Model Selection for Reliability Estimation in Series Systems

Alex Towell
lex@metafunctor.com

Abstract

This paper explores model selection for reliability estimation of components in multi-component series systems. We assess the sensitivity of a likelihood model based on ? that includes masked data (right-censoring and candidate sets indicative of masked failure causes) to deviations from a well-designed series system. We also explore how a reduced model with homogeneous component shapes simplifies analysis but may be inadequate when the deviations are too great. Appropriateness of the reduced model is assessed using likelihood ratio tests. Findings suggest the reduced model excels for well-designed systems. More complex models are favored given divergent component properties or large samples. Proper model specification balances simplicity against representativeness.

Contents

1 Introduction **1**

2 Likelihood Model **5**

3 Simulation Study: Sensitivity Analysis to Changing System Design **6**
3.1 Scenario: Assessing the Impact of Changing the Scale Parameter of Component 3 6
3.2 Scenario: Assessing the Impact of Changing the Shape Parameter of Component 3 8

4 Weibull Series Homogenous Shape Model **10**
4.1 Assessing the Appropriateness of the Reduced Model 10
4.2 Simulation Study: Full Weibull Model vs Reduced (Homogenous Shape) Model 11
4.3 Implications and Recommendations 12

5 Conclusion **12**

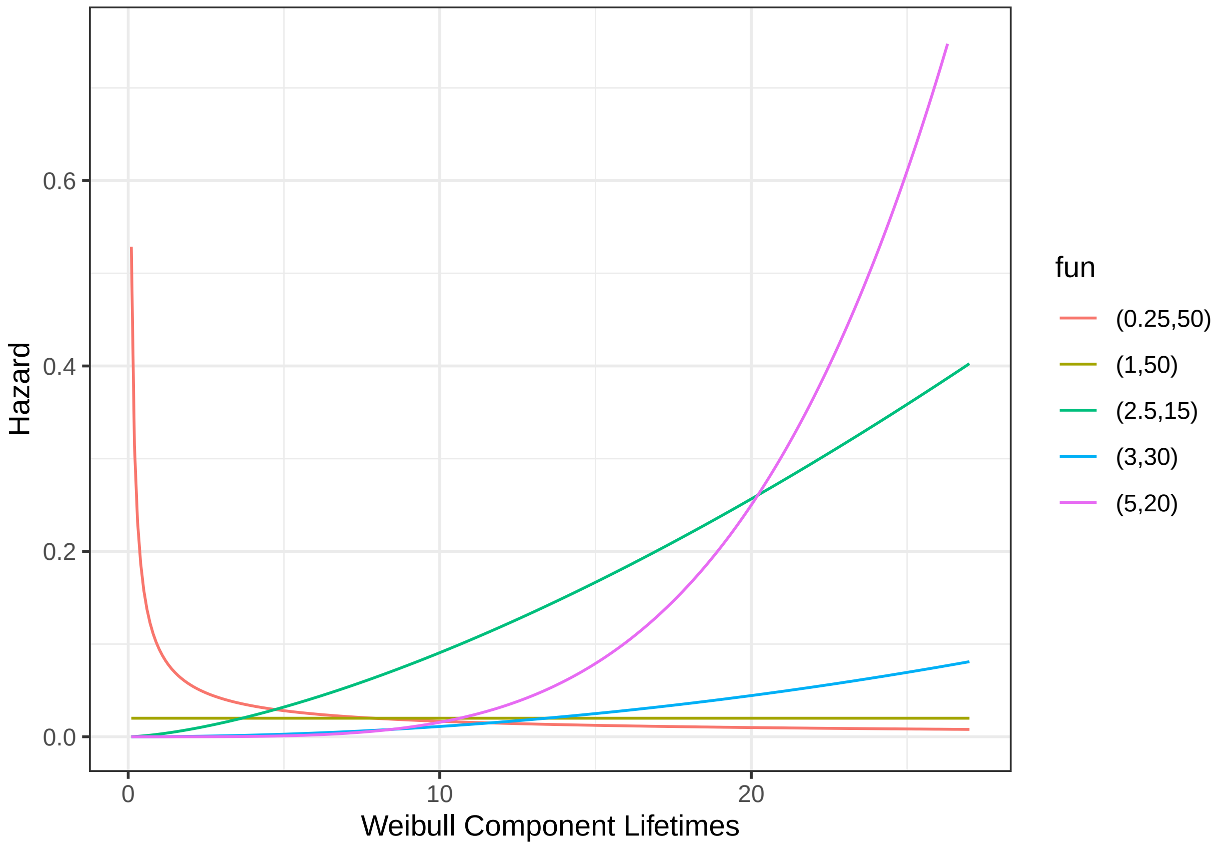
1 Introduction

Estimating reliability of individual components in multi-component systems is challenging when only system-level failure data is observable. A likelihood model incorporating right-censoring and candidate sets was previously developed to enable component inference from such masked data [1]. Simulation studies revealed estimator behavior and sensitivity to modeling assumptions given limited samples.

A key question is choosing an appropriate model complexity. A reduced model assuming homogeneous component shapes simplifies analysis as the system becomes Weibull. However, deviations in component properties impact adequacy. Proper model specification requires balancing simplicity against representativeness.

This paper explores model selection for component reliability estimation in series systems. The likelihood model from is summarized. Simulation studies demonstrate estimator sensitivity and assess reduced model appropriateness using likelihood ratio tests. Findings provide guidance on suitable models based on system design and available data.

```
## Warning: Removed 3662 rows containing missing values (`geom_line()`).
```



Plots of five different components with Weibull distributed lifetimes. Key observations:

- Components with shapes < 1 have decreasing hazards, e.g., component 1.
- Components with shapes > 1 have increasing hazards, e.g., components 3, 4, and 5.
- Components with shapes $= 1$ have constant hazards, e.g., component 2.

The shape parameter k may be understood in the following way:

- If $k < 1$, then the hazard function decreases with respect to system lifetime, which may occur if defective items fail early and are weeded out.
- If $k > 1$, then the hazard function is increases with respect to time, which may occur as a result of an aging process.
- If $k = 1$, then the failure rate is constant, which means it is exponentially distributed.

See Figure **??** for plots of the hazard and pdf functions of five different Weibull distributed components. We will use these plots to illustrate the different shapes of the hazard and pdf functions. The first component has a shape parameter $k = 0.25$, which is less than 1, and so the hazard function decreases with respect to time. The second component has a shape parameter $k = 1$, and so the hazard function is constant. The third, fourth, and fifth components have shape parameters $k = 2.5$, $k = 3$, and $k = 5$, respectively, and so the hazard functions increase with respect to time.

The lifetime of the series system composed of m Weibull components has a reliability function given by

$$R(t; \boldsymbol{\theta}) = \exp\left\{-\sum_{j=1}^{m}\left(\frac{t}{\lambda_j}\right)^{k_j}\right\}. \tag{1}$$

Proof. By Theorem **??**,

$$R(t; \boldsymbol{\theta}) = \prod_{j=1}^{m} R_j(t; \lambda_j, k_j).$$

Plugging in the Weibull component reliability functions obtains the result

$$
\begin{aligned}
R(t;\boldsymbol{\theta}) &= \prod_{j=1}^{m} \exp\left\{-\left(\frac{t}{\lambda_j}\right)^{k_j}\right\} \\
&= \exp\left\{-\sum_{j=1}^{m}\left(\frac{t}{\lambda_j}\right)^{k_j}\right\}.
\end{aligned}
$$

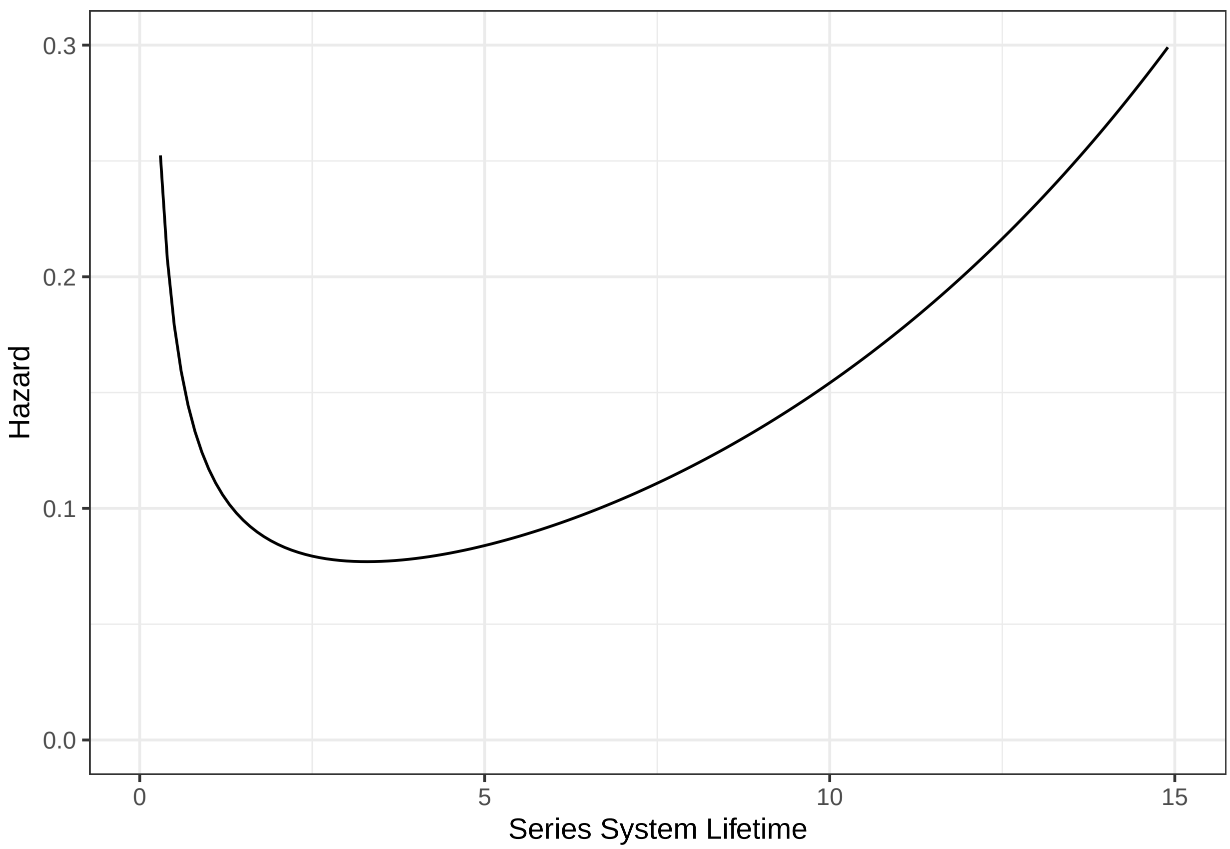
□

The Weibull series system's hazard function is given by

$$
h(t;\boldsymbol{\theta}) = \sum_{j=1}^{m} \frac{k_j}{\lambda_j}\left(\frac{t}{\lambda_j}\right)^{k_j-1}, \tag{2}
$$

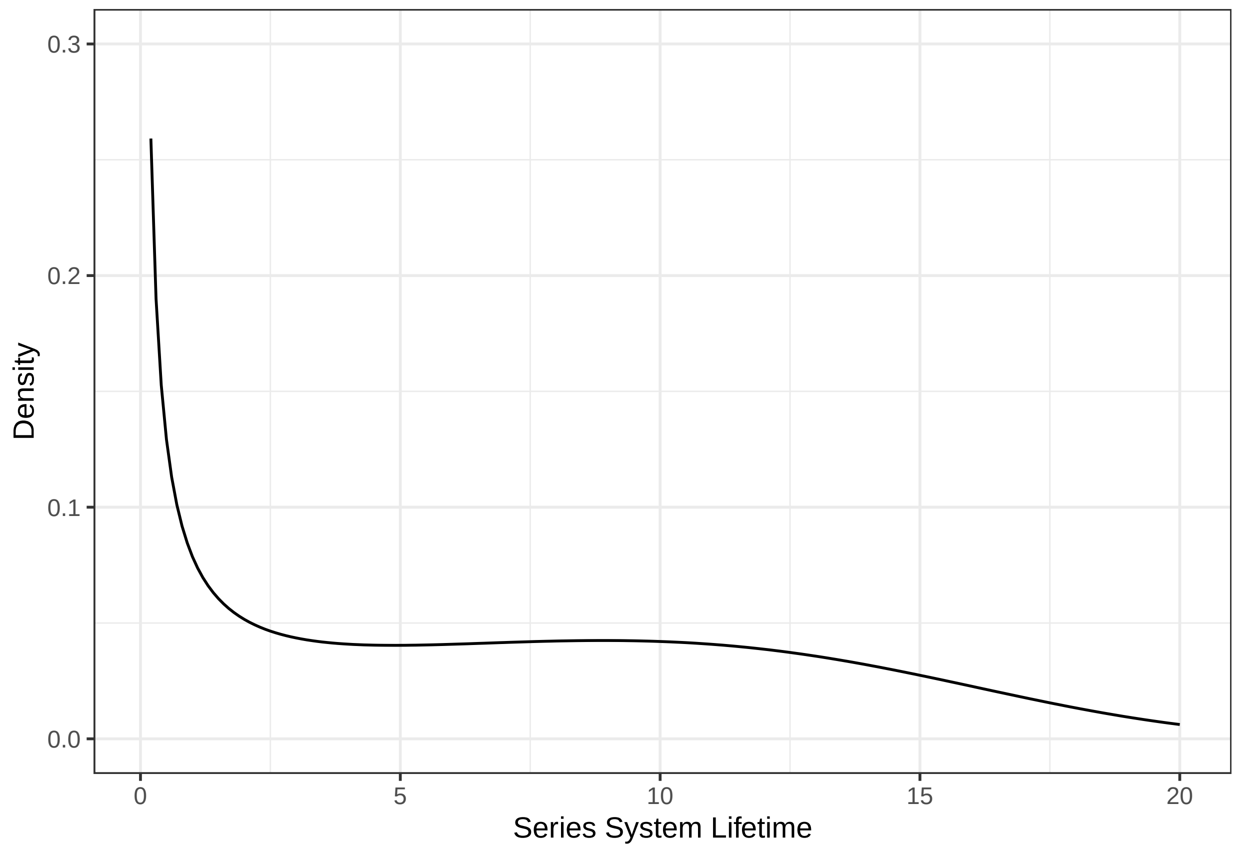
whose proof follows from Theorem **??**.

```
#fig.cap = "System lifetime plots", fig.align = "center", echo=F
h.series.ts <- hazard_wei_series(ts,
    scales=c(scale1,scale2,scale3,scale4,scale5),shapes=c(shape1,shape2,shape3,shape4,shape5))
df.series.haz <- data.frame(t=ts, y=h.series.ts)
ggplot(df.series.haz,aes(x=t, y=y)) +
    geom_line() +
    ylim(0,.3) +
    xlim(0.02,15) +
    theme_bw() +
    labs(y = "Hazard",
         x = "Series System Lifetime")
## Warning: Removed 854 rows containing missing values (`geom_line()`).
```



```r
f.series.ts <- dwei_series(ts,
    scales=c(scale1,scale2,scale3,scale4,scale5),shapes=c(shape1,shape2,shape3,shape4,shape5))

df.series.f <- data.frame(t=ts, y=f.series.ts)
ggplot(df.series.f,aes(x=t, y=y)) +
    geom_line() +
    labs(y = "Density",
         x = "Series System Lifetime") +
    ylim(0,.3) +
    xlim(.1,20) +
    theme_bw()
## Warning: Removed 802 rows containing missing values (`geom_line()`).
```



Plots of the hazard function and the pdf of a series system with the previously discussed Weibull components.

Key observations: 1. The hazard is initially large but decreases to some minimum before increasing again, exhibiting both a high infant mortality rate and an aging process. This is a pattern we see in nature (e.g., humans). 2. The pdf has a rather unusual form, a result of being a combination of Weibull distributions.

In Figure **??**, we plot the hazard function and the pdf of the Weibull series system with the component lifetime parameters considered earlier,

$$
\begin{aligned}
T_{i1} &\sim \mathrm{WEI}(3, 30)\\
T_{i2} &\sim \mathrm{WEI}(2, 50)\\
T_{i3} &\sim \mathrm{WEI}(0.5, 15)\\
T_{i4} &\sim \mathrm{WEI}(5, 20)\\
T_{i5} &\sim \mathrm{WEI}(0.25, 50)
\end{aligned}
$$

for the i-th series system where $i = 1, \ldots, n$. By Theorem **??**, the series system has a random lifetime given by

$$T_i = \min\{T_{i1}, \ldots, T_{i5}\}.$$

where $\boldsymbol{\theta} = (k_1, \lambda_1, \ldots, k_5, \lambda_5)$.

The series system, due to being a mixture of Weibull components with different shapes, has both a high infant mortality rate and an aging process, which is reflected in the plot of the hazard function. The hazard is initially high then decreases to some minimum before increasing again. This is a pattern we see in nature, e.g., electronic appliances may fail early due to defects, but those that survive the initial period of high failure rate can be expected to last for a long time before finally wearing out due to an aging process.

The pdf of the series system also appears to be multimodal, where the modes correspond to the high infant mortality rate and the aging process.

The pdf of the series system is given by

$$f(t; \boldsymbol{\theta}) = \left\{ \sum_{j=1}^{m} \frac{k_j}{\lambda_j} \left(\frac{t}{\lambda_j} \right)^{k_j - 1} \right\} \exp\left\{ -\sum_{j=1}^{m} \left(\frac{t}{\lambda_j}\right)^{k_j} \right\}. \tag{3}$$

The series system, due to being a mixture of Weibull components with different shapes, has both a infant mortality and aging phases. The hazard is initially decreasing but eventually increasing. This is a pattern we see in nature, e.g., electronic appliances may fail early due to defects, but those that survive the initial period of high failure rate can be expected to last for a long time before finally wearing out due to an aging process. The pdf of the series system also appears to be multi-modal due to the different modes of the components.

2 Likelihood Model

Consider a system with m components in series, where the j^{th} component has a Weibull distributed lifetime with scale λ_j and shape k_j. The system fails when any component fails. The system lifetime distribution is determined by component properties.

Only system failure times are observed, potentially right-censored. For failures, a candidate set indicates possible failed components. This masked data enables component inference despite lack of direct component observations.

A likelihood model for the masked data was derived in, accounting for right-censoring and candidate sets. The model relies on assumptions that candidate sets contain the failed component and components are equally likely to be masked given failure time and cause.

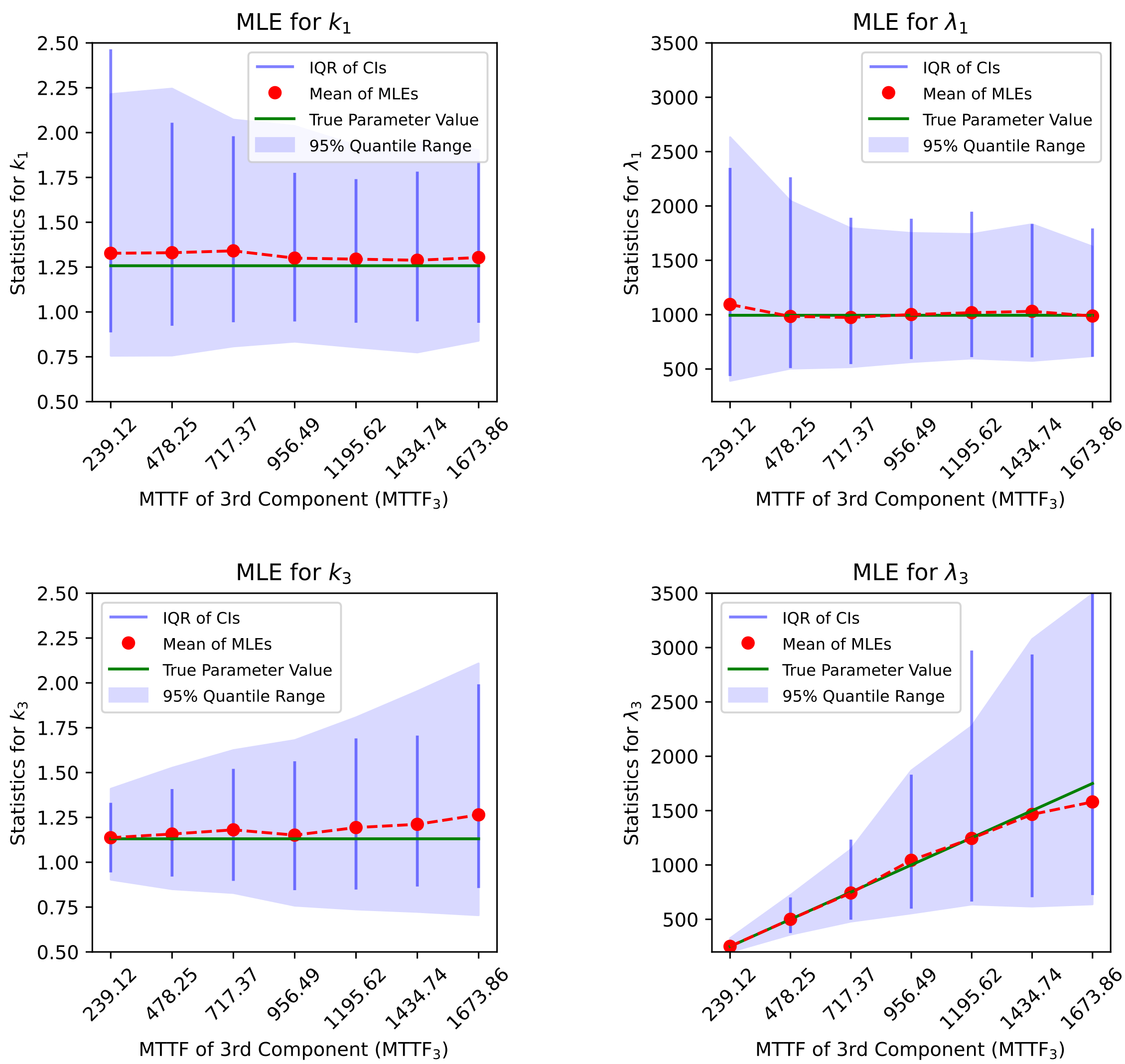
Maximum likelihood estimation produced accurate results despite small samples and significant masking and censoring. However, shape parameters were more variable and challenging to estimate precisely. A reduced model with homogeneous shapes simplified analysis while reducing estimator variability.

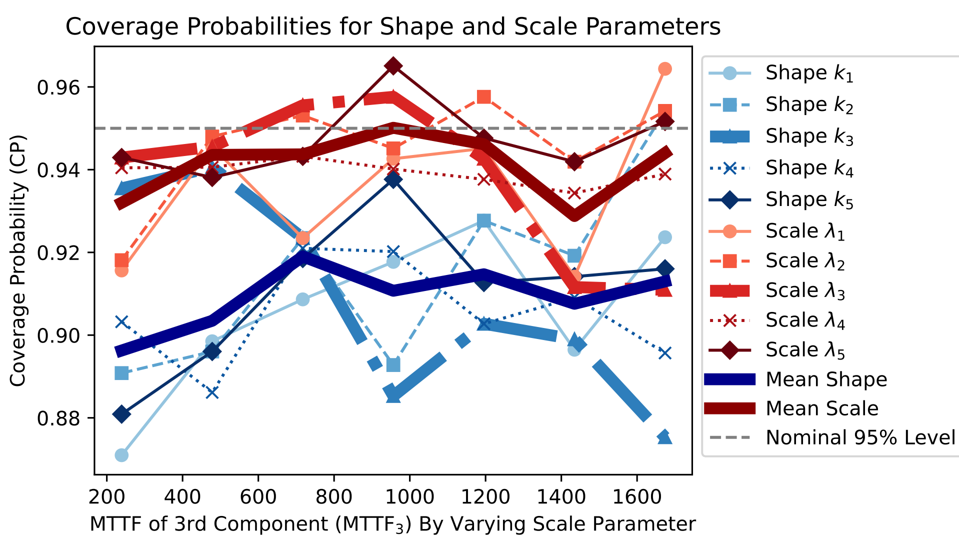
3 Simulation Study: Sensitivity Analysis to Changing System Design

3.1 Scenario: Assessing the Impact of Changing the Scale Parameter of Component 3

By Equation (**??**), we see that MTTF_j is proportional to the scale parameter λ_j, which means when we decrease the scale parameter of a component, we proportionally decrease the MTTF. In this scenario, we start with the well-designed series system described in Table **??**, and we will manipulate the MTTF of component 3, MTTF_3, by changing its scale parameter, λ_3, and observing the effect this has on the MLE. Since the other components had a similiar MTTF, we will arbitrarily choose component 1 to represent the other components. The bottom plot shows the coverage probabilities for all parameters.

In Figure **??**, we show the effect of changing the scale parameter of component 3, $lambda_3$, but map λ_3 to MTTF_3 to make it more intuitive to reason about. We vary the MTTF of component 3 from 300 to 1500 and the other components have their MTTFs fixed at around 900, as shown in Table **??**. We fix the masking probability to $p = 0.215$ (moderate masking), the right-censoring quantile to $q = 0.825$ (moderate censoring), and the sample size to $n = 100$ (moderate sample size).


MLE for k_1
MLE for λ_1
MLE for k_3
MLE for λ_3
IQR of CIs
Mean of MLEs
True Parameter Value
95% Quantile Range
Statistics for k_1
Statistics for λ_1
Statistics for k_3
Statistics for λ_3
239.12
478.25
717.37
956.49
1195.62
1434.74
1673.86
MTTF of 3rd Component (MTTF$_3$)


Coverage Probabilities for Shape and Scale Parameters
Coverage Probability (CP)
MTTF of 3rd Component (MTTF$_3$) By Varying Scale Parameter
Shape k_1
Shape k_2
Shape k_3
Shape k_4
Shape k_5
Scale λ_1
Scale λ_2
Scale λ_3
Scale λ_4
Scale λ_5
Mean Shape
Mean Scale
Nominal 95% Level

3.1.1 Key Observations

Coverage Probability (CP) When MTTF of component 3 is much smaller than other components, the CP for k_3 is very well calibrated (approximately obtaining the nominal level 95%) while the CP for other componentns are around 90%, which is still reasonable. (This is the case even though the width of the CI for k_3 is extremely narrow compared to the others). As MTTF_3 increases, the CP for k_3 decreases, while the CP for the other components increase slightly. The scale parameters are generally well-calibrated for all of the components, except for component 3 when its MTTF is large and it dips down to 90%. Despite the individual differences, the mean of the CPs for shape and scale parameters hardly change.

Dispersion of MLEs For component 3, as its MTTF decreases, the dispersion of MLEs narrows, indicating more precise estimates. Conversely, dispersion for other components widens. As MTTF of component 3 increases, its dispersion widens while others narrow. This is consistent with the fact that the smaller MTTF of component 3 means that, in this well-designed system at least, it is more likely to be the component cause of failure, and so we have more information about its parameters and are able to estimate them more accurately.

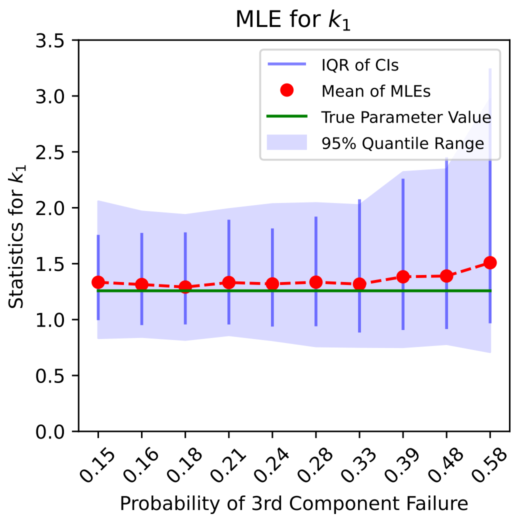
IQR of Bootstrapped CIs The dark blue vertical lines representing IQR are consistent with the dispersion of MLEs, which is the ideal behavior, and suggests that the BCa confidence intervals are performing well.

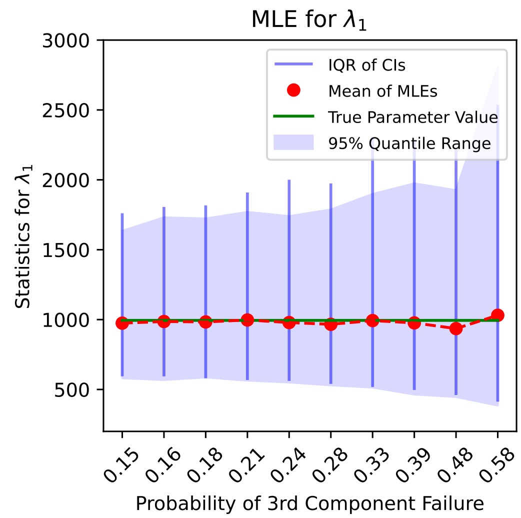
Bias of MLEs For component 3, the bias of MLE for the scale parameter becomes slightly more negatively biased as MTTF_3 increases, and the bias of the MLE for the shape parameter becomes slightly more positively biased. The MLE for the shape and scale parameters for component 1 have a very small bias, if any, and are not affected by the MTTF_3. The scale parameters are easier to estimate than the shape parameters, and so they are less sensitive to changes in scale than the shape parameters, as we will show in the next scenario.

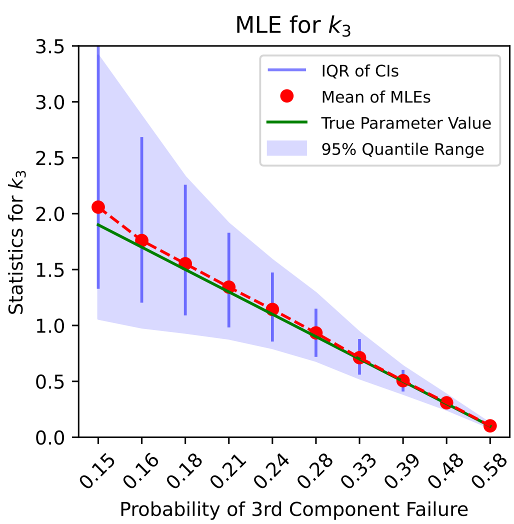
3.2 Scenario: Assessing the Impact of Changing the Shape Parameter of Component 3

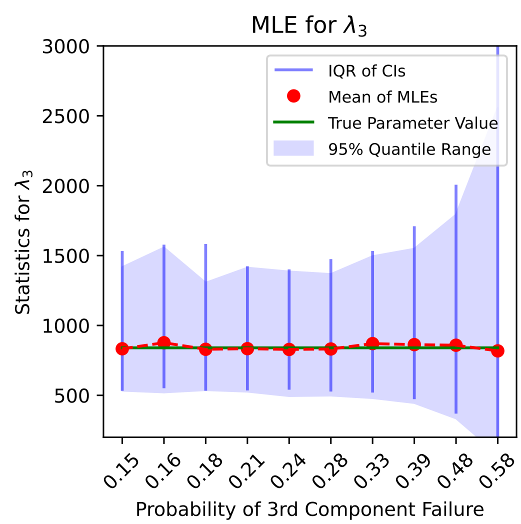
The shape parameter determines the failure characteristics. We vary the shape paramenter of component 3 from 0.1 to 3.5 and observe the effect it has on the MLE. When $k_3 < 1$, this indicates infant mortality, and when $k_3 > 1$, this indicates wear-out failures.

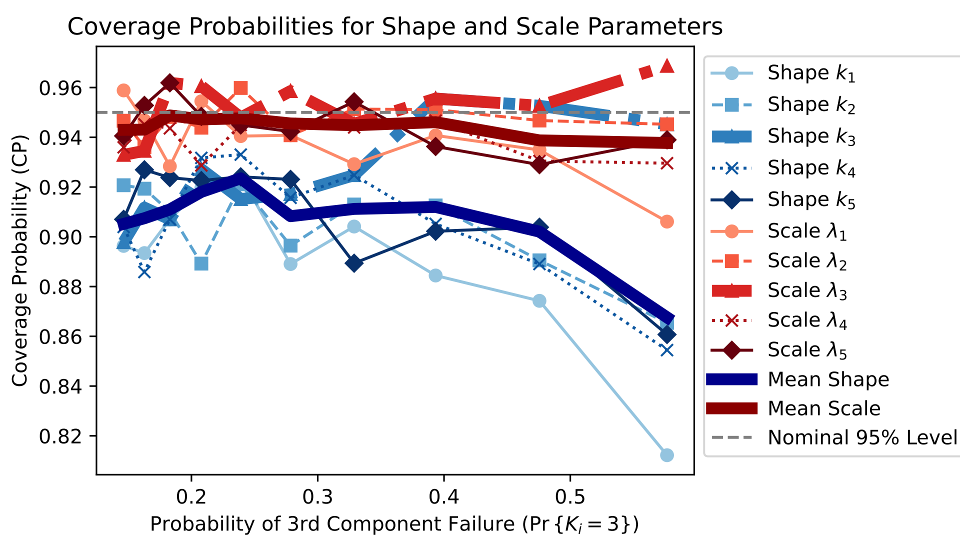
We analyze the effect of component 3's shape parameter on the MLE and the bootstrapped confidence intervals for the shape and scale parameters of components 1 and 3 (the component we are varying). First, we look at the effect on the scale parameter.


MLE for k_1
Statistics for k_1
IQR of CIs
Mean of MLEs
True Parameter Value
95% Quantile Range
0.15 0.16 0.18 0.21 0.24 0.28 0.33 0.39 0.48 0.58
Probability of 3rd Component Failure


MLE for λ_1
Statistics for λ_1
IQR of CIs
Mean of MLEs
True Parameter Value
95% Quantile Range
0.15 0.16 0.18 0.21 0.24 0.28 0.33 0.39 0.48 0.58
Probability of 3rd Component Failure


MLE for k_3
Statistics for k_3
IQR of CIs
Mean of MLEs
True Parameter Value
95% Quantile Range
0.15 0.16 0.18 0.21 0.24 0.28 0.33 0.39 0.48 0.58
Probability of 3rd Component Failure


MLE for λ_3
Statistics for λ_3
IQR of CIs
Mean of MLEs
True Parameter Value
95% Quantile Range
0.15 0.16 0.18 0.21 0.24 0.28 0.33 0.39 0.48 0.58
Probability of 3rd Component Failure


Coverage Probabilities for Shape and Scale Parameters
Coverage Probability (CP)
Shape k_1
Shape k_2
Shape k_3
Shape k_4
Shape k_5
Scale λ_1
Scale λ_2
Scale λ_3
Scale λ_4
Scale λ_5
Mean Shape
Mean Scale
Nominal 95% Level
Probability of 3rd Component Failure ($\Pr\{K_i = 3\}$)

3.2.1 Key Observations

Coverage Probability (CP) The CP for the scale parameters are well-calibrated and close to the nominal level of 0.95 for all values of $\Pr\{K_i = 3\}$. For the the shape parameter of component 3 (k_3) in bold orange colors, we see that it is well-calibrated for all values of $\Pr\{K_i = 3\}$, but actually may become too large for extreme values of $\Pr\{K_i = 3\}$. The CP for the shape parameters of the other components decreases with Pr\{K_i = 3\}$, dipping below 90% for $\Pr\{K_i = 3\} > 0.4$. At a sample size of $n = 100$, the CP for the shape parameters of the other components is generally not well-calibrated for $\Pr\{K_i = 3\} > 0.4$.

Dispersion of MLEs The dispersion of the MLE for the shape and scale parameters of component 1, k_1 and λ_1, is fairly steady but begins to increase rapdily at the extreme values of $\Pr\{K_i = 3\}$. This is indicative of having less information about the failure characteristcs of component 1 as component 3 begins to dominate the component cause of failure. The dispersion of the shape parameter k_3 is initially quite large, indicative of having very little information about the failure characteristcs of component 3 since it is unlikely to be the component cause of failure, but its dispersion rapidly decreases as $\Pr\{K_i = 3\}$ increases and more information is available about component 3's failure characteristics. In fact, it nearly becomes a point at $\Pr\{K_i = 3\} = 0.6$. The dispersion of the the scale parameter of component 3, λ_1, is quite steady and is less spread out than the MLE for λ_1, but at extreme values of $\Pr\{K_i = 3\}$, it also begins to rapidly increase, suggesting some complex interactions between the shape and scale parameters of component 3.

IQR of Bootstrapped CIs The CIs precisely track the dispersion of the MLEs, which is the ideal behavior, and suggests that the BCa confidence intervals are performing well.

Bias of MLEs The MLE for the scale parameters are nearly unbiased and generally seem unaffected by changes in $\Pr\{K_i = 3\}$. As $\Pr\{K_i = 3\}$ increases the MLE is adjusting k_1 to be more positively biased, decreasing its infant morality rate to make it less likely to be the component cause of failure, and adjusting k_3 to be less positively biased, increasing its infant mortality rate, to make it more likely to be the component cause of failure.

4 Weibull Series Homogenous Shape Model

In the sensitivity analysis in Section 3, we found that the MLE is very sensitive to deviations from a well-designed system. In this section, we develop a reduced model that assumes homogeneity in the shape parameters of the components, which simplifies analysis and reduces estimator variability.

Here, our focus shifts to a sensitivity analysis aimed at understanding when it is appropriate to use the reduced model that assumes homogeneity in the shape parameters of the components. The reduced model offers interpretability (the series system is itself Weibull) and reduced estimator variability (only $m + 1$ parameters instead of $2m$), but it is must adequately describe the data.

Define homogenous shape model here... pull from other paper.

4.1 Assessing the Appropriateness of the Reduced Model

In order to determine if a reduced model (e.g., Weibull series system in which all of the shape parameters are homogeneous) is appropriate, a hypothesis test test may be conducted to determine if there is statistically significant evidence in support of the null hypothesis H_0, e.g., that all of the shape parameters are homogeneous.

The likelihood function of the reduced model is related to the likelihood function of the full model. We denote the full model likelihood function by L_F and the reduced model likelihood by L_R. The reduced model is obtained by setting the shape parameter of each component to be the same, i.e., $k_1 = \cdots = k_m = k$. Thus, the reduced model likelihood function is given by

$$L_R(k, \lambda_1, \lambda_2, \cdots, \lambda_m | D) = L_F(k, \lambda_1, k, \lambda_2, \ldots, k, \lambda_m | D),$$

The same may be done for the score and hessian of the log-likelihood functions.

Given that we employ a well-defined likelihood model, the likelihood ratio test (LRT) is a good choice. The LRT statistic is given by

$$\Lambda = -2(\log L_R(\hat{\theta}_R|D) - \log L_F(\hat{\theta}|D))$$

where L_R is the likelihood of the reduced (null) model evaluated at its MLE $\hat{\theta}_R$ given a random sample D of masked data and L_F is the likelihood of the full model evaluated at its MLE $\hat{\theta}$ given the same set of data D. Under the null model, the LRT statistic is asymptotically distributed chi-squared with $m-1$ degrees of freedom, where m is the number of components in the series system,

$$\Lambda \sim \chi^2_{m-1}.$$

If the LRT statistic is greater than the critical value of the chi-squared distribution with $m-1$ degrees of freedom, $\chi^2_{m-1,1-\alpha}$, where α denotes the significance level, then we find the data to be incompatible with the null hypothesis H_0.

4.2 Simulation Study: Full Weibull Model vs Reduced (Homogenous Shape) Model

We aim to assess the appropriateness of the reduced model under varying sample sizes and shape parameters of the third component (k_3). We employ a simulation study using the likelihood ratio test (LRT) for this purpose, where the null hypothesis, H_0, assumes homogenous shape parameters.

We take the well-designed series system described in Table **??**, and manipulate the shape parameter of the third component (k_3) to cause the components to have different failure characteristics. Recall that $k_3 = 1.1308$ corresponds to a *well-designed* series system, where component shapes are reasoanbly aligned. We also vary the sample size n to assess the impact of sample size on the appropriateness of the reduced model.

Figure 1 provides a contour plot with varying sample sizes along the x-axis, shape of component 3 along the y-axis, and median p-value along the color scale. The contour lines corresponding to p-values of 0.05 and 0.1 are often used as a threshold for statistical significance. Points outside of these contours in dark blue are indicative of significant evidence against the null model and points inside of these contours in light blue for 0.1 and green for 0.05 are indicative of insufficient evidence against the null model.

Figure 2 provides a plot of the median p-value against the sample size for the well-designed system, where the shape parameter of component 3 is fixed at 1.1308. The 95th percentile of the p-values is also provided as a more stringent criterion for statistical significance.

Sensitivity to Sample Size (n)

- The sample size is an essential aspect of hypothesis testing, as it affects the test's power, which is the probability of correctly rejecting the null hypothesis when it is false. In the contour plot in Figure 1, as n increases, the contours trend lower. This indicates that larger samples result in smaller median p-values, implying that the power of the LRT increases with the sample size. However, its power is quite low for small samples, particularly for values of k_3 somewhat close to the shape parameters of the other components in the system.

- Recall that in the well-designed series system, $k_3 = 1.1308$. In this case, even very large sample sizes do not produce evidence against the null model, indicating robust compatibility.

- In Figure 2, we fix k_3 at 1.1308 and vary the sample size. The median p-value only manages to drop below the 0.05 theshold with sample sizes around 10000. In the more stringent criterion given by the 95th percentile of the p-values, nearly 30000 observations are necessary to reject the null hypothesis in 95% of the simulations.

Sensivity to Shape Parameter (k_3)

- In Figure 1, for a given shape parameter, increasing the sample size tends to decrease the median p-value. Larger samples provide more information about the parameters, which increases the power of the LRT.

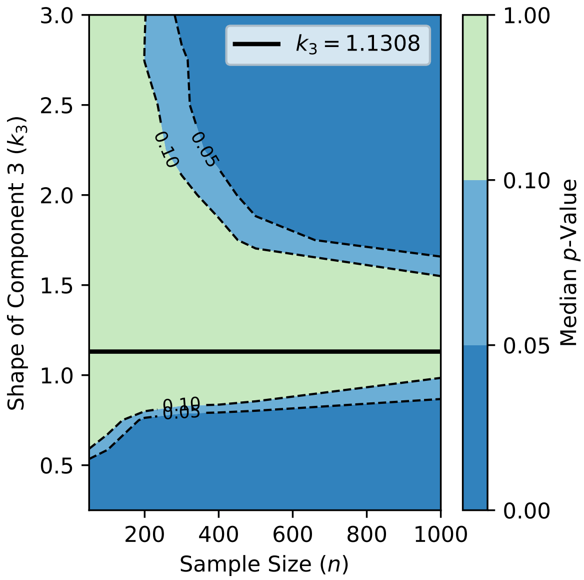


Figure 1: p-Value vs Sample Size and Shape k_3

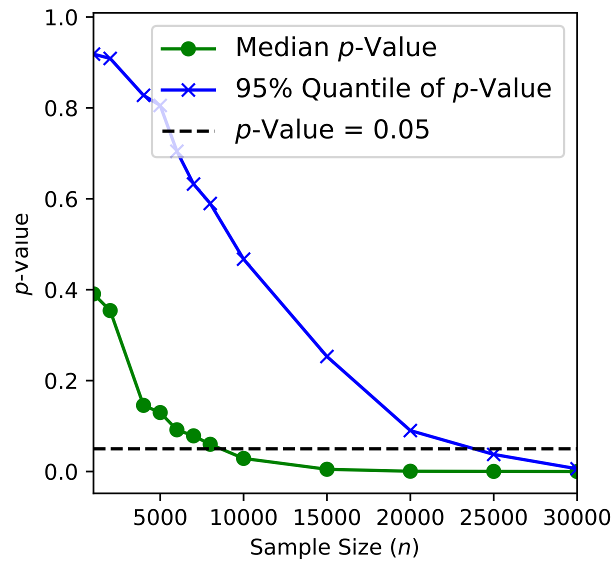


Figure 2: p-Value vs Sample Size for Well-Designed System

- The median p-values in the vicinity of the line $k_3 = 1.1308$ are high across various sample sizes, indicating that the null model is a good fit. As k_3 deviates from this line, the median p-value diminishes, indicating increasing evidence against the null model.

4.3 Implications and Recommendations

The power of the test for a well-designed series system is quite low, requiring many thousands of observations before the test has sufficient power to reject the null hypothesis. But, this is not necessarily a bad thing. The reduced model is quite simple and interpretable, and is by definition a good fit for a well-designed series system.

The findings suggest that the reduced model is particularly apt when the system is well-designed, even for very large samples. Practitioners should weigh the trade-offs between the simplicity of the reduced model and the adequacy in describing the data, with consideration of the available sample size and the characteristics of the system being modeled.

For systems believed to be well-designed, employing the null model is supported both statistically and practically due to its simplicity, reduced estimator variability, and analytical tractability. In the absence of prior information, or if the shape parameter significantly diverges from the well-designed value, the choice between models should be undertaken with caution. More complex models may be favorable, especially with large sample sizes.

5 Conclusion

In this study, we employed simulation techniques and Likelihood Ratio Tests (LRTs) to assess the adequacy and sensitivity of Maximum Likelihood Estimators (MLEs) for reliability assessment in 5-component series systems with Weibull component lifetimes. Two main models were examined: a more complex model with heterogeneous shape parameters and a reduced model assuming homogeneous shape parameters for all components.

The reduced model improves interpretability by rendering the system Weibull and reduces estimator variability, thus appearing statistically and practically favorable for well-designed systems. These well-designed systems are characterized by similar but non-identical failure characteristics among components, without a single weak point. Even for large samples, the reduced model showed excellent fit in these cases.

However, our simulations revealed that varying a single component's scale or shape parameter quickly provided evidence against the reduced model's adequacy. This suggests that more complex models may be preferable in systems with divergent component properties or when sample sizes are large.

Estimator performance was found to be sensitive but robust, particularly concerning the challenges introduced by limited, right-censored, and masked failure data. Bootstrap confidence intervals proved valuable in characterizing estimator uncertainty. As the sample size increased, estimator dispersion reduced, and confidence intervals narrowed, although small samples exhibited bias, particularly in shape parameters. Masking probability also played a role, expanding confidence intervals to maintain coverage while largely leaving scale parameters unbiased.

In summary, the choice between the reduced and more complex models should weigh the trade-offs between simplicity and representativeness. Our findings offer practical guidance for reliability assessments, particularly when dealing with limited system failure data. Proper model specification ultimately requires a nuanced understanding of both system characteristics and estimator behavior.