist one in terms of coverage probability and average length of the intervals. Section 3 also illustrates and compares the prediction intervals using some real life data. Finally, in Section 4, we give a discussion of the Box-Cox prediction interval in the linear model

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

framework. To facilitate comparisons, we now summarize the prediction intervals for common lifetime distributions.

Essentially, any distribution with the positive half real line as its domain can be regarded as a lifetime distribution. We concentrate on common two-parameter lifetime distributions and treat both parameters unknown. Obviously, this is the case that is of the greatest practical interest. Generically, we will use $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$ to represent a sample of past observations from a certain lifetime distribution having probability density function (pdf) $g(\cdot|\alpha,\beta)$ with parameters α and β, and Y^0 a single future observation from the same population. The mean and standard deviation of the past sample are denoted as usual by $\bar{Y}$ and s.

The gamma distribution is denoted by GA(α, β), where β is the scale parameter. When α = 1, the gamma becomes an **exponential** distribution where an exact prediction interval can be found from Hahn and Nelson (1972). This interval can be modified to give a prediction interval for gamma distribution for the case of α known. Though the gamma distribution has been extensively studied, the simple frequentist prediction interval for the case of both parameters unknown does not seem to have been discovered. Johnson, *et al.* (1995) gave a summary for the gamma distribution. **The Weibull distribution** is denoted by WB(α, β), where β is the scale parameter. Engelhardt and Bain (1979) presented a pivotal quantity from which one-sided or two-sided prediction intervals can be derived. However, the percentage points of this pivotal quantity has to be either simulated (Fertig, *et al.*, 1980; Mee and Kushary, 1994) or approximated (Engelhardt and Bain, 1982). We will follow the approximation (2) of Engelhardt and Bain (1982) for simulations and calculations. **The lognormal distribution** is denoted by LN(α, β), where α and β are the mean and standard deviation of log(Y). A prediction interval for $\log(Y^0)$ can be constructed using the formula for normal distribution, which is inverted to give an interval for Y^0. **The inverse Gaussian distribution** is denoted by IG(α, β), where α is the mean parameter. Chhikara and Guttman (1982) gave an exact prediction interval for Y^0. This interval is two-sided only and the upper limit sometimes has to be set to infinity. Padgett (1982) proposed an approximate prediction interval, and Padgett and Tsoi (1986) showed via simulation that the approximate

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

interval is superior to the exact one when n is larger than 15. The importance of the inverse Gaussian distribution in lifetime inferences is reflected by the number of articles in the literature and the two specialized monographs by Chhikara and Folks (1989) and Seshadri (1993). **The Birnbaum-Saunders distribution** is denoted by BS(α, β), where β is the scale parameter. Desmond and Yang (1995) gave an approximate prediction interval for Y^0.

2. Box-Cox Prediction Interval and Its Theoretical Properties

We have just summarized the problem of predicting Y^0 based on $(Y_1, Y_2, \ldots, Y_n)$ when g is one of the common lifetime distributions. It is seen that simple frequentist prediction intervals do not exist for gamma and Weibull distributions. For the inverse Gaussian distribution, it is available but can be very wide (there are cases where the upper limit has to be set to infinity), which is virtually of no practical value when warranty limits are desired. Also in practice, it is often difficult to decide which lifetime distribution is the true one (or suitable one) for the data. A goodness of fit test can help, but sometimes it fails to discriminate the two closely related distribution. Hence it is desirable to have a unified method that works for any life distribution and yet gives a reasonable approximation. Box and Cox (1964) introduced a family of power transformation for a nonnegative random variable Y

$$Y(\lambda) = \begin{cases} \dfrac{Y^{\lambda}-1}{\lambda}, & \lambda \neq 0, \\ \log Y, & \lambda = 0, \end{cases} \tag{2.1}$$

in a desire to improve the validity of normal theory inferences, including prediction. Let $Y_i(\lambda)$, $i = 1, 2, \ldots, n$, and $Y^0(\lambda)$ be transformed past and future observations. Box and Cox assumed that there exists a λ such that $Y_i(\lambda)$ is $N(\mu,\sigma)$ for some μ and σ. Under this assumption, the log-likelihood function of (μ, λ, σ), given $\mathbf{Y} = \mathbf{y} = (y_1, y_2, \ldots, y_n)$, is

$$\ell(\mu,\lambda,\sigma|\,\mathbf{y}) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i(\lambda)-\mu]^2 + (\lambda-1)\sum_{i=1}^{n}\log(y_i). \tag{2.2}$$

Maximizing (2.2) results the so called **Box-Cox estimates** $(\hat{\mu}, \hat{\lambda}, \hat{\sigma})$ of (μ, λ, σ). Clearly, $\hat{\mu}$, $\hat{\lambda}$ and $\hat{\sigma}$ are not the maximum likelihood estimates unless $Y_i(\lambda)$ is exactly $N(\mu,\sigma)$. Box-Cox then recommended using $\hat{\lambda}$ for the transformation and treat the so-transformed

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

observations $Y_i(\hat{\lambda})$, $i = 1, \ldots, n$, and $Y^0(\hat{\lambda})$ as normal, hence all the normal theory inferences can be applied. This method is usually termed as the Box-Cox transformation method and its validity has been investigated by many authors such as, Hinkley and Runger (1984), Cox and Reid (1987), Duan (1993), Yang (1996, 1998), Hooper and Yang (1997), etc. In predictive inferences, this method suggests that the following predictive pivot

$$T_{\mathrm{BC}}(\hat{\lambda}) = \frac{Y^0(\hat{\lambda}) - \bar{Y}(\hat{\lambda})}{s(\hat{\lambda})\sqrt{1+n^{-1}}} \tag{2.3}$$

should be referred to a t_{n-1} distribution, where $\bar{Y}(\hat{\lambda})$ and $s(\hat{\lambda})$ are the sample mean and sample standard deviation in $\hat{\lambda}$ scale. A prediction interval for $Y^0(\hat{\lambda})$ follows immediately from (2.3), and after a simple inverse transformation we have a prediction interval for Y^0:

$$\left\{\left[1+\hat{\lambda}\left(\bar{Y}(\hat{\lambda}) \pm t_{n-1}(\delta/2)s(\hat{\lambda})\sqrt{1+n^{-1}}\right)\right]^{1/\hat{\lambda}}\right\} \tag{2.4}$$

When $\hat{\lambda} = 0$, the interval becomes $\exp\left\{\bar{Y}(\hat{\lambda}) \pm t_{n-1}(\delta/2)s(\hat{\lambda})\sqrt{1+n^{-1}}\right\}$.

Clearly, (2.4) works for any continuous distribution with range being the positive half real line, but not necessarily a lifetime distribution. The interval (2.4) will be referred to in this article as the **Box-Cox prediction interval**. Hahn and Meeker (1991, p72) mentioned this interval and Collins (1991) discussed it in the framework of a linear model. However, no considerations have been given for the properties of the Box-Cox prediction interval. Specifically, (i) the effect of non-normality of the transformed observations needs to be assessed, and (ii) the effect of estimating λ needs to be quantified. The first issue is critical for a successful application of the Box-Cox method for prediction; if a distribution can not even be transformed to anywhere near normal then it is almost surely that the interval (2.4) will behave poorly. If a given distribution can be transformed closely to normal for some λ, then $T_{\mathrm{BC}}(\lambda)$ will behave very closely to t_{n-1} and the resulted interval will also behave well. In this case, what left is to investigate the effect of replacing λ by the Box-Cox estimate $\hat{\lambda}$.

2.1. Transforming a Known Distribution

Let $g(\cdot)$ be the known pdf of Y_i, $f(\cdot|\lambda)$ be the true pdf of $Y_i(\lambda)$ and $\varphi(\cdot|\mu,\sigma)$ be a normal pdf with mean μ and standard deviation σ. We are looking for a λ such that $f(\cdot|\lambda)$

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

is "closest" to $\varphi(\cdot|\mu,\sigma)$ for some μ and σ in a certain sense. We choose the *Kullback-Leibler information number* as the measure of discrepancy between $f(\cdot|\lambda)$ and $\varphi(\cdot|\mu,\sigma)$, i.e.,

$$I(\lambda,\mu,\sigma) = \int f(t|\lambda)\log\left\{\frac{f(t|\lambda)}{\varphi(t|\mu,\sigma)}\right\}dt, \tag{2.5}$$

Minimizing $I(\lambda,\mu,\sigma)$ over (λ, μ, σ) results a pdf, $f(\cdot|\lambda)$, that is closest to a normal pdf. Hernandze and Johnson recommended (1981) to first find the values of μ and σ that best approximate $\varphi(\cdot|\mu,\sigma)$ by $f(\cdot|\lambda)$ and then to search for the value of λ that minimizes the remaining distance. They showed that under mild conditions the values of μ and σ that minimize $I(\lambda,\mu,\sigma)$ are $\mu(\lambda)=E_g[Y_1(\lambda)]$ and $\sigma^2(\lambda)=\mathrm{VAR}_g[Y_1(\lambda)]$, which gives

$$I_{\min}(\lambda) = \tfrac{1}{2}[\log(2\pi)+1] + E_g\{\log[g(Y_1)]\} + (1-\lambda)E_g[\log(Y_1)] + \tfrac{1}{2}\log\{\mathrm{VAR}_g[Y_1(\lambda)]\}. \tag{2.6}$$

They indicated that the value of λ that minimizes (2.6) is independent of the scale parameter. Instead of (2.6), one can also work with the equation below if $I_{\min}(\lambda)$ has a unique minimum and if the differentiation and integration are interchangeable,

$$\frac{dI_{\min}(\lambda)}{d\lambda} = \frac{\mathrm{COV}_g\left[Y_1(\lambda),\dot{Y}_1(\lambda)\right]}{\mathrm{VAR}_g[Y_1(\lambda)]} - E_g\left[\log(Y_1)\right] = 0, \tag{2.7}$$

where $\dot{Y}_1(\lambda)$ is the derivative of $Y_1(\lambda)$. When $\lambda = 0$, the derivative simplifies to

$$\left.\frac{dI_{\min}(\lambda)}{d\lambda}\right|_{\lambda=0} = \frac{E_g\left\{\log(Y_1)-E_g\left[\log(Y_1)\right]\right\}^3}{E_g\left\{\log(Y_1)-E_g\left[\log(Y_1)\right]\right\}^2}.$$

Hence if $E_g\left\{\log(Y_1)-E_g\left[\log(Y_1)\right]\right\}^3 = 0$, then the log-transformation will be the optimal one. This condition is clearly satisfied by the lognormal distribution. It is also satisfied by the Birnbaum-Saunders distribution since $\log[\tfrac{1}{2}\alpha Z + (\tfrac{1}{4}\alpha^2 Z^2+1)^{1/2}]$ is an odd function of the standard normal random variable Z and $Y_1 = \beta[\tfrac{1}{2}\alpha Z + (\tfrac{1}{4}\alpha^2 Z^2+1)^{1/2}]^2$. Rieck and Nedelman (1991) discussed this property in a general framework of the sinh-normal distribution. Hence for the lognormal and Birnbaum-Saunders distributions, the best transformation to normality is the log-transformation, irrespective the values of α and β. The former has $I_{\min}(0) = 0$, i.e., the log-transformation results exact normality, and the latter gives a non-zero $I_{\min}(0)$,

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

which means that the best-transformed distribution still deviates from normality. A general result concerning the best transformation and the amount of discrepancy from normality seems impossible. However, for the five popular lifetime distributions, it is sufficient to give some calculations that cover various parameter configurations. The powerful numerical integration and minimum-finding features of MATHEMATICA 3.0 make these calculations handy.

Table 2.1 lists out the values of the best transformation λ, the mean $\mu(\lambda)$ and standard deviation $\sigma(\lambda)$ of the random variable transformed according to λ, the Kullback-Leibler information number $I_{\min}(1)$ and $I_{\min}(\lambda)$, and the 10% and 5% tail probabilities of $f(\cdot|\lambda)$. Notice that $I_{\min}(1)$ measures the distance between $g(\cdot)$ and a closest normal, and the difference between $I_{\min}(1)$ and $I_{\min}(\lambda)$ represents the improvement to normality by transformation. The 5% tail probability, for example, means the probability that the standardized transformed variable falls outside of (–1.96, 1.96), the 95% bounds of $N(0, 1)$. For the gamma, Birnbaum-Saunders and Weibull distributions, the parameter β is a scale parameter, hence does not affect the determination of the λ value. For the inverse Gaussian distribution λ depends only on α/β. For the Weibull distribution, the best transformation is $\lambda = 0.2655\times\alpha$, and for all λ, $I_{\min}(\lambda) = 0.0028$. The Birnbaum-Saunders has the largest skewness 3.9355. Thus, the cases considered in Table 2.1 should be extensive enough. Certain cases were also considered by Hernandze and Johnson (1981). Our calculations agree with theirs, showing that the MATHE-MATICA 3.0 performs well in numerical integration and minimization. Notice that in performing numerical integrations, the integration limits for Y are taken to be ε and ε^{-1} to avoid numerical overflow or underflow, where ε is a small positive number.

Table 2.1. Summary of limiting parameter values, KL information numbers and tail probabilities of f for 10% and 5% nominal levels

$g(\cdot\|\alpha,\beta)$	γ	λ	μ	σ	$I_{\min}(1)$	$I_{\min}(\lambda)$	10%	5%

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

IG(1, 5)	1.34	–0.0258	–0.0940	0.4284	0.1374	0.0005	0.1004	0.0485
IG(1, 3)	1.73	–0.0379	–0.1512	0.5405	0.2179	0.0011	0.1005	0.0476
IG(1, 1)	3	–0.0746	–0.3932	0.8611	0.5420	0.0042	0.1002	0.0444
IG(4, 1)	6	–0.1291	0.3492	1.2009	1.3844	0.0112	0.0996	0.0400
BS($\frac{1}{4}$, 1)	0.75	0.0	0.0	0.2481	0.0452	0.0001	0.1002	0.0495
BS(1, 1)	2.52	0.0	0.0	0.9147	0.5023	0.0077	0.0988	0.0411
BS(2, 1)	3.40	0.0	0.0	1.5914	1.1626	0.0382	0.0862	0.0230
GA(3, 1)	1.15	0.3120	1.1491	0.8138	0.1267	0.0002	0.1004	0.0493
GA(1$\frac{1}{2}$,1)	1.63	0.2884	0.1648	0.9205	0.2607	0.0010	0.1009	0.0485
GA(1,1)	2	0.2654	–0.3641	1.0079	0.4189	0.0028	0.1014	0.0474
WB(3, 1)	0.17	0.7963	–0.1214	0.3360	0.0074	0.0028	0.1014	0.0474
WB(2, 1)	0.63	0.5309	–0.1821	0.5039	0.0540	0.0028	0.1014	0.0474
WB(1, 1)	2	0.2655	–0.3640	1.0078	0.4189	0.0028	0.1014	0.0474

Results of this section show that all the common life distributions can be transformed very closely to a normal distribution. The largest discrepancy occurs in the Birnbaum-Saunders distribution with large skewness. The more skewed the population is, the larger the improvement to normality through transformation. Tail probabilities are all very close to the nominal level. The results are certainly encouraging and they set a foundation for a valid application of Box-Cox transformation to prediction.

2.2. The Limiting Behavior of $T_{\rm BC}(\hat{\lambda})$

To assess the effect of estimating transformation, we first look at the limiting behavior of $T_{\rm BC}(\hat{\lambda})$. The following theorem shows that when n is large the discrepancy between $T_{\rm BC}(\hat{\lambda})$ and t_{n-1} parallels the discrepancy between $f(\cdot|\lambda)$ and $\varphi(\cdot|\mu,\sigma)$.

Theorem 2.1. Assume $\hat{\lambda} \xrightarrow{p} \lambda$, and $E[\dot{Y}_i(\lambda)]$ and $E[Y_i(\lambda)\dot{Y}_i(\lambda)]$ exist. Then,

$$T_{\rm BC}(\hat{\lambda}) = \frac{Y^0(\hat{\lambda}) - \bar{Y}(\hat{\lambda})}{s(\hat{\lambda})\sqrt{1+n^{-1}}} \xrightarrow{d} \frac{Y^0(\lambda)-\mu}{\sigma}.$$

Proof: $Y^0(\hat{\lambda})$ is a continuous function of $(Y^0, \hat{\lambda})$ that converges in distribution, hence $Y^0(\hat{\lambda}) \xrightarrow{d} Y^0(\lambda)$ (see Serfling, 1980, p24). Now a first -order Taylor expansion gives

$$\bar{Y}(\hat{\lambda}) = \bar{Y}(\lambda) + (\hat{\lambda}-\lambda)[\tfrac{1}{n}\textstyle\sum_{i=1}^n \dot{Y}_i(\lambda) + R_n].$$

The law of large number ensures that $\frac{1}{n}\sum_{i=1}^n \dot{Y}_i(\lambda)$ converges in probability, hence $R_n \xrightarrow{p} 0$ as $\hat{\lambda} \xrightarrow{p} \lambda$, so is the second term. Also $\bar{Y}(\lambda) \xrightarrow{p} \mu$, which gives $\bar{Y}(\hat{\lambda}) \xrightarrow{p} \mu$. Now

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

$$s^2(\hat{\lambda}) = \tfrac{1}{n-1}\sum_{i=1}^{n}\left[Y_i(\hat{\lambda}) - \bar{Y}(\hat{\lambda})\right]^2 = \tfrac{1}{n-1}\sum_{i=1}^{n}\left[Y_i^2(\hat{\lambda})\right] - \tfrac{n}{n-1}\bar{Y}^2(\hat{\lambda}).$$

The last term above converges in probability to μ^2 and the first term becomes by a first-order Taylor expansion around λ,

$$\tfrac{1}{n-1}\sum_{i=1}^{n}Y_i^2(\hat{\lambda}) = \tfrac{1}{n-1}\sum_{i=1}^{n}Y_i^2(\lambda) + (\hat{\lambda}-\lambda)[\tfrac{2}{n-1}\sum_{i=1}^{n}Y_i(\lambda)\dot{Y}_i(\lambda) + R_n'].$$

The law of large number shows that $\tfrac{2}{n-1}\sum_{i=1}^{n}Y_i(\lambda)\dot{Y}_i(\lambda) \xrightarrow{p} E[Y_i(\lambda)\dot{Y}_i(\lambda)]$ and $\tfrac{1}{n-1}\sum_{i=1}^{n}Y_i^2(\lambda) \xrightarrow{p} \sigma^2 + \mu^2$. Hence $\tfrac{1}{n-1}\sum_{i=1}^{n}Y_i^2(\hat{\lambda}) \xrightarrow{p} \sigma^2 + \mu^2$ and $s^2(\hat{\lambda}) \xrightarrow{p} \sigma^2$. Finally, an application of Slusky's Theorem gives the result.

Hernandze and Johnson (1981) showed that, under certain regularity conditions, $\hat{\lambda}$ converges almost surely to λ, the value that minimizes the Kullback-Leibler information number. Theorem 2.1 tells that $T_{\mathrm{BC}}(\hat{\lambda})$ does not converge to the limit of the reference distribution, the standard normal, unless the original observations are taken from lognormal. However, for the common lifetime distributions, the distribution of $[Y^0(\lambda) - \mu]/\sigma$ is seen to be very close to standard normal, hence it is expected that this effect will be generally small. Besides, other transformation might exist, which may work better for certain individual lifetime distribution than the Box-Cox power transformation.

2.3. The Effect of Estimating Transformation

We continue to explore the higher-order performance of $T_{\mathrm{BC}}(\hat{\lambda})$. In particular, we want to compare $T_{\mathrm{BC}}(\hat{\lambda})$ with $T_{\mathrm{BC}}(\lambda)$, the λ-known predictive pivot, to quantify the pure effect of estimating transformation. The following result shows that this effect is negligible for moderate *n*. Let $\dot{T}_{\mathrm{BC}}(\lambda)$ and $\ddot{T}_{\mathrm{BC}}(\lambda)$ be the first- and second-order derivatives of $T_{\mathrm{BC}}(\hat{\lambda})$ evaluated at λ.

Theorem 2.2. Assume $\hat{\lambda}-\lambda = O_p(n^{-1/2})$, $\dot{T}_{\mathrm{BC}}(\lambda) = E[\dot{T}_{\mathrm{BC}}(\lambda)] + O_p(n^{-1/2})$ and $\ddot{T}_{\mathrm{BC}}(\ell) = O_p(1)$, where ℓ is a point interior of the interval joining λ and $\hat{\lambda}$. Then, as $n \to \infty$, $T_{\mathrm{BC}}(\hat{\lambda})$ differs from $T_{\mathrm{BC}}(\lambda)$ at most on third order, i.e., the error of approximating $T_{\mathrm{BC}}(\hat{\lambda})$ by $T_{\mathrm{BC}}(\lambda)$ is at most $O_p(n^{-1})$.

Proof. Under the assumptions, we have by Taylor's Theorem

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

$$T_{\mathrm{BC}}(\hat{\lambda}) = T_{\mathrm{BC}}(\lambda) + E[\dot{T}_{\mathrm{BC}}(\lambda)](\hat{\lambda}-\lambda) + O_p(n^{-1}).$$

The result follows by showing that $E[\dot{T}_{\mathrm{BC}}(\lambda)]$ is of order $O(n^{-1/2})$ or smaller, which is done by some straightforward calculations and some reasonable approximations.

Theorem 2.2 shows that the first- and second-order effect of estimating λ are zero. Similarly, one can study the third- and even higher-order effects by carrying out expansions to the corresponding order. However, the results of Theorem 2.2 almost guarantees a good agreement between $T_{\mathrm{BC}}(\hat{\lambda})$ and $T_{\mathrm{BC}}(\lambda)$ for a moderate *n*, which is sufficient for the most of the practical purposes, we decide not to proceed further. Instead, we will use Monte Carlo simulations to do further investigations.

3. Monte Carlo Simulations and Real Data Illustrations

In this section, Monte Carlo simulations are employed to simulate the distribution of the Box-Cox predictive pivot $T_{\mathrm{BC}}(\hat{\lambda})$ and to compare the Box-Cox prediction interval with the corresponding frequentist interval, if it exits. All the simulations are performed using F90 on a Cray J916 Supercomputer with system UNICOS 9.0.2.6. The following IMSL subroutines are used: RNGAM for generating gamma random numbers, RNWEI for Weibull random numbers, and RNNOR for standard normal random numbers that are converted to lognormal random numbers by exponentiating, to Birnbaum-Saunders random numbers by their relationship with standard normal random numbers, and to inverse Gaussian random numbers by an algorithm described in Chhikara and Folks (1989, p52). The simulation process can be described simply as follows. In each run, a random sample of size *n*+1 is generated from a life-time distribution. The first *n* random numbers are used to calculate $\hat{\lambda}$, $T_{\mathrm{BC}}(\hat{\lambda})$ or the prediction intervals, and the last one acts as a future observation. For each configuration, 10,000 samples are generated, giving 10,000 values of $T_{\mathrm{BC}}(\hat{\lambda})$ which are used to calculate various summary statistics such as mean, standard deviation, etc. In the case of prediction interval, the 10,000 samples give 10,000 pairs of prediction intervals, upon which average lengths of the intervals and the proportion of the intervals that cover the future observation are calculated.

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

3.1. The Distribution of $T_{\mathrm{BC}}(\hat{\lambda})$

Simulation results concerning the distribution of $T_{\mathrm{BC}}(\hat{\lambda})$ are summarized in Table 3.1. Obviously, skewness (γ) is an important factor with regard to the performance of $T_{\mathrm{BC}}(\hat{\lambda})$, so we have configured parameters to give lightly skewed to heavily skewed populations. Four different sample sizes are considered. The simulated mean, standard deviation, upper and lower 2.5% and 5% values of $T_{\mathrm{BC}}(\hat{\lambda})$ are reported. The lognormal distribution is included mainly for simulating the pure effect of estimating transformation. Simulation for other distributions reflects the combined effect of non-normality and transformation estimation.

The results in Table 3.1 show a general excellent agreement between the distributions of $T_{\mathrm{BC}}(\hat{\lambda})$ and t_{n-1}. When populations (other than lognormal) are very skewed, $T_{\mathrm{BC}}(\hat{\lambda})$ seems to have slightly smaller tail values than t_{n-1}. This agrees with the results in Sections 2.1 and 2.2. When sample size is very small, $T_{\mathrm{BC}}(\hat{\lambda})$ seems slightly skewed to the right in the gamma and Weibull cases. The results for the lognormal distribution indicate that the pure effect of estimating λ is negligible unless n is very small. In any situation, estimating λ does not seem to introduce any serious problem. This agrees with Theorem 2.2. Finally, the 2.5% tail values seem to decrease slightly with the increase of the population skewness.

Table 3.1. Simulation results for the distribution of $T_{\mathrm{BC}}(\hat{\lambda})$

$g(\cdot\vert\alpha,\beta)$	γ	n	Mean	Sd.Dev.	Lower 2.5%	Lower 5%	Upper 5%	Upper 2.5%

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

IG(1, 9)	1	20	–0.0070	1.1404	–2.1659	–1.7693	1.7612	2.1561
		40	0.0191	1.0549	–2.0301	–1.6713	1.7307	2.0892
		80	0.0161	1.0312	–2.0167	–1.6787	1.6940	2.0416
IG(1, 1)	3	10	–0.0096	1.4503	–2.3221	–1.8486	1.8421	2.3598
		20	–0.0029	1.1097	–2.0757	–1.7328	1.7156	2.1068
		40	–0.0019	1.0395	–1.9922	–1.6839	1.6747	1.9878
		80	–0.0272	1.0200	–1.9961	–1.7046	1.6236	1.9508
IG(4, 1)	6	10	–0.0046	1.3455	–2.3311	–1.8394	1.7780	2.2342
		20	0.0019	1.0841	–2.0427	–1.7382	1.6705	1.9599
		40	–0.0160	1.0487	–1.9551	–1.7045	1.6704	1.9561
		80	–0.0067	1.0232	–1.9762	–1.7031	1.6328	1.8882
BS($\frac{1}{4}$, 1)	0.75	20	0.0015	1.1180	–2.1934	–1.7922	1.7417	2.1195
		40	–0.0110	1.0528	–2.1059	–1.7003	1.6783	1.9916
BS(1, 1)	2.52	10	0.0116	1.2651	–2.2186	–1.7813	1.8289	2.2912
		20	0.0041	1.0832	–2.0494	–1.6923	1.7304	2.0634
		40	–0.0051	1.0304	–1.9569	–1.6623	1.6662	1.9595
BS(2, 1)	3.40	10	0.0014	1.2677	–2.0471	–1.6740	1.7228	2.1000
		20	–0.0054	1.0531	–1.9087	–1.6609	1.6431	1.8915
		40	–0.0132	1.0236	–1.8454	–1.6414	1.6341	1.8414
LN(1, $\frac{1}{2}$)	1.75	10	–0.0005	1.5546	–2.5034	–1.8951	1.9048	2.4876
		20	0.0031	1.1294	–2.1769	–1.7550	1.7597	2.1305
		40	0.0162	1.0584	–2.0219	–1.6683	1.7324	2.1217
LN(1, 1)	6.18	10	0.0147	1.5398	–2.4533	–1.8815	1.9289	2.5257
		20	0.0157	1.1341	–2.1420	–1.7214	1.7866	2.1816
		40	0.0037	1.0670	–2.0729	–1.7179	1.7478	2.0731
LN(1,2)	414.4	20	0.0041	1.1194	–2.1085	–1.7372	1.7274	2.1329
		40	–0.0023	1.0480	–2.0477	–1.7079	1.6867	2.0478
GA(4,1)	1	10	0.0124	1.4668	–2.3645	–1.8826	1.8959	2.7059
		20	0.0051	1.1096	–2.1042	–1.7279	1.7736	2.2203
		40	–0.0026	1.0623	–2.0582	–1.7006	1.7078	2.0405
GA(1,4)	2	10	0.0759	1.7057	-2.2059	-1.7590	1.9419	2.5884
		20	0.0212	1.1344	–2.1503	–1.7183	1.7840	2.2355
		40	0.0221	1.0460	–1.9711	–1.6457	1.7574	2.0974
GA($\frac{1}{4}$,1)	4	20	0.0080	1.0711	–1.8246	–1.5736	1.7849	2.0518
		40	–0.0053	1.0191	–1.7803	–1.5517	1.7423	1.9809
WB(3,1)	0.17	20	–0.0010	1.1135	–2.0554	–1.7092	1.7666	2.1379
		40	0.0111	1.0378	–1.9535	–1.6556	1.7301	2.0514
WB(2,1)	0.63	10	0.0545	1.7635	–2.1971	–1.7689	1.9675	2.5790
		20	0.0293	1.1004	–1.9996	–1.6359	1.7610	2.1297
		40	0.0088	1.0639	–2.0054	–1.6722	1.7856	2.0992
WB(1,1)	2	10	0.0785	1.8943	–2.1184	–1.7164	1.9008	2.5155
		20	0.0115	1.1019	–2.0078	–1.6628	1.7620	2.1231
		40	0.0135	1.0559	–1.9304	–1.6448	1.7301	2.0478

Note: for n = 10, 20, 40 and 80, $t_{n-1}(0.05)$ = 1.883, 1.729, 1.685, and 1.6645; $t_{n-1}(0.025)$ = 2.262, 2.093, 2.023, and 1.9905, and sd(t_{n-1}) = 1.1338, 1.0572, 1.0267, and 1.0129.

3.2. Monte Carlo Comparisons of Prediction Intervals

A good agreement between the distributions of $T_{\rm BC}(\hat{\lambda})$ and t_{n-1} ensures a good coverage property of the Box-Cox prediction interval. However, the length of the prediction

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

interval is another important factor that needs to be assessed and compared with that of the frequentist interval. We choose four lifetime distributions for which the exact or approximate frequentist prediction intervals exist. Again, parameters are configured to give different population skewness. The 90%, 95% and 99% intervals are considered. The selected results for the 90% interval are reported in Table 3.2. More extensive results are available from the author.

Table 3.2a presents the results for comparing the Box-Cox prediction interval with the corresponding exact inverse Gaussian prediction interval. The results clearly favor the Box-Cox prediction interval in general in terms of coverage probability and average length. Also, the Box-Cox approach allows for one-sided prediction intervals which are useful in certain practical situations. The discrepancy between two intervals increases as skewness γ increases. When γ is small to moderate, the two intervals are comparable.

Table 3.2a Simulated average lengths (upper entries) and coverage probabilities for the 90% IG and BC prediction intervals when samples are from an IG(α, β)

		n = 15		n = 30		n = 50	
(α, β)	γ	IG	BC	IG	BC	IG	BC
1, 9	1	1.2776	1.1948	1.1985	1.1139	1.1767	1.0939
		0.8946	0.8916	0.8897	0.8942	0.9008	0.8996
1, $4\frac{1}{2}$	2	1.9414	1.6837	1.7807	1.5478	1.7343	1.5133
		0.9063	0.8989	0.8993	0.8982	0.8997	0.8956
1, 1	3	6.9687	3.5863	5.1342	3.0328	4.7291	2.8925
		0.8976	0.8981	0.9008	0.8960	0.8973	0.8950
1, $\frac{9}{16}$	4	29.73911	4.6868	9.0992	3.8040	7.7466	3.5622
		0.8942	0.8931	0.8998	0.9052	0.8997	0.9035
4, 1	6	∞	29.3343	∞	19.8301	77.6350	17.8370
		0.8986	0.9031	0.8951	0.8952	0.9069	0.9012

The simulation results in Table 3.2b suggest an almost equivalent performance of the Box-Cox and Birnbaum-Saunders prediction intervals, especially when populations are not too skewed. The skewness of the Birnbaum-Saunders distribution depends only on the shape parameter α. If $\alpha \to 0$, then $\gamma \to 3.9355$. When population is very skewed, the Box-Cox intervals are longer, but coverage probabilities are higher than the Birnbaum-Saunders intervals. Both approach allow one-sided prediction intervals.

Table 3.2b. Simulated average lengths (upper entries) and coverage probabilities

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

of the 90% BS and BC prediction intervals when samples are from a BS(α, β)

		n = 15		n = 30		n = 50	
(α, β)	γ	BS	BC	BS	BC	BS	BC
$\frac{1}{4}$, 1	0.75	0.9293	0.9305	0.8799	0.8764	0.8631	0.8609
		0.9040	0.8942	0.8972	0.8948	0.9047	0.9013
$\frac{1}{2}$, 1	1.45	2.0142	2.0461	1.8770	1.8811	1.8371	1.8363
		0.9050	0.8980	0.8976	0.8947	0.8953	0.8937
1, 1	2.52	5.0386	5.4953	4.5937	4.7315	4.4556	4.5401
		0.8993	0.8995	0.8977	0.9010	0.8991	0.9022
2, 1	3.40	16.0173	22.1222	14.0239	16.4519	13.5751	15.2794
		0.8937	0.9198	0.8947	0.9148	0.8981	0.9175

Table 3.2c concerns the lognormal distribution where when λ is known $T_{\mathrm{BC}}(\lambda)$ has the exact t_{n-1} distribution. Hence the simulation in this case reflects the pure effect of estimating λ. The results show that this effect is generally very small unless the population is extremely skewed and a high coverage (99%) is desired. In this case simulation may run into difficulty as there is a possibility that $1+\hat{\lambda}[\overline{Y}(\hat{\lambda}) \pm t_{n-1}(\delta/2)s(\hat{\lambda})\sqrt{1+n^{-1}}]$ is negative, especially when n is small. However, this is not a problem of practical concern as seen in the next subsection.

Table 3.2c. Simulated average lengths (upper entries) and coverage probabilities of the 90% LN and BC prediction intervals when samples are from an LN(α, β)

		n = 15		n = 30		n = 50	
(α, β)	γ	LN	BC	LN	BC	LN	BC
1, $\frac{1}{4}$	0.78	2.5209	2.5476	2.4111	2.4061	2.3605	2.3536
		0.8962	0.8889	0.8957	0.8917	0.9039	0.9011
1, $\frac{1}{2}$	1.75	5.6529	5.7564	5.3100	5.3112	5.1650	5.1479
		0.8965	0.8891	0.9016	0.8980	0.8981	0.8945
1, 1	6.18	17.2046	18.4606	15.2314	15.3474	14.4710	14.4859
		0.9035	0.8949	0.8984	0.8955	0.9014	0.8980
1, 1$\frac{1}{2}$	33.47	48.5716	56.3198	38.7353	39.5036	35.7174	35.9523
		0.8996	0.8910	0.9017	0.8982	0.8945	0.8908

The results in Table 3.2d show that the Box-Cox prediction interval is generally comparable with the corresponding Weibull interval: the Weibull interval is slightly shorter, but the Box-Cox interval has a slightly higher coverage. The Box-Cox interval is much easier to implement than the Weibull interval. Beside increasing the length of the interval, changing the skewness does not seem to affect much of the relative performance of the two

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

intervals. In all above situations, increasing sample size reduces the discrepancy between the two intervals.

Table 3.2d. Simulated average lengths (upper entries) and coverage probabilities of the 90% WB and BC prediction intervals when samples are from a WB(α, β)

		n = 20		n = 50		n = 100	
(α, β)	γ	WB	BC	WB	BC	WB	BC
3, 1	.17	1.0746	1.1353	1.0720	1.0916	1.0613	1.0804
		0.8862	0.8977	0.8962	0.8994	0.8923	0.8978
2, 1	.63	1.5037	1.6171	1.5010	1.5442	1.4869	1.5263
		0.8808	0.8952	0.8917	0.8958	0.8894	0.8969
1, 1	2.0	2.9705	3.4149	2.9469	3.1276	2.8956	3.0502
		0.8860	0.8968	0.8904	0.8934	0.8897	0.8948

3.3. Real Data Illustrations

In this subsection, we consider some real data sets to illustrate the Box-Cox prediction interval and to compare it with the existing ones. The four data sets considered have been extensively used for illustrating the applications of the popular lifetime distributions in the context of reliability and life testing.

Vehicle Failure Data. The following data set, analyzed by Bilikan *et al.* (1979) and by Cheng and Iles (1990), are the number of miles to failure of a type of vehicle: 184, 250, 439, 444, 450, 478, 487, 524, 688, 850, 1048, 1280, 1364, 1488, 1513, 1860, 1947, 1991, 2200, 2446.

Ball Bearings Data. The following data set, the endurance of deep groove ball bearings, is probably one of the most frequently used data set in literature for illustrating the applications of lifetime distributions. It can be found from Chhikara and Folks (1989, p73): 17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.48, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40.

Repair Time Data. Repair times (in hours) for an airborne communication transceiver were reported by Von Alven (1964, p156) and subsequently analyzed by Chhikara and Folks (1977), Padgett (1982), Chhikara and Folks (1989, p139) where a '0.5' was missed from the data set, etc., using inverse Gaussian distribution. The data is omitted from this paper.

Fatigue Life Data. The data considered in this example are the fatigue lives of 101 aluminum coupons (Birnbaum and Saunders, 1969, p343). We omit the data as it is long.

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

Six different types of prediction intervals are calculated (the last one corresponds to the sample percentiles) and the results are summarized in Table 3.3, where $\hat{\gamma}$ denotes the sample skewness. Results from these data sets all show a good performance of the Box-Cox prediction interval. It gives the shortest 99% prediction interval for the fatigue life data. In many cases, the Weibull interval is the shortest, but has a slightly lower coverage than nominal levels as seen from the results of earlier subsection. The repair time data is the most skewed data set among the four, where the largest discrepancy among the intervals is observed. The fatigue life data is least skewed and the largest in size, where a very similar performance among the intervals is observed. The inverse Gaussian interval can be much wider than others even when data is not much skewed, as seen from the vehicle failure data.

Table 3.3. Prediction intervals for real life data

P.I.	90%	95%	99%

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

Vehicle Failure Data, $\hat{\lambda} = 0.2727$, $n = 20$, $\hat{\gamma} = 0.4032$						
BC	(183.08,	2826.00)	(116.12,	3443.32)	(35.01,	5059.85)
IG	(274.49,	4926.91)	(215.70,	6655.38)	(137.42,	12559.26)
BS	(222.37,	3172.33)	(175.38,	4022.04)	(112.12,	6291.83)
LN	(220.67,	3330.65)	(165.83,	4432.11)	(90.75,	8098.61)
WB	(162.68,	2471.25)	(97.78,	2816.89)	(28.92,	3531.09)
EP	(187.30,	2433.70)	(184.00,	2246.00)	(184.00,	2246.00)
Ball Bearings Data, $\hat{\lambda} = 0.1905$, $n = 23$, $\hat{\gamma} = 0.9206$						
BC	(23.55,	152.61)	(18.55,	179.26)	(10.85,	247.07)
IG	(27.88,	195.29)	(23.31,	238.30)	(16.46,	357.23)
BS	(24.72,	160.85)	(20.68,	192.28)	(14.60,	272.40)
LN	(24.90,	161.73)	(20.50,	196.46)	(13.66,	294.82)
WB	(17.96,	143.06)	(12.23,	158.20)	(4.90,	188.22)
EP	(28.92,	128.04)	(17.88,	173.40)	(17.88,	173.40)
Repair Time Data, $\hat{\lambda} = -0.1014$, $n = 46$, $\hat{\gamma} = 2.8568$						
BC	(0.3209,	14.3872)	(0.2419,	23.0032)	(0.1336,	63.9771)
IG	(0.4268,	35.5374)	(0.3140,	52.2935)	(0.1837,	114.9859)
BS	(0.3135,	13.0724)	(0.2376,	17.2432)	(0.1471,	27.8582)
LN	(0.2915,	12.8025)	(0.2000,	18.6561)	(0.0934,	39.9391)
WB	(0.1236,	12.1760)	(0.0542,	15.2975)	(0.0079,	22.7251)
EP	(0.5000,	10.3000)	(0.3000,	22.0000)	(0.2000,	24.5000)
Fatigue Life Data, $\hat{\lambda} = 0.5805$, $n = 101$, $\hat{\gamma} = 0.3288$						
BC	(97.94,	172.34)	(91.63,	180.53)	(79.68,	197.19)
IG	(100.53,	178.05)	(95.14,	188.19)	(85.35,	209.96)
BS	(99.15,	175.25)	(93.84,	185.16)	(84.22,	206.32)
LN	(99.24,	169.52)	(93.89,	185.20)	(84.13,	206.68)
WB	(91.40,	169.52)	(81.97,	174.89)	(63.74,	184.63)
EP	(100.00,	166.00)	(96.00,	174.00)	(70.00,	212.00)

4. Prediction based on a Transformed Linear Model

Theoretical and empirical results suggest that even if observations can not be transformed to exact normal, the Box-Cox procedure still provides a reasonable approximation to prediction intervals. This is very useful when the exact prediction intervals are not available or when the available intervals are wide. The importance of Box-Cox procedure also lies in its nonparametric feature. In practice, one rarely knows exactly which distribution the observations came from, hence it is desirable to have a prediction procedure that is robust against misspecification of the parent distribution. An easy generalization to general linear model prediction is another main advantage of the Box-Cox transformation

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

method for prediction. This is based on the popular Box-Cox transformed linear model of the form

$$Y(\lambda) = \beta_0 + X\beta + e \tag{4.1}$$

where X is an $n \times p$ matrix containing the values of p predictors, β is a $p \times 1$ vector of regression coefficients, and e is a vector of independent errors, assumed to have the same normal distribution with mean zero and standard deviation σ. Suppose that the values of predictors are all centered at their averages, and the future value x^0 $(p \times 1)$ of the predictor-vector is also centered at this point. Thus, the prediction interval (2.4) can be easily extended to give a prediction interval for a future observation Y^0 at the predictor value x^0**:**

$$\left\{1 + \hat{\lambda}\left(\bar{Y} + \hat{\beta}x^0 \pm t_{n-p-1}(\delta/2)\hat{\sigma}\sqrt{1 + n^{-1} + (x^0)'(X'X)^{-1}x^0}\right)\right\}^{1/\hat{\lambda}}, \tag{4.2}$$

where $\hat{\beta}$, $\hat{\lambda}$ and $\hat{\sigma}$ are the mle's of β, λ and σ under the assumptions that the errors in (4.1) are normal. The interval (4.2) is easy to compute and accounts for uncertainty in β_0, β and σ but ignores uncertainty about λ (Collins, 1991). Similar to the one-sample situation, the validity of the prediction interval (4.2) also depends on the validity of ignoring uncertainty about λ and the validity of the normality assumption of model (4.1). This can be investigated in a similar way as in the one-sample case. Carroll and Ruppert (1981) studied similar problems of predicting the median or mean of a future observation. Their results indicate that the effect of estimating transformation is small. Our results together with those of Carroll and Ruppert hint a good performance of the interval (4.2); hence, we will not pursue further the properties of (4.2) in this article.

Acknowledgements

The author is grateful to Editor-in-Chief, Dr. Mei-Ling Ting Lee, and the referees for helpful comments that lead to improvements of this article.

References

1. J. E. Bilikan, A. H. Moore and G. L. Petrick, "*K* sample ML ratio test for change of shape parameter," *IEEE Trans. Reliab.*, vol. 28, pp. 47-50, 1979.
2. Z. W. Birnbaum and S. C. Saunders, "Estimation for a family of life distributions with application to fatigue life," *J. Appl. Prob.*, vol. 6, pp. 328-347, 1969.

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

3. G. E. P. Box and D. R. Cox, "An analysis of transformation (with discussion)," *J. R. Statist. Soc.* B, vol. 26, pp. 211-252, 1964.

4. R. J. Carroll and D. Ruppert, "On prediction and the power transformation family," *Biometrika*, vol. 68, pp. 609-615, 1981.

5. R. C. H. Cheng and T. C. Iles, "Embedded models in three parameter distributions and their estimation," *J. R. Statist. Soc.* B, vol. 52, pp. 135-149, 1990.

6. R. S. Chhikara and J. L. Folks, "The inverse Gaussian distribution as a lifetime model," *Technometrics*, vol. 19, pp. 461-468, 1977

7. R. S. Chhikara and J. L. Folks, *The Inverse Gaussian Distribution. Theory, Methodology, and Applications*, Marcel Dekker, Inc.: New York and Basel, 1989.

8. R. S. Chhikara and I. Guttman, "Prediction limits for the inverse Gaussian distribution," *Technometrics*, vol. 24, pp. 319-324, 1982.

9. S. Collins, "Prediction techniques for Box-Cox regression models," *J. Bus. Econ. Stat.*, vol. 9, pp. 267-277, 1991.

10. D. R. Cox and N. Reid, "Parameter orthogonality and approximate conditional inference (with discussion)," *J. Roy. Statist. Soc.* B, vol. 49, pp. 1-39, 1987.

11. A. F. Desmond and Z. Yang, "Shortest prediction intervals for the Birnbaum-Saunders distribution." *Commu. Statist. - Theory Meth.*, vol. 24, pp. 1383-1401, 1995.

12. N. Duan, "Sensitivity analysis for Box-Cox power transformation model: contrast parameters," *Biometrika*, vol. 80, pp. 885-897, 1993.

13. M. Engelhardt and L. J. Bain, "Prediction limits and two-sample problems with complete or censored Weibull data," *Technometrics*, vol. 21, pp. 233-237, 1979.

14. M. Engelhardt and L. J. Bain, "On prediction limits for samples from a Weibull or extreme-value distribution," *Technometrics*, vol. 24, pp. 147-150, 1982.

15. K. W. Fertig, M. E. Meyer and N. R. Mann, "On constructing prediction intervals for samples from a Weibull or extreme value distribution," *Technometrics*, vol. 22, pp. 567-573, 1980;

16. G. J. Hahn and W. Q. Meeker, *Statistical Intervals. A Guide for Practitioners*, John Wiley & Sons, Inc., 1991.

17. G. J. Hahn and W. Nelson, "A survey of prediction intervals and their applications," *J. Qual. Tech.*, vol. 5, pp. 178-188, 1972.

18. F. Hernandze and R. A. Johnson, "The large sample behavior of transformations to normality," *J. Amer. Statist. Assoc.*, vol. 75, pp. 855-861, 1981.

19. D. V. Hinkley and G. Runger, "The analysis of transformed data (with discussion)," *J. Amer. Statist. Assoc.*, vol. 79, pp. 302-320, 1984.

20. P. M. Hooper and Z. Yang, "Confidence intervals following Box-Cox transformation,"

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

Can. J. Statist., vol. 25, pp. 401-416, 1997.

21. N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions,* 2nd ed, John Wiley & Sons, Inc., 1995.
22. R. W. Mee and D. Kushary, "Prediction limits for the Weibull distribution utilizing simulation," Comput. Statist. Data An., vol. 17, pp. 327-336, 1994
23. W. J. Padgett, "An approximate prediction interval for the mean of future observations from the inverse Gaussian distribution," *J. Statist. Comput. Simul.*, vol. 14, pp. 191-199, 1982.
24. W. J. Padgett and S. K. Tsoi, "Prediction intervals for future observations from the inverse Gaussian distribution," *IEEE Trans. Reliab.*, vol. R-35, pp. 406-408, 1986.
25. J. K. Patel, "Prediction intervals - a review," *Commin. Statist. - Theory Meth.*, vol. 18, pp. 2393-2465, 1989.

26 J. R. Rieck and J. R. Nedelman, "A log-linear model for the Birnbaum-Saunders distribution," *Technometrics,* vol. 33, pp. 51-60 ,1991)

27. R. J. Serfling, *Approximation Theorems of Mathematical Statistics* John Wiley & Sons, Inc., 1980.
28. V. Seshadri, *The Inverse Gaussian Distribution. A Case Study in Exponential Families.*
Clearendon Press, Oxford, 1993.
29. W. H. Von Alven (ed.), *Reliability Engineering by ARINC,* Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1964.
30. Z. Yang, "Some asymptotic results on Box-Cox transformation methodology," *Commu. Statist.-Theory Meth.*, vol. 25, pp. 403-414, 1996.
31. Z. Yang, "On robustness of usual confidence region under transformation misspecification," *J. Statist. Comput. Simul.*, vol. 61, pp. 175-190, 1998.

Predicting a Future Lifetime through Box-Cox Transformation

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

Keywords: Box-Cox transformation; coverage probability; Kullback-Leibler information; lifetime distributions; prediction interval.

Address for correspondance:

ZHENLIN YANG
Department of Statistics and Applied Probability,
c/o Department of Economics
National University of Singapore,
10 Kent Ridge Crescent,
Singaproe 119260

Tel. No.: (65) 874-6829 (O)
(65) 569-2303 (H)

Fax No.: (65) 775-2646 (O)
(65) 569-2303 (H)

e-mail: ecsyzl@nus.edu.sg

Vehicle Failure Data. The following data set, analyzed by Bilikan *et al.* (1979) and by Cheng and Iles (1990), are the number of miles to failure of a type of vehicle: 184, 250, 439, 444, 450, 478, 487, 524, 688, 850, 1048, 1280, 1364, 1488, 1513, 1860, 1947, 1991, 2200, 2446.

Ball Bearings Data. The following data set, the endurance of deep groove ball bearings, is probably one of the most frequently used data set in literature for illustrating the applications of lifetime distributions. It can be found from Chhikara and Folks (1989, pp. 73): 17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.48, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40.

Repair Time Data. Repair times (in hours) for an airborne communication transceiver were reported by Von Alven (1964, pp. 156) and subsequently analyzed by Chhikara and Folks (1977), Padgett (1982), Chhikara and Folks (1989, 139) where a '0.5' was missed from the data set, etc., using inverse Gaussian distribution. The 46 observations are: 0.2, 0.3, 0.5, 0.5, 0.5, 0.5, 0.6, 0.6, 0.7, 0.7, 0.7, 0.8, 0.8, 1.0, 1.0, 1.0, 1.0, 1.1, 1.3, 1.5, 1.5, 1.5, 1.5, 2.0, 2.0, 2.2,

Appeared in: . Lifetime Data Analysis, 1999, 5, 265-279.

2.5, 2.7, 3.0, 3.0, 3.3, 3.3, 4.0, 4.0, 4.5, 4.7, 5.0, 5.4, 5.4, 7.0, 7.5, 8.8, 9.0, 10.3, 22.0, 24.5.

Fatigue Life Data. The following data (Birnbaum and Saunders, 1969, pp. 343), are the fatigue lives of 101 aluminum coupons: 70, 90, 96, 97, 99, 100, 103, 104, 104, 105, 107, 108, 108, 108, 109, 109, 112, 112, 113, 114, 114, 114, 116, 119, 120, 120, 120, 121, 121, 123, 124, 124, 124, 124, 124, 128, 128, 129, 129, 130, 130, 130, 131, 131, 131, 131, 131, 132, 132, 132, 133, 134, 134, 134, 134, 134, 136, 136, 137, 138, 138, 138, 139, 139, 141, 141, 142, 142, 142, 142, 142, 142, 144, 144, 145, 146, 148, 148, 149, 151, 151, 152, 155, 156, 157, 157, 157, 157, 158, 159, 162, 163, 163, 164, 166, 166, 168, 170, 174, 196, 212.

All these four data sets have been extensively used in the context of reliability and life testing. Six different types of prediction intervals are calculated (the last one corresponds to the sample percentiles) and the results are summarized in Table 3.3. It is seen from the table that the Box-Cox prediction interval is always one of the shorter ones while the inverse Gaussian is always the longest among six, which can be more than twice as long as the Box-Cox interval.