

Improved Maximum Likelihood Estimation for the Weibull Distribution Under Length-Biased Sampling

David E. Giles

Department of Economics, University of Victoria

Revised

May 2021

Abstract:

We consider the estimation of the parameters of the Weibull distribution when the data arise from “length-biased” sampling. Specifically, the appropriate weighted density is formulated and we analyze the finite-sample properties of the maximum likelihood estimators for its parameters. The analytic Cox-Snell “corrective” approach is used to reduce the biases of these estimators, and we find that this can be done effectively and without detrimental consequences for the mean squared errors. Bootstrap bias-correction is also found to be effective. Simulation results also illustrate the severe consequences of failing to allow for “length-biased” sampling, even in very large samples.

Author contact:

David E. Giles, 58 Rock Lake Court, L'Amable, ON, Canada K0L 2L0; dgiles@uvic.ca

1. Introduction

Many of the inferential procedures used by econometricians can be justified by having desirable properties in large samples. A good example of this is the Maximum Likelihood Estimator (MLE) and its associated tests. However, these same estimators and tests frequently lack optimality when the sample size is small or modest. For example, MLE's frequently exhibit finite-sample bias, inefficiency, and poor coverage probability for confidence intervals.

This “small-sample” problem has generated a vast literature in statistics, including in econometrics. A classic example is the seminal contribution of A. L. Nagar (1959), in which he derived expansions that approximate the first two moments of the k-class estimators for structural equations. In the spirit of honouring Nagar, this paper examines the small-sample properties of the MLE for the parameter vector of the Weibull distribution, when a particular non-standard type of sampling is used. We focus on reducing the order of the bias of this estimator analytically, even though the MLE cannot be expressed in closed form, and on the effect that bias reduction may have on the estimator's precision.

Many economic data are gathered *via* surveys, some of them quite complex in nature. For example, a survey may comprise several levels of stratification and/or clustering. This has important implications for the construction and properties of estimators and tests. This is discussed, for example, by Ullah and Breunig (1998). Even if sufficient information about the nature of the complex sampling is available, it is often largely ignored during estimation. An exception would be applied micro-econometricians' preoccupation with reporting cluster-adjusted standard errors, but even this ignores the much more fundamental issue of the construction of the likelihood function itself, and the adverse effects on the properties of the estimators.

We will consider a more subtle sampling issue that is relevant for a variety of economic applications, namely “size-biased” sampling. This type of sampling occurs when the probability of a unit being selected is proportional to a predetermined weight function, which depends on the observed value of the data for that unit. See the seminal contributions of Wicksell (1925), Fisher (1934), Rao (1965), and Cox (1969). A more formal definition is given in section 3, but if the probability of inclusion of an observation in a sample is directly proportional to the size of that observation, we have so-called “length-biased” sampling. This is important because it affects the formulation of the appropriate data density, and hence the likelihood function. If this is ignored the likelihood is mis-specified and inferences will be invalid, *even asymptotically*. In addition, simulation results indicate the estimator is inconsistent. An appropriate “weighted density” (or “moment density”) is required – for a detailed discussion see, *e.g.*, Patil and Ord (1976).

In particular, length-biased sampling often arises with duration data – see Salant (1977) and Gove (2000). The frequency of observing survival data can in itself generate length-biased sampling, as is noted by Shen *et al.* (2009, p.1192). In this paper we investigate the (two-parameter) Weibull distribution under length-biased sampling. The corresponding “weighted Weibull” distribution has been applied in various fields, including forestry (Gove, 2000), ecology (Gove, 2003), and meteorology (Das and Roy, 2011). The MLE for its parameters is biased in finite samples, but there has been no investigation into the issue of “improving” this MLE by reducing the order of its bias analytically. That is our objective.

The rest of this paper proceeds as follows. In the next section we provide motivation for our study by illustrating how length-biased sampling arises with economic data. Section 3 sets up the statistical framework; and section 4 presents the main theoretical results for Maximum Likelihood estimation and (analytic) bias reduction. Section 5 describes some simulation results that assess the importance of allowing for length-biased sampling, and the gains from bias reduction. Two illustrative applications are presented in section 6; and section 7 contains some concluding remarks

2. Economic motivation

Although the importance of the size-biased sampling has been widely recognized, its relevance in economics is less well documented. Some economic examples should be mentioned by way of motivation for the readers of this paper. These examples arise in environmental and resource economics, labour economics, and more peripherally in marketing and real estate studies.

Length-biased sampling is common in willingness-to-pay studies for non-market goods in resource economics. This is recognized by Nowell *et al.* (1988), for example, in their use of a 1985 Yellowstone National park survey. As they note (p. 369), “All of the interviews were conducted in the Park's camping areas. The sample is length biased due to the fact that a camper who stayed in the campgrounds for a long period of time had a higher probability of being interviewed than a camper who spent only a short amount of time at the campground.” The use of the Weibull distribution in such studies is also widespread – *e.g.*, Alberini *et al.* (2005).

Salant (1977) appears to have been the first to draw attention to the issue of length-biased sampling in labor market surveys, such as those used by the Bureau of Labor Statistics (BLS) for their monthly “duration of unemployment” reports. He observed (p. 41), “... it is spells with longer than average full lengths that are more likely to be in progress at the time of the BLS survey - a phenomenon known as sampling from a “length-biased” population. If, for ex-ample, full spells of X and $2X$ weeks are equally likely to occur, the longer spells will be twice as likely to be in progress at the time of the survey, since the interval in which the longer spells might have commenced is twice as wide.” Importantly, there have been more applications

of survival models (such as the Weibull) to such employment data than to any other type of data in economics.

The use of shopping center (mall) “intercept surveys” is common in marketing. These samples are usually length-biased, for the same reason as outlined in the previous two examples. This has been recognized by Bush and Hair (1985, p.164) and Nowell *et al.* (1991), for example.

In real estate economics, housing surveys of the number of days on the market are prone to length-biased sampling, though this does not appear to have been discussed in the literature. Clearly this is relevant for duration analysis (and hence the Weibull distribution) in the context of such data.

Finally, size-biased sampling can also arise in the context of survival analysis in economics in the modeling of export duration spells. Nicita *et al.* (2013), provide one such study, but it ignores the possibility of length-biased sampling, as do other similar studies.

3. Statistical framework

If a random variable, X , has a density function $f(x; \theta)$, for some parameter (vector) θ , then the corresponding “weighted” (or “moment”) density, to allow for size-biased sampling of order c , is defined as

$$f_c(x; \theta) = f(x; \theta)x^c / m_c' \quad , \quad (1)$$

where $m_r' = E[X^r]$ is the r^{th} (raw) moment of X . This moment is assumed to exist, and clearly (1) is a proper density.

The case $c = 1$ corresponds to “length biased” sampling; $c = 2$ corresponds to “area biased”, *etc.* Motivated by the discussion in section 2, we will be concerned only with length-biased sampling in this paper.

If the random variable, Y , follows a (two-parameter) Weibull distribution, then its density function is:

$$f(y; k, \lambda) = \left(\frac{k}{\lambda}\right) \left(\frac{y}{\lambda}\right)^{k-1} \exp(-(y/\lambda)^k) \quad ; \quad y > 0 \quad (2)$$

where $k (> 0)$ is the shape parameter, and $\lambda (> 0)$ is the scale parameter.¹

For this distribution, $m_r' = \lambda^r \Gamma(1 + r/k)$, so the mean of Y is

$$m_1' = E[Y] = \lambda \Gamma(1 + 1/k) \quad , \quad (3)$$

and the length-biased weighted Weibull distribution has a density function,

$$f_1(x; k, \lambda) = \left(\frac{k}{\lambda}\right) \left(\frac{x}{\lambda}\right)^k \exp(-(x/\lambda)^k) / \Gamma(1 + 1/k) \quad ; \quad x > 0 . \quad (4)$$

Using a simple change of variable, the c.d.f. of X follows immediately as

$$F_1(x; k, \lambda) = \gamma((k + 1)/k, (x/\lambda)^k) / \Gamma(1 + 1/k) , \quad (5)$$

where $\gamma(s, z) = \int_0^z t^{s-1} e^{-t} dt$ is the usual incomplete gamma function. It is worth noting that the associated hazard function can be increasing, decreasing, or uni-modal, depending on the parameter values.

Although this length-biased weighted Weibull distribution has been mentioned previously in the literature (e.g., Gove, 2000, 2003), apparently it has not been recognized that (4) is also the density for the generalized gamma distribution (Stacy, 1962), with Stacy's three² parameters assigned as $a = \lambda$, $p = k$, and $d = k + 1$. The generalized gamma distribution encompasses the gamma, chi-square, exponential, Weibull, and log-normal distributions as special cases, but not all of these arise with our particular *constrained* parameterization. The density in (4) is illustrated in Figure 1, with the corresponding two-parameter Weibull density. Of course, when $k = 1$, the latter distribution is exponential.

The fact that our weighted distribution is generalized gamma in form is helpful for the simulation experiment discussed in sections 5 and 6. Also, it immediately provides the raw moments associated with (4), namely

$$\mu_r' = \lambda^r \Gamma[(k + 1 + r)/k] / \Gamma[(k + 1)/k] ; r = 1, 2, \dots \quad (6)$$

In particular,

$$\mu_k' = E[X^k] = \lambda^k (k + 1)/k . \quad (7)$$

In section 4.1 we will also require the following results³:

$$E_1 = E[X^k \ln(X)] = [\lambda^k \ln(\lambda)(k + 1)/k] + \lambda^k \left[(k + 1) \psi\left(\frac{1}{k}\right) + k(k + 2) \right] / k^2 \quad (8)$$

$$\begin{aligned} E_2 &= E[X^k (\ln(X))^2] \\ &= [\lambda^k / k] \left\{ (k + 1) (\ln(\lambda))^2 + 2 \ln(\lambda) \left[(k + 1) \psi\left(\frac{1}{k}\right) + k(k + 2) \right] / k \right. \\ &\quad \left. + \left[(k + 1) \left(\psi\left(\frac{1}{k}\right) \right)^2 + 2k(k + 2) \psi\left(\frac{1}{k}\right) + (k + 1) \psi'\left(\frac{1}{k}\right) + 2k^2 \right] / k^2 \right\} \end{aligned} \quad (9)$$

Here, for any real z , $\psi(z) = d\ln\Gamma(z)/dz$ is the usual digamma function, and $\psi'(z) = d\psi(z)/dz$ is the trigamma function.

4. Maximum likelihood estimation

4.1 Basic results

Based on a sample of n independent observations⁴ from the distribution in (4), the log-likelihood function is

$$l(\lambda, k) = -n\ln\Gamma\left[1 + \frac{1}{k}\right] + k \sum_{i=1}^n \ln(x_i) - \lambda^{-k} \sum_{i=1}^n x_i^k - n(k+1)\ln(\lambda) + n\ln(k) , \quad (10)$$

and the two likelihood equations are

$$\frac{\partial l}{\partial \lambda} = k\lambda^{-(k+1)} \sum_{i=1}^n x_i^k - \frac{n(k+1)}{\lambda} = 0 \quad (11)$$

$$\frac{\partial l}{\partial k} = \frac{n\psi\left(1 + \frac{1}{k}\right)}{k^2} + \sum_{i=1}^n \ln(x_i) + \lambda^{-k} \ln(\lambda) \sum_{i=1}^n x_i^k - \lambda^{-k} \sum_{i=1}^n x_i^k \ln(x_i) - n(\ln(\lambda) - 1/k) = 0 . \quad (12)$$

The MLE's, $\hat{\lambda}$ and \hat{k} , are obtained by solving the non-linear simultaneous equations, (11) and (12). This particular problem is simplified by noting that we can solve (11) to obtain:

$$\hat{\lambda} = \left[\frac{k \sum_{i=1}^n x_i^k}{n(k+1)} \right]^{1/k} . \quad (13)$$

Substituting this expression into (10) we obtain the profile log-likelihood function for k :

$$l_p(k) = -n\ln\Gamma\left[1 + \frac{1}{k}\right] + k \sum_{i=1}^n \ln(x_i) + n\ln(k) - n(k+1) \left[\ln\left(\frac{k \sum_{i=1}^n x_i^k}{n(k+1)}\right) + 1 \right] / k . \quad (14)$$

It can be shown that (14) is unimodal, and although it is non-linear, this one-parameter function is easily maximized numerically by standard methods to obtain the MLE, \hat{k} . The MLE for λ is obtained by substituting \hat{k} for k in (13). The density in (4) satisfies the usual regularity conditions, so the MLEs are consistent and Best Asymptotically Normal. However, their finite sample properties are not obvious, and this is our primary focus in this paper.

The second-order derivatives of the (full) log-likelihood function are easily obtained as:

$$\frac{\partial^2 l}{\partial \lambda^2} = \frac{n(k+1)}{\lambda^2} - k(k+1)\lambda^{-(k+2)} \sum_{i=1}^n x_i^k \quad (15)$$

$$\frac{\partial^2 l}{\partial \lambda \partial k} = \frac{\partial^2 l}{\partial k \partial \lambda} = \lambda^{-(k+1)} \left[(1 - k\ln(\lambda)) \sum_{i=1}^n x_i^k + k \sum_{i=1}^n x_i^k \ln(x_i) \right] - n/\lambda \quad (16)$$

$$\begin{aligned} \frac{\partial^2 l}{\partial k^2} = & -\frac{2n\psi\left(1+\frac{1}{k}\right)}{k^3} - \frac{n\psi'\left(1+\frac{1}{k}\right)}{k^4} - \lambda^{-k}(\ln(\lambda))^2 \sum_{i=1}^n x_i^k + 2\lambda^{-k}\ln(\lambda) \sum_{i=1}^n x_i^k \ln(x_i) - \\ & \lambda^{-k} \sum_{i=1}^n x_i^k (\ln(x_i))^2 - n/k^2. \end{aligned} \quad (17)$$

Equations (6) – (8) can then be used to derive the following cumulants⁵:

$$\kappa_{11} = E \left[\frac{\partial^2 l}{\partial \lambda^2} \right] = -nk(k+1)/\lambda^2, \quad (18)$$

$$\kappa_{12} = \kappa_{21} = E \left[\frac{\partial^2 l}{\partial \lambda \partial k} \right] = n \{ [(1 - k \ln(\lambda))(k+1)/\lambda + k\lambda^{-(k+1)}E_1] - 1/\lambda \} \quad (19)$$

$$\kappa_{22} = E \left[\frac{\partial^2 l}{\partial k^2} \right] = -n \left\{ \frac{2\psi\left(1+\frac{1}{k}\right)}{k^3} + \frac{\psi'\left(1+\frac{1}{k}\right)}{k^4} + \frac{(k+1)(\ln(\lambda))^2}{k} + \lambda^{-k}(E_2 - 2\ln(\lambda)E_1) + 1/k^2 \right\}, \quad (20)$$

where E_1 and E_2 are defined in (8) and (9). Finally, the expected information matrix is $K = \{-\kappa_{ij}\}$, for $i, j = 1, 2$. It is easily checked that this matrix is positive definite, and it provides the basis for constructing asymptotic standard errors for $\hat{\lambda}$ and \hat{k} , and the usual Wald, LM and likelihood ratio test statistics.

4.2 Bias reduction

Although our two parameter MLE's have optimal large-sample properties, their small-sample quality is uncertain. Assessing these small-sample properties *analytically* is non-trivial, especially as we do not have a closed-form expression for the *estimator*, \hat{k} . We only have a numerical *estimate* for k , obtained by maximizing (14). However, we might anticipate that these MLE's will be biased in small samples. One reason for this conjecture is that the MLE's of k and λ are upward biased (k more than λ) in the regular Weibull case (2). For example, see the simulation results in Table 2 of Chen *et al.*, 2017. Also, Gove (2000) reports *simulated* (positive) biases for the MLE's of the parameters of the area-weighted Weibull distribution. However, there has been no research into *analytic* MLE bias expressions, or the construction of bias-corrected MLE's, in the context of the size-weighted Weibull distribution(s).

Notwithstanding the fact that we do not have an analytic expression for \hat{k} itself, *analytic* expressions for the first-order biases of our two MLE's can still be obtained by using the “corrective” method⁶ suggested by Cox and Snell (1968). In our case, this involves using the second-order cumulants in (17) – (19), and the additional cumulants

$$\kappa_{111} = E \left[\frac{\partial^3 l}{\partial \lambda^3} \right]; \kappa_{222} = E \left[\frac{\partial^3 l}{\partial k^3} \right]; \kappa_{112} = \kappa_{211} = \kappa_{121} = E \left[\frac{\partial^3 l}{\partial \lambda^2 \partial k} \right]; \kappa_{221} = \kappa_{212} = \kappa_{122} = E \left[\frac{\partial^3 l}{\partial \lambda \partial k^2} \right],$$

together with the following cumulant derivatives:

$$\kappa_{11}^{(1)} = \left(\frac{\partial \kappa_{11}}{\partial \lambda} \right); \kappa_{11}^{(2)} = \left(\frac{\partial \kappa_{11}}{\partial k} \right); \kappa_{22}^{(1)} = \left(\frac{\partial \kappa_{22}}{\partial \lambda} \right); \kappa_{22}^{(2)} = \left(\frac{\partial \kappa_{22}}{\partial k} \right); \kappa_{12}^{(1)} = \kappa_{21}^{(1)} = \left(\frac{\partial \kappa_{12}}{\partial \lambda} \right);$$

$$\kappa_{12}^{(2)} = \kappa_{21}^{(2)} = \left(\frac{\partial \kappa_{12}}{\partial k} \right).$$

Manipulating the original Cox-Snell results, Cordeiro and Klein (1994) show that the biases of the MLE's of the p elements of a general parameter vector, θ can be written in a more convenient form than stated originally, namely:

$$Bias(\hat{\theta}_s) = \sum_{i=1}^p \kappa^{si} \sum_{j=1}^p \sum_{l=1}^p [\kappa_{ij}^{(l)} - 0.5 \kappa_{ijl}] \kappa^{jl} + O(n^{-2}); \quad s = 1, 2, \dots, p \quad (21)$$

where κ^{ij} is the $(i, j)^{th}$ element of the *inverse* of the (expected) information matrix, K . The cumulants and their derivatives are assumed to be $O(n)$, which holds in our case⁷.

Define: $a_{ij}^{(l)} = \kappa_{ij}^{(l)} - (\kappa_{ijl}/2)$, for $i, j, l = 1, 2, \dots, p$

$$A^{(l)} = \{a_{ij}^{(l)}\}; \quad i, j, l = 1, 2, \dots, p$$

$$A = [A^{(1)} | A^{(2)} | \dots | A^{(p)}]. \quad (22)$$

Cordeiro and Klein (1994) show that the bias of $\hat{\theta}$ can be re-written as:

$$Bias(\hat{\theta}) = K^{-1} A vec(K^{-1}) + O(n^{-2}), \quad (23)$$

and a “bias-corrected” MLE for θ is defined as:

$$\tilde{\theta} = \hat{\theta} - \hat{K}^{-1} \hat{A} vec(\hat{K}^{-1}), \quad (24)$$

where $\hat{K} = (K)|_{\hat{\theta}}$ and $\hat{A} = (A)|_{\hat{\theta}}$. It can be shown that the bias of $\tilde{\theta}$ is $O(n^{-2})$. Clearly, (22) and (23) can be evaluated *even when the likelihood equation does not admit a closed-form analytic solution*, and the MLE has to be obtained by numerical methods.

We can now apply these results to our problem, where $p = 2$, and $\theta = (\lambda, k)'$. As the second-order cumulants in (18) and (19) involve the expectations, E_1 and E_2 , in (8) and (9), it is clear that the expressions for their derivatives and for the third-order cumulants are exceptionally tedious. However, there is no gain in reporting them here, for two reasons. First, (at least in principle) they can be obtained using the generic Maple code (Maplesoft, 2020) supplied by Stošić and Cordeiro (2009). Of even more practical interest, they can be implemented without prior derivation by using the R package, “mle.tools” (Mazucheli, 2017; Mazucheli *et al.*, 2017). *In short, practitioners do not need the details of these*. However, for the record, they have been derived and are available as an online supplement to this paper.⁸

5. A simulation study

We have undertaken a limited simulation experiment to investigate the biases and mean squared errors (MSE's) of the MLE's and bias-corrected MLE's of k and λ . All of the computations were undertaken with the R software (R Core Team, 2020).⁹ The “ggamma” package (Saldanham and Suzuki, 2019) was used to generate the random variates; the “optimize” routine in base R was used to maximize the profile log-likelihood function; and the “mle.tools” package was used to obtain the Cox-Snell bias corrections. We found that the results are invariant to the true values of the scale parameter, so we assigned $\lambda = 1$.

In addition to the MLE and Cox-Snell bias-adjusted MLE, we also considered the (parametric) bootstrap-bias-corrected estimator¹⁰. In the case of λ , this is obtained as $\check{\lambda} = 2\hat{\lambda} - (1/N_B)[\sum_{j=1}^{N_B} \hat{\lambda}_{(j)}]$, where $\hat{\lambda}_{(j)}$ is the MLE of λ obtained from the j^{th} of the N_B bootstrap samples. A corresponding expression applies for the estimator of the shape parameter, k . See Efron (1982, p.33).

Bootstrap methods have been used by a number of other authors in the context of various inference problems associated with length-biased sampling. For example, see Bergeron *et al.* (2008), Ojeda *et al.* (2008), and Zamini (2017).

Each part of the experiment relating to the MLE's and Cox-Snell biased-corrected MLE's uses 50,000 Monte Carlo replications. This was found to be sufficient to obtain stable results. For the bootstrap-corrected estimators, the same number of Monte Carlo replications was used, with 1,000 bootstrap samples *per* replication. We did not encounter any instances of non-convergence of the maximization algorithm. The results in Table 1 are *percentage* biases and MSE's. The latter are defined as $100 \times (\text{MSE} / \lambda^2)$, *etc.*, and appear in square brackets below the corresponding percentage biases.

5.2 Results

Several patterns emerge in the main simulation results in Table 1. First, the MLE's for both parameters are positively biased, more so (in percentage terms) for the shape parameter than for the scale parameter. The % bias of \hat{k} decreases, while that for $\hat{\lambda}$ increases, as the true value of k increases. Second, the % MSE's for both MLE's decrease as k increases – especially in the case of $\hat{\lambda}$. In addition, for $k > 1$, $\%MSE(\hat{k}) > \%MSE(\hat{\lambda})$ for a fixed sample size, n . Third, the analytic bias correction reduces the (absolute) % bias of both estimators, often by up to two orders of magnitude. In general, “over correction” occurs, and the % biases become negative¹¹. A positive implication of these results is that the *direction* of the biases of the estimators in small samples is known. Fourth, $\%MSE(\check{k}) \leq \%MSE(\hat{k})$, always. In contrast, the analytic

bias correction generally leads to a small *increase* in % MSE in the case of the scale parameter. Of course, the % biases and % MSE's for both estimators of both parameters become negligible as n grows, reflecting the consistency of the MLE's.

The bootstrap bias adjustment results in larger (absolute) % biases than those for the analytically adjusted MLEs, for both parameters, in very small samples. However, the situation reverses for moderate to large sample sizes. In addition, the bootstrap correction results in % MSE's that are very comparable to those associated with the analytic bias correction.

We have also performed a second simulation study to investigate the consequences of wrongly fitting a standard Weibull distribution to length-biased sample data. The set-up for this study is as described above, but the results for only a very limited range of parameters and sample sizes are reported, given their obvious nature. The focus here is purely on mis-specification, so no bootstrap results are included in Table 2.

In that table, \check{k} and, $\check{\lambda}$ are the mis-specified MLE's, and $\bar{\bar{k}}$ and $\bar{\bar{\lambda}}$ are their Cox-Snell bias-adjusted counterparts. It is immediately clear from those results that failing to allow for length-bias when applying the Weibull model has drastic consequences for estimator bias and MSE. This is not resolved by applying the Cox-Snell bias adjustment and, in fact, this adjustment makes matters even worse when estimating the shape parameter (k). The situation *does not improve* as the sample size is increased, reflecting the *inconsistency* of the MLE's of the parameters in this context.

6. Empirical applications

We present two economics-related empirical applications of the length-biased Weibull distribution. One uses survey data relating to a natural resource, so it illustrates the first motivating example in section 2. The other uses stock price data and is chosen to illustrate a different situation in which length-biased sampling can occur. This is one where the frequency of observation is lower than the frequency at which the data are generated (Shen *et al.*, 2009), as mentioned briefly in section 1 above.

6.1 Algonquin Park fishing survey

Algonquin Park is the oldest provincial park in Ontario, Canada. Covering an area of 7,630 km², it has 1,500 lakes and 1,200 kilometres of streams. Ontario Ministry of Natural Resources *et al.* (2010) conducted a survey¹² of anglers who registered for permits when visiting the park in 2010. Angling parties intending to fish for trout were invited to complete the survey during their visit, and 409 of those 1,200 surveyed responded. Information gathered included party numbers and sizes, fishing hours, lakes and streams fished, and trout catches. We have analyzed the duration data for rod-hours per angler, for the "Top 40"

waterbodies in the survey¹³. One might anticipate length bias in the sampling – greater rod-hours may reflect more devotion to the sport, and more concern for information to assist in conservation.

The MLE results for the two-parameter Weibull and the length-biased (unadjusted and bias-adjusted) weighted Weibull models appear in Table 3. Because of the relatively small sample size, bootstrap standard errors¹⁴ are reported in addition to those based on the (asymptotic) information matrix. The three fitted models are compared with the survey data in Figure 2(a). Length-biased weighting reduces the parameter point estimates somewhat, but has less impact on their significance. Kolmogorov-Smirnov (K-S) goodness-of-fit test results are also reported, with the “jitter” function in R used to deal with ties in the data. The K-S results support an allowance for length-biased sampling when using the Weibull distribution here. The analytic bias adjustment reduces the point estimates for k and λ by 4.3% and 0.8% respectively, and increases the mean life estimate by 1.3%. However, it does not have a significant effect on the conclusions in this case.

6.2 Tesla share price

During the period 31 December 2018 to 30 October 2020 the share price for Tesla (TSLA) rose by over 600%, from US\$63.54 to US\$388.04. In that period there were 23 “spells” when week-over-week prices rose for one or more successive weeks¹⁵. Of course, within such weeks, the share price typically fell and rose several times. Length-biased sampling arises here if only weekly price observations are available. We have divided the 23 spell duration values by 4 to make them non-integer, so their units are very roughly in months. This is important, as the estimation of Weibull models from discrete data leads to its own biases (*e.g.*, Røed and Zhang, 2000).

The MLE estimates for the length-biased (unadjusted and bias-adjusted) weighted Weibull models are compared with those for the two-parameter Weibull in Table 3. Both (asymptotic) analytic and bootstrap standard errors are again reported. We see that the MLE’s for the parameters when length-bias is ignored are an order of magnitude larger than when it is taken into account, and there is a substantial reduction in the standard error for the scale parameter estimate in the latter case. Bias correction leads to reductions of 8.8% and 14.9% in the point estimates of k and λ and an increase of 4.8% in the mean life estimate under length-biased weighting. The K-S results overwhelmingly reject the standard Weibull model, while supporting the weighted Weibull specification. This is totally expected from Figure 2(b). Finally, in both applications, failure to account for the length-biased sampling leads to an overstatement in the mean life estimates and this serves as an important warning to practitioners.

7. Further discussion

Our results illustrate the importance of allowing properly for length-biased sampling when analyzing economic duration data. Our focus was on the Weibull distribution, but our broad conclusions can be expected to hold for alternative models. The analytic Cox-Snell methodology is very effective in dramatically reducing the small-sample bias of the MLE for the weighted Weibull distributions parameters, with little or no negative impact on relative MSE. The bootstrap bias correction is also very effective, but with additional computational cost. The Cox-Snell correction implicitly assumes that the bias function is approximately *flat*.¹⁶ This assumption can be relaxed to one of approximate *linearity*, as is discussed by Godwin and Giles (2019). It would be interesting to generalize our results in this direction, but that is left for future research.

Table 1. Percentage biases and MSE's

n	$\%Bias(\hat{k})$ [%MSE(\hat{k})]	$\%Bias(\tilde{k})$ [%MSE(\tilde{k})]	$\%Bias(\check{k})$ [%MSE(\check{k})]	$\%Bias(\hat{\lambda})$ [%MSE($\hat{\lambda}$)]	$\%Bias(\tilde{\lambda})$ [%MSE($\tilde{\lambda}$)]	$\%Bias(\check{\lambda})$ [%MSE($\check{\lambda}$)]
$k = 1.0 ; \lambda = 1.0$						
25	8.260 [6.450]	0.040 [5.082]	-0.317 [5.124]	6.775 [10.907]	0.102 [11.210]	0.845 [11.370]
50	3.880 [2.599]	-0.101 [2.299]	-0.204 [2.330]	3.358 [5.361]	-0.120 [5.448]	0.070 [5.521]
100	1.853 [1.175]	-0.108 [1.106]	-0.037 [1.091]	1.634 [2.658]	-0.141 [2.682]	0.065 [2.671]
250	0.730 [0.445]	-0.048 [0.434]	-0.085 [0.434]	0.657 [1.074]	-0.061 [1.078]	0.108 [1.065]
500	0.364 [0.217]	-0.024 [0.215]	-0.001 [0.214]	0.331 [0.534]	-0.030 [0.535]	0.017 [0.532]
1000	0.176 [0.107]	-0.018 [0.107]	0.010 [0.106]	0.148 [0.265]	-0.033 [0.266]	0.050 [0.266]
$k = 2.0 ; \lambda = 1.0$						
25	7.281 [5.187]	0.031 [4.138]	-0.254 [4.172]	0.980 [1.581]	-0.056 [1.648]	0.155 [1.657]
50	3.424 [2.113]	-0.091 [1.883]	-0.175 [1.908]	0.475 [0.793]	-0.058 [0.810]	0.004 [0.823]
100	1.636 [0.961]	-0.097 [0.907]	-0.030 [0.896]	0.224 [0.398]	-0.046 [0.402]	0.022 [0.402]
250	0.645 [0.366]	-0.043 [0.357]	0.074 [0.356]	0.089 [0.162]	-0.020 [0.163]	0.026 [0.161]
500	0.321 [0.179]	-0.021 [0.177]	-0.001 [0.176]	0.046 [0.081]	-0.009 [0.081]	-0.002 [0.081]
1000	0.155 [0.088]	-0.016 [0.088]	0.009 [0.087]	0.015 [0.040]	-0.012 [0.040]	0.005 [0.040]
$k = 4.0 ; \lambda = 1.0$						
25	6.644 [4.407]	0.026 [3.535]	-0.196 [3.598]	0.045 [0.311]	-0.028 [0.317]	0.041 [0.318]
50	3.127 [1.806]	-0.084 [1.614]	-0.158 [1.635]	0.016 [0.155]	-0.021 [0.157]	0.004 [0.159]
100	1.494 [0.823]	-0.089 [0.778]	-0.026 [0.769]	0.005 [0.078]	-0.014 [0.078]	0.009 [0.078]
250	0.589 [0.314]	-0.039 [0.308]	0.067 [0.306]	0.002 [0.032]	-0.006 [0.032]	0.005 [0.031]
500	0.293 [0.153]	-0.020 [0.152]	-0.001 [0.151]	0.001 [0.016]	-0.002 [0.016]	0.001 [0.016]
1000	0.141 [0.076]	-0.015 [0.076]	0.008 [0.075]	-0.003 [0.008]	-0.005 [0.008]	0.002 [0.008]

Table 2. Percentage biases and MSE's

n	$\%Bias(\hat{k})$ [%MSE(\hat{k})]	$\%Bias(\tilde{k})$ [%MSE(\tilde{k})]	$\%Bias(\ddot{k})$ [%MSE(\ddot{k})]	$\%Bias(\bar{k})$ [%MSE(\bar{k})]	$\%Bias(\hat{\lambda})$ [%MSE($\hat{\lambda}$)]	$\%Bias(\tilde{\lambda})$ [%MSE($\tilde{\lambda}$)]	$\%Bias(\ddot{\lambda})$ [%MSE($\ddot{\lambda}$)]	$\%Bias(\bar{\lambda})$ [%MSE($\bar{\lambda}$)]
$k = 1.0 ; \lambda = 1.0$								
25	8.260 [6.450]	0.040 [5.082]	57.910 [40.230]	109.709 [129.300]	6.775 [10.907]	0.102 [11.21]	121.956 [158.748]	58.229 [40.631]
50	3.880 [2.599]	-0.101 [2.299]	52.892 [30.906]	115.865 [138.982]	3.358 [5.361]	-0.120 [5.448]	121.990 [153.823]	53.053 [31.083]
100	1.853 [1.175]	-0.108 [1.106]	50.555 [26.926]	118.882 [143.779]	1.634 [2.658]	-0.141 [2.682]	121.944 [151.221]	50.635 [27.008]
250	0.730 [0.445]	-0.048 [0.434]	49.178 [24.712]	120.731 [146.749]	0.657 [1.074]	-0.061 [1.078]	121.955 [149.734]	49.210 [24.743]
500	0.364 [0.217]	-0.024 [0.215]	48.729 [24.007]	121.342 [147.736]	0.331 [0.534]	-0.030 [0.535]	121.954 [149.228]	48.745 [24.022]
1000	0.176 [0.107]	-0.018 [0.107]	48.498 [23.651]	121.637 [148.206]	0.148 [0.265]	-0.033 [0.266]	121.943 [148.953]	48.506 [23.659]
$k = 3.0 ; \lambda = 1.0$								
25	6.870 [4.680]	0.028 [3.748]	25.688 [10.904]	-64.733 [41.948]	0.248 [0.594]	-0.038 [0.610]	11.980 [1.878]	275.340 [796.638]
50	3.233 [1.914]	-0.087 [1.709]	21.913 [6.686]	-63.656 [40.544]	0.115 [0.297]	-0.032 [0.301]	12.125 [1.692]	264.918 [718.731]
100	1.544 [0.872]	-0.092 [0.824]	20.169 [4.947]	-63.121 [39.855]	0.052 [0.149]	-0.023 [0.150]	12.185 [1.596]	260.105 [684.450]
250	0.609 [0.332]	-0.040 [0.325]	19.159 [4.009]	-62.796 [39.439]	0.020 [0.061]	-0.010 [0.061]	12.230 [1.540]	257.318 [665.167]
500	0.303 [0.162]	-0.020 [0.161]	18.834 [3.715]	-62.689 [39.301]	0.011 [0.030]	-0.004 [0.030]	12.244 [1.521]	256.424 [659.042]
1000	0.146 [0.080]	-0.016 [0.080]	18.666 [3.567]	-62.635 [39.233]	0.000 [0.015]	-0.007 [0.015]	12.249 [1.511]	255.958 [655.894]

Table 3. Application results*

MLE results						
	(a)	(b)	(c)	(a)	(b)	(c)
	Fishing survey ($n = 40$)			Tesla share price ($n = 23$)		
k	2.357	1.863	1.782	1.206	0.800	0.730
	(0.311)	(0.042)	(0.041)	(0.227)	(0.207)	(0.196)
	[0.272]	[0.008]	[0.008]	[0.176]	[0.026]	[0.022]
λ	9.738	7.396	7.334	2.715	0.215	0.183
	(0.686)	(0.802)	(0.842)	(0.491)	(0.104)	(0.103)
	[0.693]	[0.507]	[0.521]	[0.500]	[0.876]	[0.961]
Mean life	8.630	8.608	8.721	2.551	0.625	0.655
K-S	0.350	0.150	0.150	0.609	0.217	0.217
	{0.014}	{0.766}	{0.766}	{0.000}	{0.660}	{0.660}

* (a) = Weibull; (b) = length-biased weighted Weibull; (c) = bias-adjusted length-based weighted Weibull.

Bootstrap standard errors appear in parentheses. Analytic *asymptotic* standard errors appear in brackets. Two-sided p-values appear in braces.

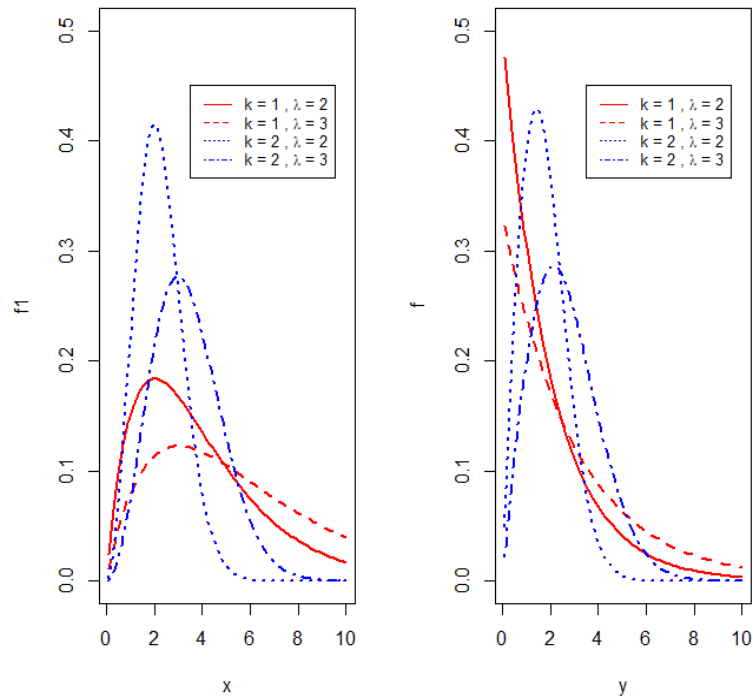


Figure 1: (a) Weighted Weibull (b) Weibull

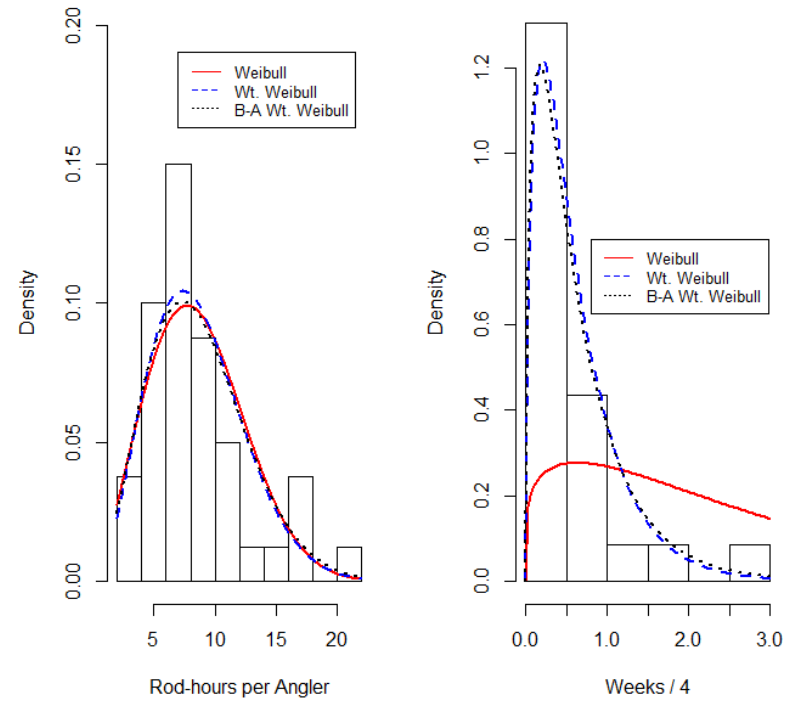


Figure 2*: (a) Fishing survey (b) Tesla share price

* “Wt.” = “weighted”; “B-A” = “Bias-adjusted”

References

- Alberini, A., P. Rosato, A. Longo, and V. Zanatta, 2005. Information and willingness to pay in a contingent valuation study: The value of S. Erasmo in the Lagoon of Venice. *Journal of Environmental Economics and Management*, 48, 155-175.
- Bergeron, P.-J., M. Asgharian and D. B. Wolfson, 2008. Covariate bias induced by length-biased sampling of failure times. *Journal of the American Statistical Association*, 103, 737-742.
- Bush, A. J. and J. F. Hair, 1985. An assessment of the mall intercept as a data collection method. *Journal of Marketing Research*, 22, 158-67.
- Cordeiro, G. M. and F. Cribari-Neto, 2012. *An Introduction to Bartlett Corrections and Bias Reduction*. Springer, New York.
- Cordeiro, G. M. and R. Klein, 1994. Bias correction in ARMA models. *Statistics and Probability Letters*, 19, 169–176.
- Cordeiro, G. M., E. C. Da Rocha, J. G. C. Da Rocha and F. Cribari-Neto, 1996. Bias-corrected maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*, 58, 21-35.
- Cordeiro, G. M. and P. McCullagh, 1991. Bias correction in generalized linear models. *Journal of the Royal Statistical Society, B*, 53, 629-643.
- Cox, D. R., 1969. Some sampling problems in technology. In N. L. Johnson and H. Smith (eds.), *New Developments in Survey Sampling*, Wiley, New York.
- Cox, D. R. and E. J. Snell, 1968. A general definition of residuals (with discussion). *Journal of the Royal Statistical Society, B*, 30, 248–275.
- Cribari-Neto, F. and K. L. P. Vasconcellos, 2002. Nearly unbiased maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*, 72, 107-118.
- Das, K. K. and T. D. Roy, 2011. On some length-biased weighted Weibull distributions. *Advances in Applied Science Research*, 2, 465-475
- Efron, B., 1982. *The Jackknife, the Bootstrap, and Other Resampling Plans*. Society of Industrial Mathematics, Philadelphia, PA.
- Firth, 1993. Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27-38.
- Fisher, R. A., 1934. The effect of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics*, 6, 13–25.

- Giles, D. E., 2012. Bias reduction for the maximum likelihood estimator of the parameters in the half-logistic distribution. *Communications in Statistics – Theory & Methods*, 41, 212-222.
- Giles, D. E., H. Feng and R. T. Godwin, 2013, “On the bias of the maximum likelihood estimator for the two-parameter Lomax distribution. *Communications in Statistics – Theory and Methods*, 42, 1934-1950.
- Giles, D. E., H. Feng and R. T. Godwin, 2016. Bias-corrected maximum likelihood estimation of the parameters of the generalized Pareto distribution. *Communications in Statistics – Theory and Methods*, 45, 2465-2483.
- Godwin, R. T. and D. E. Giles, 2019. Improved analytic bias correction for maximum likelihood estimators. *Communications in Statistics – Simulation and Computation*, 48, 15-26.
- Gove, J. H., 2000. Some observations on fitting assumed diameter distributions to horizontal point sampling data. *Canadian Journal of Forest Research*, 30, 521-33.
- Gove, J. H., 2003. Moment and maximum likelihood estimators for Weibull distributions under length- and area-biased sampling. *Environmental and Ecological Statistics*, 10, 455-467.
- Ontario Ministry of Natural Resources, Algonquin Fisheries Assessment Unit and Ontario Parks, 2010. *The 2010 Algonquin Park Trout Fishing Survey*. www.algonquinpark.on.ca/pdf/fish_survey_2010_final.pdf .
- Maplesoft (2020). Maple Mathematical Software. Maplesoft, Waterloo, ON.
- Mazucheli, J., 2017. mle.tools: Expected/observed Fisher information and bias-corrected maximum likelihood estimate(s). R package version 1.0.0. <https://CRAN.R-project.org/package=mle.tools> .
- Mazucheli, J., A. F. B. Menezes, and S. Nadarajah, 2017. mle.tools: An R package for maximum likelihood bias correction. *The R Journal*, 9, 268-290.
- Nagar, A. L., 1959. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica*, 27, 573-595.
- Nicita, A., M. Shirotori, and B. T. Klok, 2013. Survival analysis of the exports of least developed countries: The role of comparative advantage. Policy Issues in International Trade and Commodities Study Series No. 54, UNCTAD, Geneva.
- Noufaily, A. and M. C. Jones, 2012. On maximization of the likelihood for the generalized gamma distribution. *Computational Statistics*, 28, 505-517.
- Nowell, C., M. A. Evans, and L. McDonald, 1988. Length-biased sampling in contingent valuation studies. *Land Economics*, 64, 367-371.

Nowell, C. and L. R. Stanley, 1991. Length-biased sampling in mall intercept surveys. *Journal of Marketing Research* 28, 475-479.

Ojeda, J. L., J. A. Cristóbal and J. T. Alcalá, 2008. A bootstrap approach to model checking for linear models under length-biased data. *Annals of the Institute of Statistical Mathematics*, 60, 519–543.

Patil, G. P. and J. K. Ord, 1976. On size-biased sampling and related form-invariant weighted distributions. *Sankhyā*, B, 38, 48-61.

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Rao, C. R., 1965. On discrete distributions arising out of methods of ascertainment. In G. P. Patil (ed.), *Classical and Contagious Discrete Distributions*. Statistical Publishing Society, Calcutta, 320-332. Also published in *Sankhyā*, A, 1965, 27, 311-324.

Røed, K. and T. Zhang, 2000. A note on the Weibull distribution and time aggregation bias. Memorandum No. 23/2000, Department of Economics, University of Oslo.

Salant, S. W., 1977. Search theory and duration data: A theory of sorts. *Quarterly Journal of Economics*, 91, 39-57.

Saldanham, J. H. and A. K. Suzuki, 2019. ggamma: Generalized gamma probability distribution. R package version 1.0.1. <https://CRAN.R-project.org/package=ggamma>

Shen, Y., J. Ning, and J. Qin, 2009. Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *Journal of the American Statistical Association*, 104, 1192–1202.

Stacy, E. W., 1962. A generalization of the gamma distribution. *Annals of Mathematical Statistics*, 33, 1187-1192.

Stošić, B. D. and G. M. Cordeiro, 2009. Using Maple and Mathematica to derive bias corrections for two parameter distributions. *Journal of Statistical Computation and Simulation*, 79, 751-767.

Ullah, A. and R. V. Breunig, 1998. Econometric analysis in complex surveys. In A. Ullah and D. E. A. Giles (eds.), *Handbook of Applied Economic Statistics*, Marcel Dekker, New York, 325-363.

Wicksell, S. D., 1925. The corpuscle problem: A mathematical study of a biometric problem. *Biometrika*, 17, 84–99.

Xiao, L. and D. E. Giles, 2014. Bias reduction for the maximum likelihood estimator of the generalized Rayleigh family of distributions. *Communications in Statistics – Theory and Methods*, 43, 1778-1792.

Zamini, R, 2017. A bootstrap approximation to L_p -statistic of kernel density estimator in length-biased model. Paper presented at the 48th Annual Iranian Mathematics Conference, Hamedan, Iran.

Appendix 1

Derivations of equations (8) and (9):

$$E_1 = E[X^k \ln(X)] = \int_0^\infty \frac{x^k \ln(x) \left(\frac{x}{\lambda}\right)^k \left(\frac{k}{\lambda}\right) \exp(-(x/\lambda)^k)}{\Gamma[1+1/k]} dx \quad . \quad (\text{A.1})$$

Let $t = (x/\lambda)^k$, so $x = \lambda t^{1/k}$, and $dx = (\lambda/k)t^{(1/k)-1} dt$.

Making this change of variable in (A.1), we have, after some re-arrangement:

$$\begin{aligned} E_1 &= \frac{\lambda^k}{\Gamma\left[1 + \frac{1}{k}\right]} \int_0^\infty t^{1+\frac{1}{k}} \left[\ln(\lambda) + \frac{1}{k} \ln(t) \right] e^{-t} dt \\ &= \frac{\lambda^k \ln(\lambda)}{\Gamma\left[1 + \frac{1}{k}\right]} \int_0^\infty t^{1+\frac{1}{k}} e^{-t} dt + \frac{\lambda^k}{k\Gamma\left[1 + \frac{1}{k}\right]} \int_0^\infty t^{1+\frac{1}{k}} \ln(t) e^{-t} dt \quad . \end{aligned} \quad (\text{A.2})$$

Using Maple (2020) it is easy to verify that the first integral in (A.2) equals $\Gamma[(2k + 1)/k]$; and the second integral in (A.2) equals $\Gamma[1/k][(k + 1)\psi(1/k) + k(k + 2)]/k^2$. Inserting these expressions in (A.2) and using the recurrence formula for the gamma function, we immediately obtain the expression for E_1 in (8).

Similarly,

$$E_2 = E[X^k (\ln(X))^2] = \int_0^\infty \frac{x^k (\ln(x))^2 \left(\frac{x}{\lambda}\right)^k \left(\frac{k}{\lambda}\right) \exp(-(x/\lambda)^k)}{\Gamma[1+1/k]} dx \quad . \quad (\text{A.3})$$

Again, let $t = (x/\lambda)^k$, so $x = \lambda t^{1/k}$, and $dx = (\lambda/k)t^{(1/k)-1} dt$.

Making this change of variable in (A.3), we have, after some simplification:

$$\begin{aligned} E_2 &= \frac{\lambda^k}{\Gamma\left[1 + \frac{1}{k}\right]} \int_0^\infty t^{1+\frac{1}{k}} \left[(\ln(\lambda))^2 + \frac{2}{k} \ln(t) \ln(\lambda) + (\ln(t))^2 / k^2 \right] e^{-t} dt \\ &= \frac{\lambda^k (\ln(\lambda))^2}{\Gamma\left[1 + \frac{1}{k}\right]} \int_0^\infty t^{1+\frac{1}{k}} e^{-t} dt + \frac{2\lambda^k \ln(\lambda)}{k\Gamma\left[1 + \frac{1}{k}\right]} \int_0^\infty t^{1+\frac{1}{k}} \ln(t) e^{-t} dt + \frac{\lambda^k}{k^2 \Gamma\left[1 + \frac{1}{k}\right]} \int_0^\infty t^{1+\frac{1}{k}} (\ln(t))^2 e^{-t} dt. \end{aligned} \quad (\text{A.4})$$

The first two integrals in (A.4) have been evaluated above. Using Maple (2020), the third integral in (A.4) equals $\Gamma[1/k] \left[(k + 1)(\psi(1/k))^2 + 2k(k + 2)\psi(1/k) + (k + 1)\psi'(1/k) + 2k^2 \right] / k^2$. Inserting the expressions for these integrals in (A.4) and using the recurrence formula for the gamma function, we immediately obtain the expression for E_2 in (9).

Footnotes

* I am grateful to two anonymous referees for their helpful comments on the previous version of this paper.

1. It should be noted that other parameterizations of this distribution are commonly encountered.
2. The constraint that links the two shape parameters (p and d) in our case is fortuitous as it reduces the number of parameters to two. This avoids the well-known difficulties generally associated with the MLE for the generalized gamma distribution (*e.g.*, see Noufaily and Jones, 2012), as we will see below.
3. See Appendix 1 for the derivation of these expectations.
4. We ignore the possibility of censored sampling in this paper.
5. Throughout this paper, λ and k will be numbered as the first and second parameters respectively. In particular, this will be reflected in the choice of subscripts and superscripts in the rest of this section.
6. The “preventive” approach suggested by Firth (1993) provides another way of eliminating first-order bias analytically. Interestingly, Nagar’s (1959) method for bias elimination in k -class estimators can also be considered “preventive” rather than “corrective”. We do not pursue Firth’s method, because the analytic expressions needed to modify the likelihood function under his procedure are too unwieldy to be practical in this case.
7. The original Cox-Snell result involved a less convenient expectation of products of derivatives. In addition, Cordeiro and Klein showed that the bias expression holds even under non-independence, which is important for the time-series data used in section 6.2.
8. The supplement is available at <https://github.com/DaveGiles1949/My-Documents/find/master> or https://web.uvic.ca/~dgiles/downloads/working_papers/Supplement.pdf.
9. The R code is available from the author, on request.
10. Extremely similar results were obtained when the non-parametric bootstrap was used for bias reduction.
11. This phenomenon arises in similar studies involving the Cox-Snell correction (*e.g.*, Cribari-Neto and Vasconcellos, 2002; Giles, 2012; and Giles *et al.*, 2013, 2016).
12. The report also includes details of the survey methodology.
13. So, $n = 40$. See the first four columns of Summary Table 1 in Ontario Ministry of Natural Resources *et al.* (2010).
14. 10,000 bootstrap replications were used to obtain “stable” results.
15. Data were accessed on 1 November 2020, from <https://finance.yahoo.com/quote/TSLA?p=TSLA>. The longest such spell lasted for 12 weeks.
16. See Cordeiro and Cribari-Neto (2012, pp. 68-69).