NEARLY UNBIASED ESTIMATORS FOR THE THREE-PARAMETER WEIBULL DISTRIBUTION WITH GREATER EFFICIENCY THAN THE ITERATIVE LIKELIHOOD METHOD

Abstract

The maximum likelihood method is the most commonly used method to estimate the parameters of the three-parameter Weibull distribution. However, it returns biased estimates. In this paper, we show how to calculate weights which cancel the biases contained in the MLE equations. The exact weights can be computed when the population parameters are known and the expected weights when they are not. Two of the three weights' expected values are dependant only on the sample size whereas the third also depends on the population shape parameters. Monte Carlo simulations demonstrate the practicability of the weighted MLE method. Compared to the iterative MLE technique, the bias is reduced seven times (irrespective of the sample size) and the variability of the parameter estimates is reduced seven times for very small sample sizes but this gain disappears for large sample sizes.

Nearly unbiased estimators for the three-parameter Weibull distribution with greater efficiency than the iterative likelihood method

The Weibull distribution is often used in experimental psychology either to describe the response time data (e.g. Palmer, 1998, Burbeck and Luce, 1982) or to test a model (e.g. Cousineau and Shiffrin, 2004). In some applications, the Weibull distribution is just a convenient tool used to highlight the properties of interests in the data (e.g. Rouder, Lu, Speckman, Sun and Yiang., 2005). In other applications however, the Weibull distribution is a direct consequence of a cognitive model (e.g. Logan, 1988, Cousineau, 2004, van Zandt and Ratcliff, 1995, Tuerlinckx, 2004).

The most commonly used technique to estimate the parameters of a data set assuming the Weibull distribution is the maximum likelihood estimation technique (hereafter called the iterative MLE). The problem with that technique is that it returns biased estimators. The exact amount of bias is unknown and depends on the sample size. As a consequence, it is not possible to compare the parameter values obtained from samples of different sizes. As an example, for a very small sample (8 observations), the shape parameter can be underestimated by more than 40 % and the scale parameter by more than 30 % (these figures were approximated using Monte Carlo simulations described later). Overall, if the three parameters are seen as a vector, the length of the estimated vector is wrong by over 50 %. For medium size samples (n = 32), the vector is still wrong by over 10%. Averaging the parameters across multiple participants (assuming they have identical distributions of RT) can be used to reduce the variability of the estimates but this cannot eliminate the biases. Because experiments in psychology rarely have the same sample sizes, this is a major obstacle for cross-experiment comparisons, an obstacle that we wish to eliminate with the present paper.

Electrical engineers, who extensively use the Weibull distribution for voltage breakage, have developed many heuristics which aim at removing the bias of the obtained MLE estimates. However, these heuristics were developed for the two-parameter (no shift) Weibull distribution and cannot be generalized to the three-parameter Weibull distribution (see Cacciari and Montanari, 1994, for a review of some of these heuristics). Likewise, Hirose (1999) proposed to create a set of polynoms which, given the obtained parameters, would return the unbiased parameters. These polynoms are based on the biases found using Monte Carlo simulations for a large number of parameter values. However, this technique requires tremendous amount of computations and must be recalibrated for each implementation of the MLE program. Further, we found the gains to be modest (Cousineau, in preparation).

In the following, we extend the MLE technique by incorporating three weights in the solutions to the maximum likelihood equations. Exact expressions for these weights are given, but being based on the (unknown) population parameters, we instead examine their expected values. In doing so, we will need to move from an iterative MLE technique to a two-step method in which two of the parameters are estimated iteratively and the last one is estimated algebraically. Finally, Monte Carlo simulations will examine the performance of the iterative MLE vs. the two-step MLE, and within the two-step methods, the impact of the weights on the bias of the parameters.

The iterative and weighted MLE equations

In psychology, the iterative MLE technique is definitely the most commonly used technique (Heathcote, 1996, Cousineau and Larochelle, 1997, Dolan, 2000, Rouder, Sun, Speckman, Lu and Zhou, 2003, Heathcote, Brown and Cousineau, 2004, see Cousineau, Brown and Heathcote, 2004, for a review and van Zandt, 2000, for alternatives).

The best-fitting shape, scale and shift parameters are found by maximizing the likelihood function (or more commonly, the log of the likelihood function), performing a search in the parameter space (see Myung, 2003, for a tutorial). Let the true parameters be denoted by γ, β and α for the shape, the scale and the shift parameters respectively. Figure 1 illustrates some Weibull distributions varying on their shape.

Insert Figure 1 about here

The domains of the parameters are given by $\gamma \in \Gamma = \mathbb{R}^+$, $\beta \in B = \mathbb{R}^+$ and $\alpha \in A = \{\alpha < min(X), \alpha \in \mathbb{R}\}$ where X denotes the sample and $\mathbb{R}^+$ excludes zero.[1] The iterative MLE are obtained by performing a search over the three parameters simultaneously:

$$\left\{\hat{\gamma}, \hat{\beta}, \hat{\alpha}\right\}_{\text{MLE}} = \left\{\underset{\gamma\in\Gamma,\beta\in B,\alpha\in A}{\text{Max}} \log\left(\ell(\gamma,\beta,\alpha|X)\right)\right\} \tag{1}$$

where $\hat{\gamma}$, $\hat{\beta}$ and $\hat{\alpha}$ are the estimated parameters, and the likelihood function ℓ is given by:

$$\ell(\gamma,\beta,\alpha|X) = \prod_{i=1}^{n} f\left(x_i|\,\gamma,\beta,\alpha\right)$$

in which n is the sample size and f is the assumed probability density function (pdf) of x, given by the Weibull distribution:

$$f(x|\gamma,\beta,\alpha) = e^{-\left(\frac{x-\alpha}{\beta}\right)^{\gamma}}(x-\alpha)^{\gamma-1}\beta^{-\gamma}\gamma.$$

Its cumulative density function (cdf) is given by:

$$F(x|\gamma,\beta,\alpha) = 1 - e^{-\left(\frac{x-\alpha}{\beta}\right)^{\gamma}}.$$

The iterative MLE technique is used by most computer programs available (RTSYS, PASTIS, DISFIT and QMPE). It is very general and can be adapted to any distribution whose pdf is known.

Instead of relying on brut computational power, we can look for an analytical solution to (1) by looking for a maximum using the derivatives with respect to each of the parameters. The result, given in Equation 2, is composed of a system of equations for $\hat{\gamma}$ and $\hat{\alpha}$ and a second equation for $\hat{\beta}$ given $\hat{\gamma}$ and $\hat{\alpha}$. We call this solution a "2-step MLE" solution.

$$\begin{aligned} \{\hat{\gamma},\hat{\alpha}\}_{\text{MLE}} &= \left\{ \begin{matrix} \underset{\gamma\in\Gamma,\alpha\in A}{\text{Min}} \left(\frac{1}{\gamma} + \frac{1}{n}\sum_{i=1}^{n}\log\left(x_i-\alpha\right) - \frac{\sum_{i=1}^{n}\log(x_i-\alpha)(x_i-\alpha)^{\gamma}}{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}} \right)^2 \\ \underset{\gamma\in\Gamma,\alpha\in A}{\text{Min}} \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i-\alpha} \times \frac{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma-1}} - \frac{\gamma}{\gamma-1} \right)^2 \end{matrix} \right\} \\ [\hat{\beta}|\hat{\gamma},\hat{\alpha}]_{\text{MLE}} &= \sqrt[\gamma]{\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}} \end{aligned} \qquad (2)$$

In the above notation, the curled braces denote that a search is required whereas square braces indicate an algebraic solution.[2] Appendix A indicates how this solution is derived.

The equations in (2) are the well-known MLE solutions for the Weibull distributions (see e.g. Rockette, Antle and Klimko, 1974). What is less known is that they can also be obtained from simple algebraic manipulations. Jacquelin (1996) showed how to derive the solution in the case of the two-parameter (no shift) Weibull distribution; in Appendix B, we derive the solution for the three-parameter Weibull distribution. We call this solution the weighted MLE solution as it is identical to the standard MLE solution except for the presence of three weights, W_1, W_2 and W_3.

Overall, the weighted MLE are given by

$$\{\hat{\gamma},\hat{\alpha}\}_W = \left\{\begin{array}{c} \underset{\gamma\in\Gamma,\alpha\in A}{\text{Min}}\left(\frac{W_2}{\gamma}+\frac{1}{n}\sum_{i=1}^{n}\log\left(x_i-\alpha\right)-\frac{\sum_{i=1}^{n}\log(x_i-\alpha)(x_i-\alpha)^\gamma}{\sum_{i=1}^{n}(x_i-\alpha)^\gamma}\right)^2 \\ \underset{\gamma\in\Gamma,\alpha\in A}{\text{Min}}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i-\alpha}\times\frac{\sum_{i=1}^{n}(x_i-\alpha)^\gamma}{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma-1}}-W_3\right)^2 \end{array}\right\}$$

$$[\hat{\beta}|\hat{\gamma},\hat{\alpha}]_W = \sqrt[\gamma]{\frac{1}{nW_1}\sum_{i=1}^{n}(x_i-\alpha)^\gamma} \qquad (3)$$

in which the quantities W_1, W_2 and W_3 are defined as:

$$W_1 = \frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{1}{1-F\left(x_i\right)}\right)$$

$$W_2 = \frac{\sum_{i=1}^{n}\log\left(\frac{1}{1-F(x_i)}\right)\log\left(\log\left(\frac{1}{1-F(x_i)}\right)\right)}{\sum_{i=1}^{n}\log\left(\frac{1}{1-F(x_i)}\right)}-\frac{1}{n}\sum_{i=1}^{n}\log\left(\log\left(\frac{1}{1-F\left(x_i\right)}\right)\right)$$

$$W_3 = W_1\frac{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{1}{1-F(x_i)}\right)\right)^{-\frac{1}{\gamma}}}{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{1}{1-F(x_i)}\right)\right)^{\frac{\gamma-1}{\gamma}}}$$

Except for the introduction of W_1 and W_2 and the replacement of $\frac{\gamma}{\gamma-1}$ by W_3, the weighted MLE equations are identical to the standard MLE equations. The terms W_i can be seen as weights and the corresponding MLE weights are 1, 1 and $\frac{\gamma}{\gamma-1}$. However, they cannot be pulled out of the equations, so that it is not possible to unbias the estimates after the search has been completed.

When the true weights W_1, W_2 and W_3 are used, the estimated parameters are precisely the true population parameters whatever the sample. Hence, the estimated parameters $\hat{\alpha}=\alpha$, $\hat{\beta}=\beta$ and $\hat{\gamma}=\gamma$. In addition, $Var\left(\hat{\alpha}\right)=Var(\hat{\beta})=Var\left(\hat{\gamma}\right)=0$. The difficulty is that computing the true weights W_1, W_2 and W_3 requires $F(X_i)$ which in turn requires the true parameters. Of course, they are unknown in practical applications.

One way around this difficulty is to realize that the weights are random variables. Hence, instead of using the exact but unknown weights, we can infer their most probable values. In the following section, we examine the mean, the median and the geometric average of W_1, W_2 and W_3.

Three propositions related to the weights W_1, W_2 and W_3

We present a certain number of propositions related to the weights W_i. We were not able to demonstrate all of them; whenever we were not able to demonstrate a proposition, we checked that Monte Carlo simulations (described later) results were congruent with the propositions. Table 1 presents an overview of some of the results.

Insert Table 1 about here

Proposition 1 (the standard MLE weights are the asymptotic values of W_1, W_2 and W_3): the limit as *n* tends to ∞ of W_1 is 1, of W_2 is 1 and of W_3 is $\frac{\gamma}{\gamma-1}$ if γ is greater than 1. When γ is equal or smaller to 1, the limit of W_3 is divergent.

To prove the limit of W_1 and some of the following propositions, it is convenient to deduce the sampling distribution of W_1. Assume a population X following a Weibull distribution with true parameters γ, β and α. The quantity $\log\left(\frac{1}{1-F(X_i)}\right)$ equals $\left(\frac{x_i-\alpha}{\beta}\right)^{\gamma}$. Defining z_i as $\left(\frac{x_i-\alpha}{\beta}\right)^{\gamma}$, we find that the z_i are distributed following an (unshifted) exponential distribution with mean parameter 1 (Smith and Rose, 2002). The quantity

$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\alpha}{\beta}\right)^{\gamma}$ equals $\sum_{i=1}^{n}\frac{z_i}{n}$ which is a convolution of *n* exponentially distributed variates with mean parameter $\frac{1}{n}$. This is known to result in a Gamma distribution with parameters $\{n, \frac{1}{n}\}$ (Townsend and Ashby, 1983, Luce, 1986). The expected value of W_1 is therefore $n \times \frac{1}{n} = 1$, whatever the sample size.

Regarding W_2 and W_3, we were unable to derive their sampling distributions and to demonstrate the propositions.

Proposition 1, if correct, shows that for large *n*, the weighted MLE (Eq. 3) are equivalent to the standard MLE (Eq. 2). Among other things, it means that asymptotically, whenever standard MLE are efficient and normally distributed, so are the weighted MLE. As shown in Smith, 1985, this is the case when $\gamma \geq 2$.

Proposition 2: The sampling distribution of W_1 and W_2 depends only on the sample size whereas the sampling distribution of W_3 depends on the sample size and the shape parameter of the population.

Regarding W_1, we already showed that the sampling distribution is Gamma with parameters $\{n, \frac{1}{n}\}$. Regarding W_2, note that the quantity $\log\left(\frac{1}{1-F(X_i)}\right)$ equals $\left(\frac{x_i-\alpha}{\beta}\right)^{\gamma}$ which follows an exponential distribution with rate parameter 1. Hence, the three sums in W_2 are random variates which only depend on *n*. For W_3, we don't have a formal demonstration. However, we remark that the numerator follows a type-II distribution of extreme (Fréchet) with scale parameter $\frac{1}{n}$ and shape parameter γ whereas the denominator follows a type-II distribution of extreme with scale parameter $\frac{1}{n}$ and shape parameter $\frac{\gamma}{\gamma-1}$ (Galambos, 1978, Gumbel, 1958).

Proposition 3: Regarding W_1, its mean value is 1 (irrespective of sample size);

its median value is the median of a Gamma distribution with parameter {*n*, $\frac{1}{n}$}; and its geometric average value is $\frac{\exp(\psi(n))}{n}$ where $\psi(n)$ is the digamma function given by $\Gamma'(n)/\Gamma(n)$. Regarding W_2, its mean value is $1-\frac{1}{n}$.

The propositions regarding W_1 are all derived from the fact that its sampling distribution is a Gamma distribution with parameter {*n*, $\frac{1}{n}$}. The geometric average of W_1 was obtained by finding the expected value of the transformed variable $y_i = \log(z_i)$, since $G(x_i) = e^{\log(G(z_i))} = e^{\frac{\sum_{i=1}^{n}\log(z_i)}{n}} = e^{E[\log(z_i)]}$ where $G(x_i) = \sqrt[n]{\prod_{i=1}^{n} x_i}$ denotes the geometric average (Rose and Smith, 2000).

The proposition regarding the mean of W_2 is not demonstrated but Monte Carlo simulations suggested this result.

In what follows, we use the following shortcuts: E_i to denote the mean value of W_i, G_i to denote the geometric mean of W_i and J_i to denote the median value of W_i. As suggested by Proposition 3, J_1, J_2, G_1, G_2 and E_2 depend on *n* only. J_3, G_3 and E_3 depend on *n* and γ and E_1 is a constant. Whenever these quantities were not available in closed form, we used Monte Carlo simulations to estimate them. To do so, we replaced in W_1, W_2 and W_3, the terms $\log\left(\frac{1}{1-F(x_i)}\right) = \left(\frac{x_i-\alpha}{\beta}\right)^{\gamma}$ with z_i which is a random variable sampled form a standard exponential distribution In one simulation, *n* random deviates were sampled and the W_i estimated. This process was repeated 2^{20} times before computing the mean, median or geometric average.[3] Tables 2, 3 and 4 provide some values as a function of sample sizes (and as a function of γ for W_3).

Insert Tables 2, 3 and 4 about here

The weights E_3, G_3 and J_3 have the advantage over the MLE weights to be everywhere defined, including at γ = 1. Figure 2 shows the median weight J_3 as a function of γ for three sample sizes, along with the corresponding MLE weight $\frac{\gamma}{\gamma-1}$. First thing to note is that there is not discontinuity at γ =1 and that the median values of W_3 are everywhere positive. Second thing to note is that the differences between the MLE weights and the median weights becomes vanishingly small past γ = 2 (this is also true for the mean weight and the geometric average weight). It demonstrates that i) the sampling distribution of W_3 tends to become symmetrical for $\gamma \geq 2$ as the three measures of central tendency become equals; ii) the solutions therefore must be asymptotically unbiased whenever $\gamma \geq 2$. In other words, the MLE solution behave as if the Weibull distribution was satisfying the regularity conditions when $\gamma \geq 2$ (which is not the case), a results formally demonstrated by Smith (1985).

Insert Figure 2 about here

The mean of W_3 is very difficult to estimate. We found using the Monte Carlo simulations described above that its sampling distribution is skewed and with a high kurtosis, resulting in a long tail to the right. The tail becomes very important when $\gamma < 1$. For example, for a γ of 0.5, the median of W_3 obtained from 10,000 Monte Carlo simulations is approximately 13, but the largest value of W_3 was almost 900,000. We nevertheless provided estimates for E_3 in Table 4, but even the first digit is not reliable when $\gamma \leq 1$. In contrast, estimating the median and the geometric average of W_3 was easier, the first two digits being reliable when obtained from 2^{20} Monte Carlo simulations Figure 3 shows the distribution of W_3 obtained from 10,000 Monte Carlo simulations for three shape parameters.

Insert Figure 3 about here

Point estimates and median bias

To compare the techniques, we need to define a measure of exactitude showing how well the true parameters are recovered from random samples. Possible definitions of exactitude are the mean bias ($b_{\mathrm{Mn}} = E[\hat{\theta}] - \theta_T$ where θ is one parameter and E is the mean), the median bias ($b_{\mathrm{Md}} = Md[\hat{\theta}] - \theta_T$) and the modal bias ($b_{\mathrm{Mo}} = Mode[\hat{\theta}] - \theta_T$).

The mean bias is not a good choice because if the distribution of the estimates is skewed (as is certainly the case for small samples), the true parameter value is more likely to be near the mode than the mean of the distribution. The modal bias is not a convenient measure as it is difficult to estimate the mode of a sample. Hence, the median bias is the preferred measure of exactitude (e.g. Cacciari & Montanari, 1994; Jacquelin, 1997). In addition, the following property holds for any continuous increasing (decreasing) transformation *U*: $\mathrm{Md}(U(x)) = U(\mathrm{Md}(x))$.[4] In the following, we will use the median bias to evaluate the estimation techniques.

Bias of the scale parameter assuming γ and α are known

Regarding the parameter β given by equation 3, if we assume that γ and α are known, we can write:

$$\mathrm{Md}(\hat{\beta}) = \mathrm{Md}\left(\sqrt[\gamma]{\frac{1}{nW_1}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}\right) = \sqrt[\gamma]{\frac{1}{n\mathrm{Md}\left(W_1\right)}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}. \qquad (4)$$

Hence, the median bias of β when the median of the weight W_1 is used (J_1) should be zero.

This was confirmed by simulations. Across simulations, we varied the sample size *n* {8, 16, 32} and the population shape γ {0.5, 1.0, 1.5, 2.0, 2.5}. The scale and shift parameters, being scaling parameters (see Rouder et al., 2005, footnote 3 for a definition), were held constant at $\{\beta = 100, \alpha = 300\}$. The shape parameters were chosen such that there are two for which the regular MLE should be both efficient and normally distributed ($\gamma \geq 2$), two for which the regular MLE should work but where the distributions of the estimates are not normal ($1 \leq \gamma < 2$) and the last case ($\gamma < 1$) correspond to a situation for which there should be no consistent estimators (Smith, 1985).

For each simulation, a random sample of size *n* was generated and the best-fitting scale parameter $\hat{\beta}$ was estimated using Equation 3 with three different weights: the expected value of W_1 (E_1, which is always 1), its median value (J_1) and its geometric average value (G_1). For each combination of $n \times \gamma$, 2^{10} such simulations were run and the median estimated $\hat{\beta}$ was computed.

The results are reported in Table 5. As seen, when the median weight is used, the median bias is very close to zero, as expected from Eq. 4. More importantly, the median bias seems independent of the sample size and of the true population parameter. Hence, the weight J_1 returns a truly unbiased estimate of β (given the assumption that γ and α are known; more on this latter). Using the mean weight E_1 produces an underestimated median estimate and using G_1, an overestimated median estimate. In the last two cases, the estimates are better when *n* is large and when γ is large. This last result is understandable because as γ increases, the Weibull distribution becomes more symmetrical and the mean, median and geometric average weights tend toward identical values.

Insert Table 5 about here

Bias of the shape and shift parameters

Isolating the second weight from Eq. 3, we can write:

$$W_2 = \gamma \left(-\frac{1}{n} \sum_{i=1}^{n} \log\left(x_i - \alpha\right) + \frac{\sum_{i=1}^{n} \left(x_i - \alpha\right)^{\gamma} \log\left(x_i - \alpha\right)}{\sum_{i=1}^{n} \left(x_i - \alpha\right)^{\gamma}} \right)$$

such that, using the geometric mean (whose following property is true for two independent random variables: $G\left(x_i y_i\right) = G\left(x_i\right) G\left(y_i\right)$), we have:

$$G_2 = G\left(W_2\right) = G\left(\gamma \left(-\frac{1}{n} \sum_{i=1}^{n} \log\left(x_i - \alpha\right) + \frac{\sum_{i=1}^{n} \left(x_i - \alpha\right)^{\gamma} \log\left(x_i - \alpha\right)}{\sum_{i=1}^{n} \left(x_i - \alpha\right)^{\gamma}} \right) \right)$$

$$= G(\gamma) \times G\left(-\frac{1}{n} \sum_{i=1}^{n} \log\left(x_i - \alpha\right) + \frac{\sum_{i=1}^{n} \left(x_i - \alpha\right)^{\gamma} \log\left(x_i - \alpha\right)}{\sum_{i=1}^{n} \left(x_i - \alpha\right)^{\gamma}} \right).$$

Sadly, it is not possible to isolate $G(\gamma)$ in the other equation defining $\hat{\gamma}$ and $\hat{\alpha}$. Hence, it is not known whether G_2 and G_3 would result in small biases or not. For that reason, we ran another series of simulations estimating the parameters γ and α. We followed the exact same procedure as in the previous section.

Table 6 presents the results. As seen, the general quality of the estimations increases with increasing sample sizes for all weights used. In addition, the shift parameter α is generally well estimated. It is best estimated using the geometric average weights G_2 and G_3. The worse estimates are obtained when E_2 and E_3 are used whereas the median weights J_2 and J_3 returns median estimates that are in-between.

Regarding the shape parameter, it is poorly estimated by the mean weights E_2 and E_3 (bias for very small sample sizes *n* = 8 of over 10%). However, the bias is reduced fivefold by using the geometric mean weights $G_1$and G_3. Using the median weights J_2 and J_3 also does a god job, the bias being about half the bias of the mean weights E_2 and

E_3.

Using G_2 and G_3, the bias on γ rarely exceed 2%, whatever the true shape and the sample size. The solution is not bias-free however since the biases are affected by sample sizes. Contrary to the estimated scales, the estimated shapes and shifts are more accurately estimated when $\gamma \leq 1$. This is caused by the parameter α: When $\gamma \leq 1$, the distribution is exponential or hyper exponential and the smallest observation will generally be very close to the true α.

Insert Table 6 about here

Simultaneous estimates of the three parameters

The above results (unbiased estimates of β using J_1, nearly unbiased estimates of γ and α using G_2 and G_3) were established independently and, regarding β, assuming the exact values of γ and α. In practical applications, the three parameters will be estimated and the estimated $\hat{\beta}$ will depends on the estimated $\hat{\gamma}$ and $\hat{\alpha}$ as well. In addition, even though $\{\gamma, \alpha\}_{G_1,G_2}$ are least biased than $\{\gamma, \alpha\}_{J_1,J_2}$, it does not mean that the $\hat{\beta}$ obtained from the former will be least biased that the $\hat{\beta}$ obtained from the latter. Because this question is difficult to answer by examining the equations, we ran another series of simulations.

We first examine the usual iterative MLE solutions (Eq. 1), which is a 1-step method since all three parameters are estimated simultaneously. We also examined the two-step method of Eq. 2. It is the same as equation 3 except that the asymptotic weights 1, 1 and $\frac{\gamma}{\gamma-1}$ are used. Finally, we tested three sets of weights: $\{J_1, J_2, J_3\}$, $\{G_1, G_2, G_3\}$ and a mixture of weights: $\{J_1, G_2, G_3\}$. This last mixture of weights was tested because

G_2 and G_3 were the best weights for estimating γ and α whereas J_1 was the best weight for estimating β. The procedure used was identical to the one previously used.

Some of the results are presented in Table 7. The median bias is seen for each parameter individually. We also report a 3D-bias which is the median bias in the three-dimensional parameter space, measured as the median length of the vector separating the true parameters from the estimated parameters relative to the parameter vector length. In equation:

$$b_{\text{Md}}^{3D} = \frac{\text{Md}\left(\|\hat{\theta} - \theta\|\right)}{\|\theta\|} \times 100\%$$

where θ is the true parameter vector and $\hat{\theta}$ is one estimated parameter vector.

As seen, the iterative MLE does a very poor job for very small samples (n = 8) with biases on γ and β exceeding 25%. To see how bad this is, note that with a medium-size sample (n = 32), the bias is worse than for the two-step MLE technique with a small sample (n = 16) and also twice as worse than with the J_1, J_2 and J_3 weights with a very small sample (n = 8).

Insert Table 7 about here

The three set of weights outperformed the two-step MLE. The set $\{J_1, J_2, J_3\}$ (reported in Table 7) was the best one, followed by the mixture $\{J_1, G_2, G_3\}$. The set $\{G_1, G_2, G_3\}$ was last with bias approximately 50% larger than those of $\{J_1, J_2, J_3\}$. Still, the biases obtained by this last set were half those obtained by the 2-step MLE technique.

Examining the 3D biases for n = 16 and n = 32, we see that on average, the set of weights $\{J_1, J_2, J_3\}$ produced estimates that are about seven times less biased than the iterative MLE estimates (average 3D bias of 0.187 for the iterative MLE vs. 0.025 for the weighted MLE) and 3.5 times les biased than the 2-step MLE technique (average 3D bias of 0.085).

In parallel to bias, we also checked the efficiency of the techniques (efficiency measures the variability of the estimates). This measure is important since the general error of estimation is a function of both the systematic bias and the variability of the estimate such that a biased estimate with high efficiency (low variability) can sometimes be preferable to an unbiased estimate with weak efficiency.

We measured efficiency using the standard deviation of the estimates. Table 8 shows the results. As seen, the iterative MLE technique has the weakest efficiency for all the sample sizes. However, the difference diminishes as sample size increases. This was to be expected since the MLE technique is asymptotically the most efficient technique. The two-step MLE and the two-step weighted MLE (using J_1, J_2 and J_3) have nearly identical efficiencies, irrespective of the sample size and the population shape parameter (the same result was also found for the other two sets of weights not shown in Table 8).

Insert Table 8 about here

Overall, the simulations showed that the two-step technique using the weights J_1, J_2 and J_3 is considerably less bias than the MLE technique while having the same (large n) or better efficiency.

General discussion

The present paper showed how to modify the regular MLE equations to obtained nearly unbiased estimates. The exact values of the weights required are not known when the true parameters are unknown, but their most probable values can be used. The weighted MLE technique (with median weights) works best for the three parameter Weibull distribution and for the two-parameter Weibull distribution as well (as discussed in Appendix C).

As we mentioned ealier, the third weight used by the standard MLE technique is undefined when γ = 1 and inconsistent when γ < 1. Using J_3, the median of the weight W_3, avoids the problem as it is defined for any γ. The Minimum Product Spacing (MPS) technique solved this inconsistency problem using a slightly different approach: instead of changing the weights, Cheng and Amin (1983) changed the probability measure. Indeed, they replaced $f\left(x_i\right)$ with $\int_{x_i-1}^{x_i} f\left(x_i\right)\, dx$. With this new measure, infinities that occurred when the smallest x_i was equal to α are avoided. Nevertheless, this technique still returns biased estimates (Cousineau, Brown and Heathcote, 2004, Cousineau, in preparation). The same also applies to the derived techniques, the Quantile Maximum Product (QMPE, Heathcote, Brown and Cousineau, 2004) and the Quantile Product Spacing (QPS, Speckman and Rouder, 2004).

This paper leaves two open questions: What are the sampling distributions of W_2 and W_3? Having them in closed-form would avoid the use of approximations (Tables 3 and 4), might increase the precision of the estimates and speed-up the fitting process. Currently, we used lookup tables for G_2, G_3, J_2 and J_3 that are accurate to two digits only. Further, for the third weight, we only have the approximations as a function of γ between γ = 0.25 and γ = 2.75 by multiples of 0.25. In between, linear interpolations are used, even though the functions E_3, G_3 and J_3 are not linear with respect to γ (J_3 was seen in Figure 2). Despite these approximations, the technique can be programmed and automatized. [5]

References

Burbeck, S. L. & Luce, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. Perception and Psychophysics, 32, 117-133.

Cacciari, M., & Montanari, G. C. (1994). Generalization of the method of Maximum Likelihood. IEEE Transactions on Dielectrics and Electrical Insulation, 1, 545-547.

Cousineau, D. (2004). Merging race models and adaptive networks: A parallel race network. Psychonomic Bulletin & Review, 11, 807-825.

Cousineau, D. & Larochelle, S. (1997). PASTIS: A Program for Curve and Distribution Analyses. Behavior Research Methods, Instruments, & Computers, 29, 542-548.

Cousineau, D., & Shiffrin, R. M. (2004). Termination of a visual search with large display size effect. Spatial Vision, 17, 327-352.

Cousineau, D., Goodman, V. & Shiffrin, R. M. (2002). Extending statistics of extremes to distributions varying on position and scale, and implication for race models. Journal of Mathematical Psychology, 46, 431-454.

Dolan, C. (2000). DISFIT version 1.0: Provisional program documentation (Technical report series No. not numbered). University of Amsterdam: Department of psychology.

Galambos, J. (1978). The Asymptotic Theory of Extreme Order Statistics. New York: John Wiley and Sons.

Gumbel, E. J. (1958). The Statistics of Extremes. New York: Columbia University Press.

Heathcote, A. (1996). RTSYS: A computer program for analysing response time data. Behavior Research Methods, Instruments, & Computers, 28, 427-445.

Heathcote, A., Brown, S., & Cousineau, D. (2004). QMPE: Estimating Lognormal, Wald and Weibull RT distributions with a parameter dependent lower bound. Behavior Research Methods, Instruments, & Computers, 36, 277-290.

Hirose, H. (1999). Bias correction for the maximum-likelihood estimates in the two-parameter Weibull distribution. IEEE Transactions on Dielectrics and Electrical Insulation, 6, 66-68.

Hopkins, G. W. & Kristofferson, A. B. (1980). Ultrastable stimulus-reponse latencies:

Acquisition and stimulus control. Perception and Psychophysics, 27, 241-250.

Jacquelin, J. (1993). Generalization of the method of maximum likelihood. IEEE Transactions on Electrical Insulation, 28, 65-72.

Jacquelin, J. (1997). Inferfence on sampling on Weibull parameter estimation. IEEE Transactions on Dielectrics and Electrical Insulation, 4, 146-146.

Leadbetter, M. R., Lindgren, G. &, Rootzén, H. (1983). Extremes and related properties of random sequences and processes. New York: springer-Verlag.

Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, 95, 492-527.

Luce, R. D. (1986). Response times, their role in inferring elementary mental organization. New York: Oxford University Press.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. Journal of Mathematical Psychology, 47, 90-100.

Palmer, J. (1998). Attentional effects in visual search: relating search accuracy and search time, in Richard D. Wright (eds.). Visual attention (pp. 348-388). New York: Oxford University Press.

Rockette, H., Antle, C., & Klimko, L. A. (1974). Maximum likelihood estimation with the Weibull model. Journal of the American Statistical Association, 69, 246-249.

Rose, C., & Smith, M. D. (2000). Symbolic maximum likelihood estimation with Mathematica. The Statistician, 49, 229-240.

Rose, C., Smith, M., D. (2001). Mathematical Statistics with Mathematica. New York: Springer-Verlag.

Rouder, J. N., Lu, J., Speckman, P., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. Psychonomic Bulletin & Review, 12, 195-223.

Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A hierarchical bayesian statistical framework for response time distributions. Psychometrika, 68, 589-606.

Smith, R. L. (1985). Maximum likelihood estimation in a classe of nonregular cases. Biometrika, 72, 67-90.

Speckman, P. L., & Rouder, J. N. (2004). A comment on Heathcote, Brown and

Mewhort's QMLE estimation method for response time distributions. Psychonomic Bulletin & Review, 11, 574-576.

Townsend, J. T. & Ashby, F. G. (1983). Stochastic Modeling of Elementary Psychological Processes. Cambridge, England: Cambridge University Press.

Tuerlinckx, F. (2004). A multivariate counting process with Weibull-distributed first-arrival times. Journal of Mathematical Psychology, 48, 65-79.

Van Zandt, T. (2000). How to fit a response time distribution. Psychonomic Bulletin & Review, 7, 424-465.

Van Zandt, T. & Ratcliff, R. (1995). Statistical mimicking of reaction time data: Single-process models, parameter variability, and mixtures. Psychonomic Bulletin & Review, 2, 20-54.

Footnote

[1] In response time studies, γ is often restricted to be smaller than 5 as response time distributions are nearly always positively skewed or sometimes symmetrical (Hopkins and Kristofferson, 1980). Hence, γ should not be much beyond 3.602 (Rouder et al., 2005).

[2] Note by the way that it is easy to see that the MLE solution is applicable only when $\gamma > 1$ (Smith, 1985). Indeed, when the shape parameter is 1, the ratio $\frac{\gamma}{\gamma-1}$ found in the second equation defining γ and $\hat{\alpha}$ is undefined. Further, when $0 < \gamma < 1$, the term $-\frac{\gamma}{\gamma-1}$ takes a positive value. However, all the remaining terms of that equation are also all strictly positive. This means that there is no MLE solution in those cases. Hence, as was shown by Rockette et al. (1974), maximizing the likelihood function can lead to inconsistent estimates.

[3] For programmers, note that computing the median of W_1 requires the computation of $\Gamma(n) = (n-1)!$ (see the cdf of the Gamma distribution, Luce, 1986). This number rapidly exceeds the capacity of a long integer or a double float.

[4] This property is also true for the mode but not for the mean of a distribution.

[5] A *Mathematica* package, FitDataJacquelin.mx, along with two demos are available on the author's web site, http://www.mapageweb.umontreal.ca/cousined/papers/35-FitData/. This package was used for the simulations resulting in Tables 7, 8 and 9.

Appendix A: The 2-step MLE for the Weibull distribution

In this appendix, we derive the MLE for the Weibull distribution. First note that for the Weibull distribution,

$$\log(\ell(\gamma,\beta,\alpha|X)) = -n\gamma\log(\beta)+n\log(\gamma)+(\gamma-1)\sum_{i=1}^{n}\log(x_i-\alpha)-\beta^{-\gamma}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}$$

The derivative with respect to the scale parameter is

$$\frac{\partial\log(\ell(\gamma,\beta,\alpha|X))}{\partial\beta} = -\frac{n\gamma}{\beta}+\beta^{-\gamma-1}\gamma\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}$$

Setting it to zero, it can be reorganized into:

$$\beta = \sqrt[\gamma]{\frac{1}{n}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}} \tag{A.1}$$

It is the only parameter which can be isolated on the left-hand side of an equation. Note that it requires the knowledge of the other two parameters to be computed.

The derivative with respect to γ yields:

$$\frac{\partial\log(\ell(\gamma,\beta,\alpha|X))}{\partial\gamma} =$$

$$\frac{n}{\gamma}-n\log(\beta)+\sum_{i=1}^{n}\log(x_i-\alpha)+\beta^{-\gamma}\log(\beta)\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}-\beta^{-\gamma}\left(\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}\log(x_i-\alpha)\right)$$

Replacing $\beta^{-\gamma}$ with $\frac{1}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}$ (eq. A.1), setting the equation equals to zero, and dividing by *n* returns:

$$\frac{1}{\gamma}+\frac{1}{n}\sum_{i=1}^{n}\log(x_i-\alpha)-\frac{\sum_{i=1}^{n}\log(x_i-\alpha)(x_i-\alpha)^{\gamma}}{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}} = 0 \tag{A.2}$$

Similarly for α, we get:

$$\frac{\partial \log(\ell(\gamma,\beta,\alpha|X))}{\partial\alpha} = -(\gamma-1)\sum_{i=1}^{n}\frac{1}{x_i-\alpha} + \beta^{-\gamma}\gamma\sum_{i=1}^{n}(x_i-\alpha)^{\gamma-1}$$

which is set equal to zero. After some reorganizations, including a replacement of β^{γ} with $\frac{1}{n}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}$, we get:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i-\alpha} \times \frac{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma-1}} - \frac{\gamma}{\gamma-1} = 0 \tag{A.3}$$

Equations A.2 and A.3 define implicitly the two parameters γ and α, which in turn can be used to derive β. This solution is called in the text a two-step MLE method.

Appendix B: Deriving the weighted MLE solution from algebraic manipulations

In this appendix, we show how to derive the weighted MLE solution for the Weibull distribution using algebraic manipulations. First, let define W_1 as

$$\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{1}{1-F\left(x_i\right)}\right) \tag{B.1}$$

where $F\left(x_i\right)$ stands for short for $F\left(x_i\middle|\,\gamma,\beta,\alpha\right)$. From the Weibull cdf, we see that

$$W_1=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\alpha}{\beta}\right)^{\gamma}=\frac{1}{n\beta^{\gamma}}\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}$$

so that

$$\beta=\sqrt[\gamma]{\frac{1}{nW_1}\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}} \tag{B.2}$$

Note that Equation B.2 is nearly identical to the standard MLE solution except for the presence of the term W_1. Similarly, we can find the second weighted MLE equation with the following manipulations. Start with

$$\log\left(\frac{1}{1-F\left(x_i\right)}\right)=\left(\frac{x_i-\alpha}{\beta}\right)^{\gamma}=\frac{\left(x_i-\alpha\right)^{\gamma}}{\frac{1}{nW_1}\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}} \tag{B.3}$$

obtained by inserting the definition of β^{γ} (Eq. B.2). Taking the logarithms, we get:

$$\begin{aligned}\log\left(\log\left(\frac{1}{1-F\left(x_i\right)}\right)\right)&=\log\left(\frac{\left(x_i-\alpha\right)^{\gamma}}{\frac{1}{nW_1}\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}}\right)\\&=\gamma\log\left(x_i-\alpha\right)-\log\left(\frac{1}{nW_1}\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}\right)\end{aligned} \tag{B.4}$$

Multiply Equations B.3 and B.4 together, and take the average:

$$\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{1}{1-F\left(x_i\right)}\right)\log\left(\log\left(\frac{1}{1-F\left(x_i\right)}\right)\right)$$

$$=\gamma W_1\sum_{i=1}^{n}\frac{\log\left(x_i-\alpha\right)\left(x_i-\alpha\right)^{\gamma}}{\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}}-W_1\sum_{i=1}^{n}\frac{\log\left(\frac{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}{nW_1}\right)\left(x_i-\alpha\right)^{\gamma}}{\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}}$$

$$=\gamma W_1\frac{\sum_{i=1}^{n}\log\left(x_i-\alpha\right)\left(x_i-\alpha\right)^{\gamma}}{\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}}-W_1\log\left(\frac{1}{nW_1}\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}\right)\times\frac{\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}}{\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}}$$

Dividing both side by W_1 and replacing W_1 by its definition (Eq. B.1), we get:

$$\frac{\sum_{i=1}^{n}\log\left(\frac{1}{1-F(x_i)}\right)\log\left(\log\left(\frac{1}{1-F(x_i)}\right)\right)}{\sum_{i=1}^{n}\log\left(\frac{1}{1-F(x_i)}\right)}$$

$$=\gamma\frac{\sum_{i=1}^{n}\log\left(x_i-\alpha\right)\left(x_i-\alpha\right)^{\gamma}}{\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}}-\log\left(\frac{1}{nW_1}\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}\right)$$

Subtracting to this equation the mean of Eq. B.4, we obtain

$$\frac{\sum_{i=1}^{n}\log\left(\frac{1}{1-F(x_i)}\right)\log\left(\log\left(\frac{1}{1-F(x_i)}\right)\right)}{\sum_{i=1}^{n}\log\left(\frac{1}{1-F(x_i)}\right)}-\frac{1}{n}\sum_{i=1}^{n}\log\left(\log\left(\frac{1}{1-F\left(x_i\right)}\right)\right)$$

$$=\gamma\frac{\sum_{i=1}^{n}\log\left(x_i-\alpha\right)\left(x_i-\alpha\right)^{\gamma}}{\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}}-\log\left(\frac{1}{nW_1}\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}\right)-\frac{\gamma}{n}\sum_{i=1}^{n}\left(x_i-\alpha\right)+\log\left(\frac{1}{nW_1}\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}\right)$$

$$=\gamma\frac{\sum_{i=1}^{n}\log\left(x_i-\alpha\right)\left(x_i-\alpha\right)^{\gamma}}{\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}}-\frac{\gamma}{n}\sum_{i=1}^{n}\left(x_i-\alpha\right)$$

Let define the lhs of the above equation W_2. We get:

$$W_2=\gamma\frac{\sum_{i=1}^{n}\log\left(x_i-\alpha\right)\left(x_i-\alpha\right)^{\gamma}}{\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}}-\frac{\gamma}{n}\sum_{i=1}^{n}\left(x_i-\alpha\right) \tag{B.5}$$

so that:

$$\frac{W_2}{\gamma}+\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\alpha\right)-\frac{\sum_{i=1}^{n}\log\left(x_i-\alpha\right)\left(x_i-\alpha\right)^{\gamma}}{\sum_{i=1}^{n}\left(x_i-\alpha\right)^{\gamma}}=0 \tag{B.6}$$

Again, this formula is pretty much the same as the standard MLE solution except for the presence of a term W_2. Finally, we derive the last equation. We first examine

$$\frac{\beta}{x_i-\alpha}=\left(\sqrt[\gamma]{\log\left(\frac{1}{1-F\left(x_i\right)}\right)}\right)^{-1}$$

$$=\left(\log\left(\frac{1}{1-F\left(x_i\right)}\right)\right)^{-\frac{1}{\gamma}}.$$

By taking the summation and dividing both sides by *n*, we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\beta}{x_i-\alpha}=\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{1}{1-F\left(x_i\right)}\right)\right)^{-\frac{1}{\gamma}}$$

The left part further simplifies to

$$\frac{\beta}{n}\sum_{i=1}^{n}\frac{1}{x_i-\alpha}=\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{1}{1-F\left(x_i\right)}\right)\right)^{-\frac{1}{\gamma}} \tag{B.7}$$

In parallel, we examine the following ratio

$$\left(\frac{x_i-\alpha}{\beta}\right)^{\gamma-1}=\left(\sqrt[\gamma]{\log\left(\frac{1}{1-F\left(x_i\right)}\right)}\right)^{\gamma-1}$$

$$=\left(\log\left(\frac{1}{1-F\left(x_i\right)}\right)\right)^{\frac{\gamma-1}{\gamma}}$$

which yields, after summation and division by *n* on both sides:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\alpha}{\beta}\right)^{\gamma-1}=\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{1}{1-F\left(x_i\right)}\right)\right)^{\frac{\gamma-1}{\gamma}}$$

$$\frac{1}{n\beta^{\gamma-1}}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma-1}=\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{1}{1-F(x_i)}\right)\right)^{\frac{\gamma-1}{\gamma}} \tag{B.8}$$

If we divide Equation B.7 by Equation B.8, we get for the right hand side:

$$\frac{\frac{\beta}{n}\sum_{i=1}^{n}\frac{1}{x_i-\alpha}}{\frac{1}{n\beta^{\gamma-1}}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma-1}}=\beta^{\gamma}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(x_i-\alpha)\frac{1}{n}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma-1}}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(x_i-\alpha)}\times\frac{\frac{1}{nW_1}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma-1}}$$

and for the left hand side:

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{1}{1-F(x_i)}\right)\right)^{-\frac{1}{\gamma}}}{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{1}{1-F(x_i)}\right)\right)^{\frac{\gamma-1}{\gamma}}}$$

Defining W_3 as

$$W_3=W_1\frac{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{1}{1-F(x_i)}\right)\right)^{-\frac{1}{\gamma}}}{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{1}{1-F(x_i)}\right)\right)^{\frac{\gamma-1}{\gamma}}},$$

the whole equation becomes

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(x_i-\alpha)}\times\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma-1}}-W_3=0 \tag{B.9}$$

which is identical to the third MLE equation except for a term W_3.

Appendix C:

The two-parameter Weibull distribution

Because many disciplines use an unshifted Weibull distribution, we also verified the ability of the weighted MLE to estimate parameters. Because there are only two unknown parameters, we do not need three equations and the one involving W_3 was dropped. The method is thus:

$$\begin{aligned} \{\hat{\gamma}\}_W &= \left\{ \underset{\gamma\in\Gamma}{\text{Min}} \left(\frac{W_2}{\gamma} + \frac{1}{n}\sum_{i=1}^{n}\log\left(x_i\right) - \frac{\sum_{i=1}^{n}\log\left(x_i\right)x_i^{\gamma}}{\sum_{i=1}^{n}x_i^{\gamma}} \right)^2 \right\} \\ [\hat{\beta}|\hat{\gamma}]_W &= \sqrt[\gamma]{\frac{1}{nW_1}\sum_{i=1}^{n}x_i^{\gamma}} \end{aligned} \quad \text{(C.1)}$$

Simulations were run as previously except that there is no shift ($\alpha = 0$) and Equation C.1 was used for estimating the parameters. Table C.1 presents the median biases obtained for the iterative MLE and the J_1, J_2 weighted MLE technique. The two-step MLE results are not presented as they were strictly identical to the iterative MLE results. As seen, by comparison with Table 7, the parameters are more accurately estimated when there are only two unknown parameters. In addition, the scale parameter is equally well estimated using both techniques. Only the shape parameter benefits from the use of a weight (J_2 in this case).

Insert Table C.1 about here

Figure Captions

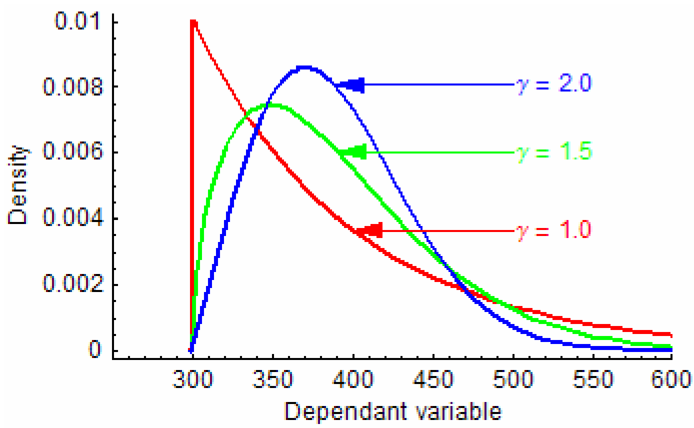

Figure 1. Examples of Weibull distributions with identical scale and shift parameters $\{\beta = 100, \alpha = 300\}$ but varying on the shape parameters. The case where $\gamma = 1$ is the shifted exponential distribution. The value of β and α were chosen so that the distribution resembles a distribution of response times of a well-trained participant.

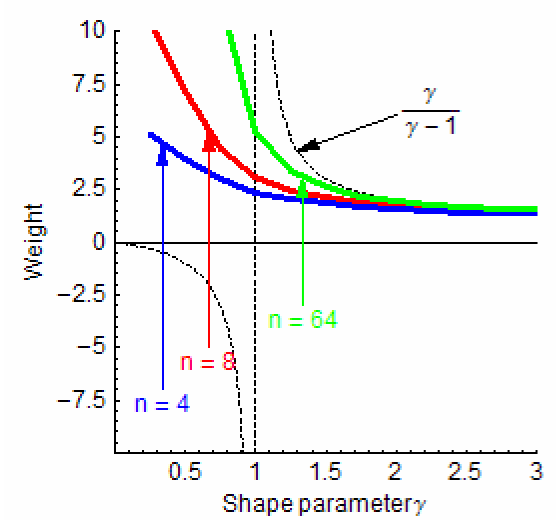

Figure 2. Median weight J_3 as a function of the shape parameter γ for three different sample sizes (n = 4, n = 8 and n = 64). The dashed line shows the MLE weight, given by $\frac{\gamma}{\gamma-1}$; this ratio is undefined at γ = 1 and negative for γ < 1.

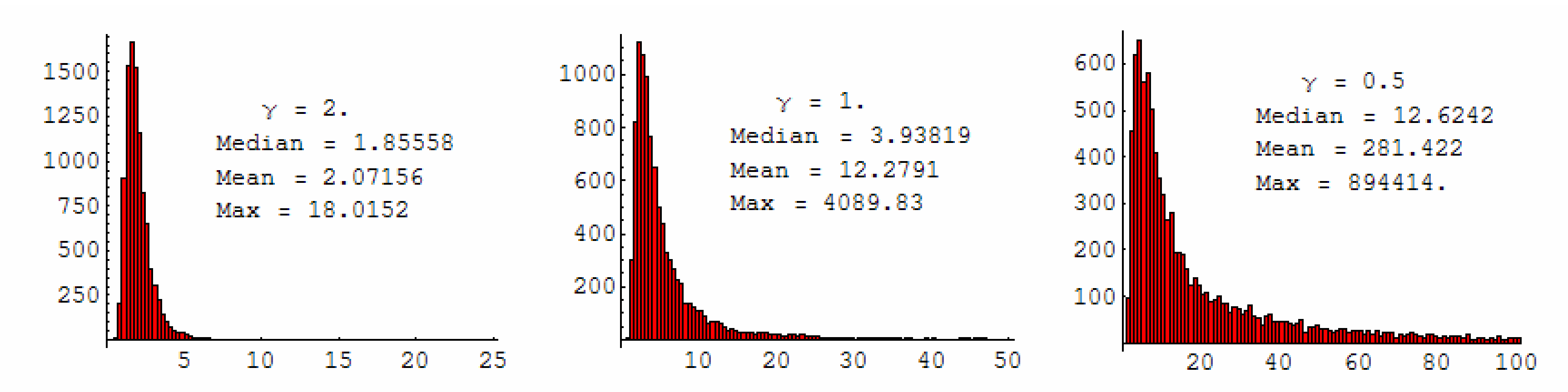

Figure 3: Distribution of the weight W_3 over 10,000 random samples of size $n = 16$. The shape parameter differs across panels. The closer γ is to 0, the more the distribution has a long right tail.

Table 1. The weights W_1, W_2 and W_3 and their mean, median and geometric average values

	W_1	W_2	W_3
Formulae	$\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{1}{1-F(x_i)}\right)$	$\frac{\sum_{i=1}^{n}\log\left(\frac{1}{1-F(x_i)}\right)\log\left(\log\left(\frac{1}{1-F(x_i)}\right)\right)}{\sum_{i=1}^{n}\log\left(\frac{1}{1-F(x_i)}\right)}-\frac{1}{n}\sum_{i=1}^{n}\log\left(\log\left(\frac{1}{1-F(x_i)}\right)\right)$	$W_1\frac{\sum_{i=1}^{n}\left(\log\left(\frac{1}{1-F(x_i)}\right)^{-\frac{1}{\gamma}}\right)}{\sum_{i=1}^{n}\log\left(\left(\frac{1}{1-F(x_i)}\right)^{\frac{\gamma-1}{\gamma}}\right)}$
MLE solution	1	1	$\frac{\gamma}{\gamma-1}$
Limit as $n \to \infty$	1	1	$\begin{cases}\infty & \text{when } \gamma \le 1\\ \frac{\gamma}{\gamma-1} & \text{when } \gamma > 1\end{cases}$
Mean E	1	$1-\frac{1}{n}$	? *
Geo. mean G	$\frac{e^{\psi(n)}}{n}$	?	?
Median J	x such that $\frac{e^{-nx}(nx)^n}{x\Gamma(n)}=\frac{1}{2}$	?	?
***E*, *G* and *J* values depends on**	*n*	*n*	*n* and γ

Note: $X=\{x_i\}$ represents a sample of size *n*;
$\Gamma(n)$ is the Gamma function; $\psi(n)$ is the polygamma function.
*: the mean of W_3 is very difficult to estimate numerically when $\gamma \le 1$.

Table 2: The quantities E_1, G_1 and J_1 for sample sizes 1 to 16. All these values are based on exact formula

n	E_1	G_1	J_1
1	1.000	0.561	0.693
2	1.000	0.763	0.839
3	1.000	0.839	0.891
4	1.000	0.878	0.918
5	1.000	0.902	0.934
6	1.000	0.918	0.945
7	1.000	0.929	0.953
8	1.000	0.938	0.959
9	1.000	0.945	0.963
10	1.000	0.950	0.967
11	1.000	0.955	0.970
12	1.000	0.959	0.972
13	1.000	0.962	0.974
14	1.000	0.965	0.976
15	1.000	0.967	0.978
16	1.000	0.969	0.979

Table 3: The quantitiesE_2, G_2 and J_2 for sample sizes 1 to 16. G_2 and J_2 are obtained through Monte Carlo simulations and are approximate to 2 digits.

n	E_2	G_2	J_2
1	0.000	0.000	0.000
2	0.500	0.163	0.275
3	0.667	0.409	0.517
4	0.750	0.553	0.638
5	0.800	0.642	0.711
6	0.833	0.702	0.759
7	0.857	0.742	0.791
8	0.875	0.775	0.817
9	0.889	0.800	0.838
10	0.900	0.820	0.853
11	0.909	0.835	0.867
12	0.917	0.849	0.877
13	0.923	0.860	0.886
14	0.929	0.871	0.895
15	0.933	0.879	0.902
16	0.938	0.887	0.908

Table 4: The quantities E_3, G_3 and J_3 for sample sizes 1 to 16 and for selected values of the shape parameter γ. All the values are obtained through Monte Carlo simulations and are approximate to 2 digits for G_3, J_3 and E_3, $\gamma > 1$; the precision of E_3, $\gamma \leq 1$ is unknown.

E_3

	γ = 0.5	γ = 1.0	γ = 1.5	γ = 2.0	γ = 2.5
1	12.429	20.157	11.371	12.483	24.796
2	19.452	11.567	4.183	3.147	2.771
3	33.320	14.372	3.701	2.596	2.225
4	73.132	36.570	3.480	2.411	2.043
5	66.530	14.812	3.431	2.309	1.948
6	87.540	11.751	3.297	2.235	1.888
7	124.230	21.331	3.270	2.198	1.852
8	99.608	13.230	3.192	2.170	1.831
9	97.148	14.622	3.178	2.154	1.808
10	447.600	19.335	3.244	2.143	1.794
11	105.660	13.879	3.195	2.120	1.779
12	164.510	13.765	3.154	2.113	1.769
13	136.390	12.762	3.109	2.109	1.759
14	342.220	14.270	3.111	2.099	1.755
15	168.430	14.737	3.110	2.091	1.746
16	198.130	13.641	3.101	2.093	1.742

G_3

	γ = 0.5	γ = 1.0	γ = 1.5	γ = 2.0	γ = 2.5
1	1.004	1.001	1.001	1.006	1.002
2	2.395	1.851	1.524	1.360	1.268
3	3.682	2.375	1.775	1.520	1.383
4	4.854	2.753	1.934	1.603	1.438
5	6.005	3.046	2.042	1.665	1.479
6	7.097	3.295	2.119	1.704	1.506
7	8.103	3.497	2.184	1.740	1.528
8	9.145	3.665	2.229	1.766	1.543
9	10.133	3.842	2.278	1.778	1.552
10	11.104	3.966	2.312	1.800	1.564
11	12.190	4.083	2.354	1.817	1.572
12	13.019	4.192	2.378	1.829	1.583
13	13.898	4.280	2.403	1.839	1.582
14	14.857	4.367	2.423	1.842	1.591
15	15.819	4.493	2.440	1.854	1.595
16	16.604	4.561	2.464	1.862	1.598

J_3

	γ = 0.5	γ = 1.0	γ = 1.5	γ = 2.0	γ = 2.5
1	1.001	0.999	0.999	0.995	0.998
2	2.096	1.668	1.456	1.339	1.262
3	3.081	2.082	1.680	1.479	1.367
4	3.950	2.381	1.822	1.567	1.428
5	4.806	2.631	1.920	1.625	1.464
6	5.631	2.808	2.004	1.669	1.492
7	6.433	2.982	2.056	1.698	1.509
8	7.150	3.114	2.105	1.722	1.525
9	7.931	3.252	2.151	1.739	1.537
10	8.643	3.365	2.180	1.758	1.552
11	9.319	3.462	2.207	1.774	1.555
12	10.051	3.560	2.239	1.782	1.565
13	10.746	3.642	2.262	1.793	1.570
14	11.379	3.713	2.285	1.804	1.578
15	12.069	3.780	2.301	1.813	1.581
16	12.743	3.854	2.324	1.820	1.586

Table 5: Median estimated scale parameter using the two-step weighted MLE technique with weights E_1, G_1 or J_1 when the true shape and shift parameters are known.

	True γ						
Weight used	0.5	1.0	1.5	2.0	2.5	Average bias	in percent
	***n* = 8**						
E_1	92.15	95.89	97.51	98.22	98.16	- 3.6	-3.6%
G_1	104.7	102.2	101.8	101.4	100.7	2.2	2.2%
J_1	100.3	100.0	100.3	100.3	99.83	*0.1*	0.1%
	***n* = 16**						
E_1	95.33	97.10	98.38	98.78	99.16	- 2.3	-2.3%
G_1	101.5	100.2	100.5	100.3	100.4	0.6	0.6%
J_1	99.61	99.76	99.76	99.82	99.99	- *0.2*	-0.2%
	***n* = 32**						
E_1	97.60	98.85	99.28	99.40	99.67	- 1.0	-1.0%
G_1	100.7	100.4	100.3	100.2	100.30	0.4	0.4%
J_1	99.67	99.89	99.97	99.93	100.08	- *0.1*	-0.1%

Note: Negative bias means that the parameter is underestimated

Table 6: Median estimated shape and shift parameters using the two-step weighted MLE technique with weights E_2 andE_3, G_2 and G_3, or J_2 and J_3.

		True γ						
Weights used		0.5	1.0	1.5	2.0	2.5	Average bias	in percent
	***n* = 8**							
E_2 & E_3	γ	0.537	1.079	1.661	2.226	2.763	0.153	10.2%
	α	300.	298.3	297.2	295.5	293.1	- 3.2	-3.2%
G₂ & G₃	γ	0.486	0.966	1.486	1.985	2.466	- 0.022	*-1.5%*
	α	300.1	300.1	299.4	298.3	296.6	- *1.1*	*-1.1%*
J_2 & J_3	γ	0.508	1.011	1.561	2.086	2.591	0.051	3.4%
	α	300.1	299.5	298.4	297.	295.	- 2.0	-2.0%
	***n* = 16**							
E_2 & E_3	γ	0.518	1.032	1.587	2.113	2.664	0.083	5.5%
	α	300.	299.1	298.3	297.	296.4	- 1.8	-1.8%
G₂ & G₃	γ	0.497	0.987	1.515	2.015	2.537	*0.010*	*0.7%*
	α	300.	299.5	299.1	298.2	297.7	- *1.1*	*-1.1%*
J_2 & J_3	γ	0.505	1.005	1.544	2.056	2.589	0.040	2.7%
	α	300.	299.4	298.8	297.7	297.1	- 1.4	-1.4%
	***n* = 32**							
E_2 & E_3	γ	0.506	1.02	1.542	2.059	2.591	0.044	2.9%
	α	300.	299.8	298.4	298.1	297.6	- 1.2	-1.2%
G₂ & G₃	γ	0.496	1.	1.508	2.01	2.528	*0.008*	*0.5%*
	α	300.	299.9	298.7	298.6	298.2	- *0.9*	*-0.9%*
J_2 & J_3	γ	0.5	1.008	1.523	2.031	2.553	0.023	1.5%
	α	300.	299.8	298.6	298.4	298.	- 1.0	-1.0%

Note: Note: Negative bias means that the parameter is underestimated
The percent bias for the parameter α is relative to the scale parameter.

Table 7: Median estimated parameters using the iterative MLE technique, the two-step MLE technique and the two-step weighted MLE technique with weights J_1, J_2 and J_3.

Method		*True* γ						
		0.5	1.0	1.5	2.0	2.5	Average bias	in percent
	***n* = 8**							
Iterative MLE	γ	0.366	0.447	0.487	1.288	1.703	- 0.642	43.4%
	β	93.45	60.78	61.39	61.95	66.8	- 31.13	31.1%
	α	300.7	307.3	316.5	320.	323.5	13.60	-13.6%
	3D	27.5 %	67.9 %	78.0 %	52.5 %	46.7 %	54.5 %	
two-step MLE	γ	0.608	1.23	1.851	2.499	3.162	0.370	-23.9%
	β	99.14	98.67	103.3	106.5	109.5	3.422	-3.4%
	α	299.7	296.9	294.4	292.8	289.6	- 5.320	5.3%
	3D	21.6 %	23.1 %	23.7 %	25.9 %	28.4 %	24.5 %	
two-step J_1, J_2, J_3	γ	0.508	1.03	1.529	2.07	2.605	0.048	*-2.8%*
	β	93.9	94.21	97.51	99.33	103.2	- 2.370	*2.4%*
	α	300.	299.7	298.8	298.3	294.6	- 1.720	*1.7%*
	3D	6.3 %	6.5 %	3.2 %	3.6 %	5.6 %	5.0 %	
	***n* = 16**							
Iterative MLE	γ	0.442	0.661	1.248	1.726	2.099	- 0.265	18.4%
	β	107.3	80.3	81.38	83.05	82.09	- 13.18	13.2%
	α	300.2	304.2	308.6	311.4	314.6	7.80	-7.8%
	3D	13.7 %	39.2 %	25.2 %	22.1 %	24.5 %	24.9 %	
two-step MLE	γ	0.55	1.094	1.667	2.243	2.811	0.173	-11.0%
	β	97.42	100.1	101.7	103.6	104.4	1.444	-1.4%
	α	300.0	299.2	297.	296.	295.6	- 2.440	2.4%
	3D	10.3 %	9.4 %	11.3 %	12.7 %	13.3 %	11.4 %	
two-step J_1, J_2, J_3	γ	0.513	1.014	1.538	2.06	2.578	0.041	*-2.5%*
	β	95.88	98.3	99.75	100.9	101.4	- 0.754	*0.8%*
	α	300.	299.8	298.4	298.	298.	- 1.160	*1.2%*
	3D	4.8 %	2.2 %	2.6 %	3.2 %	3.5 %	3.3 %	
	***n* = 32**							
Iterative MLE	γ	0.469	0.827	1.393	1.829	2.353	- 0.126	9.0%
	β	105.2	90.32	91.53	90.79	91.31	- 6.17	6.2%
	α	300.1	301.8	304.4	306.8	306.	3.82	-3.8%
	3D	8.0 %	19.8 %	11.2 %	12.8 %	10.7 %	12.5 %	
two-step MLE	γ	0.515	1.043	1.594	2.115	2.678	0.089	-5.3%
	β	100.1	99.9	101.7	101.8	101.6	1.020	-1.0%
	α	300.0	299.6	298.3	297.5	297.4	- 1.440	1.4%
	3D	3.0%	4.3 %	6.5 %	6.1 %	7.4 %	5.5 %	
two-step J_1, J_2, J_3	γ	0.498	1.007	1.535	2.033	2.569	0.028	*-1.4%*
	β	99.6	99.36	100.8	100.8	100.5	0.212	*-0.2%*
	α	300.0	299.7	298.8	298.3	298.4	- 0.960	*1.0%*
	3D	0.5 %	0.9 %	2.5 %	1.9 %	2.8 %	1.7 %	

Note: The percent bias for the parameter α is relative to the scale parameter;
Negative bias means that the parameter is underestimated;
3D represents the median distance between the true vector parameter and the estimated vector parameter relative to the length of the true vector parameter.

Table 8: Efficiency of the estimates (measured by the standard deviation) using the iterative MLE technique, the two-step MLE technique and the two-step weighted MLE technique with weights J_1, J_2 and J_3.

Method		True γ 0.5	1.0	1.5	2.0	2.5	Average efficiency
	***n* = 8**						
Iterative MLE	γ	0.084	0.321	0.777	1.41	1.813	0.881
	β	59.54	37.59	35.22	40.05	41.86	42.85
	α	2.569	5.425	14.54	24.68	29.15	15.27
two-step MLE	γ	0.116	0.158	0.175	0.199	0.207	0.171
	β	59.34	25.91	20.31	18.92	19.22	28.74
	α	1.853	7.211	12.69	14.52	16.55	10.56
two-step J_1, J_2, J_3	γ	0.101	0.133	0.154	0.178	0.182	*0.150*
	β	56.38	25.2	19.76	17.2	18.03	*27.31*
	α	1.833	7.127	11.82	14.07	16.06	*10.18*
	***n* = 16**						
Iterative MLE	γ	0.074	0.142	0.536	0.582	0.918	0.450
	β	41.22	20.73	18.89	20.29	26.52	25.53
	α	0.666	3.94	7.635	14.94	21.48	9.732
two-step MLE	γ	0.075	0.111	0.129	0.156	0.17	0.128
	β	36.13	18.39	13.31	12.29	12.42	18.51
	α	0.519	3.969	7.201	9.291	10.61	6.318
two-step J_1, J_2, J_3	γ	0.07	0.104	0.124	0.148	0.161	*0.121*
	β	35.67	17.94	13.01	12.	12.09	*18.14*
	α	0.535	3.991	7.057	9.158	10.35	*6.218*
	***n* = 32**						
Iterative MLE	γ	0.051	0.105	0.202	0.348	0.587	0.259
	β	31.38	11.47	11.41	12.64	16.4	16.66
	α	0.181	1.903	4.704	9.004	14.63	6.08
two-step MLE	γ	0.048	0.084	0.099	0.111	0.135	0.095
	β	25.54	12.43	8.894	8.603	7.789	12.65
	α	0.153	1.833	4.036	5.76	6.769	3.71
two-step J_1, J_2, J_3	γ	0.047	0.082	0.096	0.11	0.133	*0.094*
	β	25.43	12.39	8.927	8.603	7.847	*12.64*
	α	0.156	1.861	3.999	5.718	6.81	*3.71*

Note: The percent bias for the parameter α is relative to the scale parameter; Negative bias means that the parameter is underestimated.

Table C.1: Median estimated parameters using the iterative MLE technique and the two-step weighted MLE technique with weights J_1 and J_2 for the two-parameter Weibull distribution.

Method		0.5	1.0	1.5	2.0	2.5	Average bias	in percent
		True γ						
	n = 8							
Iterative MLE	γ	0.564	1.145	1.690	2.265	2.895	0.212	13.8%
	β	92.26	97.2	98.7	98.54	98.39	- 2.98	-3.0%
two-step J_1, J_2	γ	*0.496*	*1.01*	*1.488*	*2.004*	*2.537*	*0.007*	*0.2%*
	β	*92.6*	*96.8*	*98.5*	*98.4*	*98.2*	*- 3.10*	*-3.1%*
	n = 16							
Iterative MLE	γ	0.543	1.061	1.562	2.115	2.649	0.086	6.1%
	β	101.4	98.78	98.61	100.0	100.1	- 0.222	-0.2%
two-step J_1, J_2	γ	*0.501*	*1.001*	*1.475*	*1.990*	*2.501*	*- 0.006*	*-0.4%*
	β	*100.9*	*98.51*	*98.47*	*99.87*	*100.0*	*- 0.450*	*-0.5%*
	n = 32							
Iterative MLE	γ	0.519	1.031	1.545	2.085	2.575	0.051	3.4%
	β	98.84	99.05	99.78	99.89	99.83	- 0.522	-0.5%
two-step J_1, J_2	γ	*0.504*	*1.000*	*1.501*	*2.026*	*2.499*	*0.006*	*0.4%*
	β	*98.51*	*98.89*	*99.69*	*99.82*	*99.78*	*- 0.662*	*-0.7%*

Note: Negative bias means that the parameter is underestimated;