Sensitivity of Series System Reliability Estimation to the Non-Informative Masking Assumption

Alex Towell
lex@metafunctor.com

February 28, 2026

Abstract

Maximum likelihood estimation for series systems with masked failure data relies on the non-informative masking assumption (Condition C2): that the candidate set reported for a failed component is independent of the true cause of failure. This assumption is rarely verifiable and frequently suspect. We characterize the bias that arises when C2 is violated but assumed to hold, proving that the MLE converges to a pseudo-true parameter whose individual component parameters are biased while the system hazard remains consistent. Using a Bernoulli masking model that generates controlled C2 violations of varying severity, we conduct a systematic simulation sweep on a 5-component Weibull series system. The results establish a practical breakdown boundary: system-level reliability estimates are robust to C2 violations across the full severity range, but individual component estimates degrade with severity, with scale parameters showing order-of-magnitude bias under strong informative masking. Since the masking mechanism and component parameters are jointly non-identifiable, we recommend sensitivity analysis over model elaboration. Software is available in the `mdrelax` R package.

Keywords: Series systems, masked failure data, non-informative masking, model misspecification, sensitivity analysis, maximum likelihood estimation

1 Introduction

Estimating the reliability of individual components in a series system is a fundamental challenge when failure causes are masked. When a series system fails, diagnostic procedures may identify only a *candidate set* of components that could have caused the failure, rather than pinpointing the exact failed component [7]. Maximum likelihood estimation under masked data has been developed extensively under a standard set of assumptions [6, 8]: (C1) the true failed component is always in the candidate set; (C2) masking is non-informative, meaning the candidate set formation is independent of the failure cause given the failure time; and (C3) the masking mechanism does not depend on the component reliability parameters.

Among these, C2 is the strongest and least verifiable assumption. It requires that the diagnostic process which generates candidate sets reveals nothing about which component actually failed beyond the guarantee of C1. In practice, the masking mechanism reflects a complex diagnostic process—technician experience, test equipment sensitivity, failure mode characteristics—that is rarely known and seldom non-informative. An experienced technician may preferentially include components whose failure signatures match the observed failure time; an automated diagnostic may weight components by their estimated reliability ranking, effectively violating C2.

Prior work has addressed dependent masking by proposing alternative models for the masking probability. Lin and Guess [4] introduced proportional masking probabilities for two-component systems with closed-form maximum likelihood estimators. Guttman et al. [3] developed Bayesian inference under dependent masking for two-component systems. Mukhopadhyay and Basu [5] extended likelihood-based inference to general m-component systems where masking depends on the cause of failure. Craiu and Reiser [1] studied conditional masking probability models and their identifiability. These approaches share a common structure: they propose a specific masking model to replace C2 and estimate its parameters jointly with the component reliability parameters.

We take a complementary approach. Rather than proposing an alternative masking model, we study the *sensitivity* of the standard C1-C2-C3 estimator to violations of C2. The question is practical: when a reliability engineer fits the standard model—as is common practice—how wrong are the resulting estimates when C2 fails? Our contributions are:

1. **Misspecification characterization.** We prove that when C2 is violated but assumed to hold, the MLE converges to a pseudo-true parameter whose individual component rates absorb the masking asymmetry. The total system hazard, however, remains consistently estimated (Section 3).

2. **Simulation sweep.** Using a Bernoulli perturbation model that generates controlled C2 violations of varying severity, we map the bias, RMSE, and coverage degradation as a function of violation severity for a 5-component Weibull series system (Section 4).

3. **Practical finding.** System-level reliability estimates (total hazard, mean time to failure) are robust to C2 violations across the full severity range. Individual component parameter estimates are not. Since the masking mechanism and component rates are jointly non-identifiable, sensitivity analysis—not model elaboration—is the appropriate response (Section 5).

The remainder of this paper is organized as follows. Section 2 reviews series systems, masked data, and the standard C1-C2-C3 likelihood, and surveys prior work on dependent masking. Section 3 develops the sensitivity framework: the general likelihood under C1 alone, the Bernoulli perturbation model, the misspecification theorem, and the identifiability result. Section 4 presents the simulation sweep. Section 5 discusses practical implications, and Section 6 concludes.

2 Background

2.1 Series System Model

Consider a series system composed of m components. The lifetime of the i-th system is

$$T_i = \min\{T_{i1}, T_{i2}, \ldots, T_{im}\}, \tag{1}$$

where T_{ij} denotes the lifetime of the j-th component in the i-th system. Component lifetimes are assumed independent with parametric distributions indexed by $\boldsymbol{\theta}_j$; the full parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m) \in \boldsymbol{\Omega}$.

For component j with parameter $\boldsymbol{\theta}_j$, the reliability, density, and hazard functions are

$$R_j(t; \boldsymbol{\theta}_j) = \Pr\{T_{ij} > t\}, \quad f_j(t; \boldsymbol{\theta}_j) = -R'_j(t; \boldsymbol{\theta}_j), \quad h_j(t; \boldsymbol{\theta}_j) = \frac{f_j(t; \boldsymbol{\theta}_j)}{R_j(t; \boldsymbol{\theta}_j)}. \tag{2}$$

Independence yields the series system hazard $h_T(t; \boldsymbol{\theta}) = \sum_{j=1}^{m} h_j(t; \boldsymbol{\theta}_j)$ and survival $R_T(t; \boldsymbol{\theta}) = \prod_{j=1}^{m} R_j(t; \boldsymbol{\theta}_j)$.

Given that the system failed at time t, the conditional probability that component j caused the failure is

$$\Pr\{K_i = j \mid T_i = t\} = \frac{h_j(t;\boldsymbol{\theta}_j)}{\sum_{\ell=1}^m h_\ell(t;\boldsymbol{\theta}_\ell)}, \tag{3}$$

which follows from the joint density $f_{T,K}(t,k) = h_k(t;\boldsymbol{\theta}_k)\, R_T(t;\boldsymbol{\theta})$ and Bayes' theorem [6].

2.2 Masked Data

For each system i, we observe:

- $S_i = \min\{T_i, \tau_i\}$: right-censored system lifetime,
- $\delta_i = \mathbf{1}_{T_i \le \tau_i}$: event indicator (1 if failure observed),
- $C_i \subseteq \{1, \ldots, m\}$: candidate set (only meaningful when $\delta_i = 1$).

The true cause of failure K_i and the individual component lifetimes $(T_{i1}, \ldots, T_{im})$ are latent.

2.3 Conditions C1, C2, C3

Tractable likelihood-based inference requires conditions on the relationship between the latent cause K_i and the observed candidate set C_i [2, 7, 8]:

Condition 1 (C1: Failed Component in Candidate Set)**.** $\Pr\{K_i \in C_i\} = 1$.

Condition 2 (C2: Non-Informative Masking)**.** *For all $j, j' \in c$:*

$$\Pr\{C_i = c \mid T_i = t, K_i = j\} = \Pr\{C_i = c \mid T_i = t, K_i = j'\}. \tag{4}$$

Condition 3 (C3: Parameter-Independent Masking)**.** $\Pr\{C_i = c \mid T_i = t, K_i = j\}$ *does not depend on $\boldsymbol{\theta}$.*

C1 is a minimal correctness requirement. C3 is often reasonable when the diagnostic process is fixed independently of the reliability parameters. C2 is the strongest assumption: it requires that the diagnostic process which generates candidate sets reveals nothing about the failure cause beyond C1. This paper studies the consequences of C2 failure.

2.4 Likelihood Under C1-C2-C3

Theorem 2.1 (Likelihood Under C1-C2-C3 [6])**.** *Under Conditions C1, C2, and C3, the likelihood contribution from observation i is*

$$L_i(\boldsymbol{\theta}) \propto \prod_{\ell=1}^m R_\ell(s_i;\boldsymbol{\theta}_\ell) \cdot \left[\sum_{k\in c_i} h_k(s_i;\boldsymbol{\theta}_k)\right]^{\delta_i}. \tag{5}$$

The masking probability factors out and cancels from the likelihood, leaving a tractable expression for MLE. The complete log-likelihood for n independent systems is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \left[\sum_{j=1}^m \log R_j(s_i;\boldsymbol{\theta}_j) + \delta_i \log\left(\sum_{k\in c_i} h_k(s_i;\boldsymbol{\theta}_k)\right)\right]. \tag{6}$$

Exponential specialization. When each component has exponential lifetime $T_{ij} \sim \mathrm{Exp}(\theta_j)$ with rate $\theta_j > 0$, we have $h_j(t) = \theta_j$ and $R_j(t) = e^{-\theta_j t}$, giving

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[-s_i \sum_{j=1}^{m} \theta_j + \delta_i \log \left(\sum_{k \in c_i} \theta_k \right) \right]. \tag{7}$$

Weibull specialization. When each component has Weibull lifetime with shape k_j and scale λ_j, the hazard is $h_j(t) = (k_j/\lambda_j)(t/\lambda_j)^{k_j-1}$ and $R_j(t) = \exp[-(t/\lambda_j)^{k_j}]$. The log-likelihood becomes

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[-\sum_{j=1}^{m} \left(\frac{s_i}{\lambda_j} \right)^{k_j} + \delta_i \log \left(\sum_{k \in c_i} h_k(s_i) \right) \right], \tag{8}$$

where $\boldsymbol{\theta} = (k_1, \lambda_1, \ldots, k_m, \lambda_m)$.

2.5 Prior Work on Dependent Masking

Several authors have proposed models that relax C2 by explicitly specifying a masking mechanism that depends on the failure cause.

Lin and Guess [4] introduced proportional masking probabilities for two-component systems, deriving closed-form MLEs under the assumption that the masking probability for the true cause exceeds that for the non-failed component by a fixed ratio. Guttman et al. [3] extended two-component dependent masking to a Bayesian framework with conjugate priors. Mukhopadhyay and Basu [5] developed EM-based maximum likelihood estimation for general m-component systems where masking probabilities depend on the cause of failure, establishing conditions for identifiability of the joint parameter. Craiu and Reiser [1] studied conditional masking probability models and proved identifiability results under specific structural assumptions on the masking mechanism.

These approaches share a common strategy: they replace C2 with a specific parametric masking model and estimate its parameters jointly with the component reliability parameters. The implicit assumption is that the masking model is correctly specified. However, the masking mechanism is itself a diagnostic process that is rarely known in detail. A natural complementary question—one not addressed in the prior literature—is: *how sensitive is the standard C1-C2-C3 estimator to violations of C2 when no alternative masking model is adopted?* This is the question we study.

3 Sensitivity Framework

We now develop the tools needed to study the sensitivity of C1-C2-C3 inference to C2 violations. We proceed in four steps: the general likelihood under C1 alone (Section 3.1), a Bernoulli model for generating controlled C2 violations (Section 3.2), a misspecification theorem characterizing the resulting bias (Section 3.3), and a non-identifiability result that motivates sensitivity analysis over model estimation (Section 3.4).

3.1 Likelihood Under C1 Alone

When only C1 holds, the masking probability enters the likelihood and cannot be eliminated.

Theorem 3.1 (Likelihood Under C1 Alone)**.** *Under Condition C1 alone, the likelihood contribution from an uncensored observation* (s_i, c_i) *is:*

$$L_i(\boldsymbol{\theta}) = \prod_{\ell=1}^{m} R_\ell(s_i; \boldsymbol{\theta}_\ell) \cdot \sum_{k \in c_i} h_k(s_i; \boldsymbol{\theta}_k) \cdot \Pr\{C_i = c_i \mid T_i = s_i, K_i = k\}. \tag{9}$$

Proof. Under C1, $\Pr\{C_i = c \mid T_i = t, K_i = k\} = 0$ when $k \notin c$. Summing over K_i:

$$\begin{aligned} f_{T_i,C_i}(t, c; \boldsymbol{\theta}) &= \sum_{k=1}^{m} h_k(t; \boldsymbol{\theta}_k) \prod_{\ell=1}^{m} R_\ell(t; \boldsymbol{\theta}_\ell) \cdot \Pr\{C_i = c \mid T_i = t, K_i = k\} \\ &= \prod_{\ell=1}^{m} R_\ell(t; \boldsymbol{\theta}_\ell) \cdot \sum_{k \in c} h_k(t; \boldsymbol{\theta}_k) \Pr\{C_i = c \mid T_i = t, K_i = k\}. \end{aligned}$$

□

Under C2, the masking probability $\Pr\{C_i = c \mid T_i = t, K_i = k\}$ is constant over $k \in c$ and factors out of the sum. Under C3, it does not depend on $\boldsymbol{\theta}$ and can be dropped from the likelihood. When either condition fails, the masking probability remains inside the sum, coupling the inference about $\boldsymbol{\theta}$ to the unknown masking mechanism.

3.2 The Bernoulli Perturbation Model

To generate data with controlled C2 violations, we use a Bernoulli candidate set model. We do not claim this model describes how masking works in practice. Rather, it provides a device for producing data whose departure from C2 is known and calibrated.

Definition 3.2 (Bernoulli Masking Model)**.** *Let* $p_j(k) = \Pr\{j \in C_i \mid K_i = k\}$ *denote the probability that component* j *is included in the candidate set given that component* k *failed. Under C1,* $p_j(j) = 1$ *for all* j*. These probabilities form an* $m \times m$ *matrix:*

$$\mathbf{P} = \begin{pmatrix} 1 & p_1(2) & \cdots & p_1(m) \\ p_2(1) & 1 & \cdots & p_2(m) \\ \vdots & \vdots & \ddots & \vdots \\ p_m(1) & p_m(2) & \cdots & 1 \end{pmatrix}, \tag{10}$$

where each component is included independently. The probability of observing candidate set c *given* $K_i = k$ *is*

$$\pi_k(c) = \mathbf{1}_{k \in c} \prod_{j \in c \setminus \{k\}} p_j(k) \prod_{j \notin c} (1 - p_j(k)). \tag{11}$$

Remark 3.1 (C2 in Terms of $\mathbf{P}$). Condition C2 holds if and only if each row of $\mathbf{P}$ has constant off-diagonal entries: $p_j(k) = p_j$ for all $k \neq j$. When the columns of $\mathbf{P}$ differ, the masking mechanism "knows" which component failed, violating C2.

The likelihood under this model follows immediately from Theorem 3.1:

Corollary 3.3 (Likelihood Under the Bernoulli Model)**.** *Under the Bernoulli masking model with known* $\mathbf{P}$*, the likelihood contribution from an uncensored observation* (s_i, c_i) *is:*

$$L_i(\boldsymbol{\theta}) = \prod_{\ell=1}^{m} R_\ell(s_i; \boldsymbol{\theta}_\ell) \cdot \sum_{k \in c_i} h_k(s_i; \boldsymbol{\theta}_k) \cdot \pi_k(c_i). \tag{12}$$

Parameterizing violation severity. We generate a family of $\mathbf{P}$ matrices indexed by a scalar severity parameter $s \in [0, 1]$:

$$\mathbf{P}(s) = \mathbf{P}_0 + s\,\mathbf{D}, \tag{13}$$

where $\mathbf{P}_0$ is the uniform (C2-satisfying) matrix with all off-diagonal entries equal to a base probability p_0, and $\mathbf{D}$ is a fixed "direction" matrix with zero diagonal, whose off-diagonal entries create column-wise asymmetry. Off-diagonal entries of $\mathbf{P}(s)$ are clamped to $[0.05, 0.95]$ to maintain valid probabilities. At $s = 0$, C2 holds exactly; as s increases, the columns of $\mathbf{P}$ diverge and the masking becomes increasingly informative.

3.3 Misspecification Under C2 Violation

We now characterize what happens when the standard C1-C2-C3 model is fitted to data generated under C2 violation.

Theorem 3.4 (Bias from C2 Misspecification)**.** *Suppose data is generated under C1 and C3 with informative masking weights $\pi_{k,c}$, but estimation is performed under C1-C2-C3 (assuming non-informative masking). The MLE under the misspecified model converges in probability to a pseudo-true parameter $\boldsymbol{\theta}^\dagger$ satisfying*

$$\mathbb{E}_{\boldsymbol{\theta}^*}\left[\left.\frac{\partial \ell^{\text{wrong}}}{\partial \theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\dagger}\right] = 0, \tag{14}$$

where the expectation is taken under the true data-generating process with parameter $\boldsymbol{\theta}^*$. The pseudo-true parameter differs from $\boldsymbol{\theta}^*$ unless C2 holds.

For exponential components with rates $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$:

(a) The total system hazard is approximately preserved: $\sum_j \theta_j^\dagger \approx \sum_j \theta_j^*$.

(b) Individual rates absorb the masking asymmetry: components that are over-represented in candidate sets (due to asymmetric $\mathbf{P}$) have inflated $\theta_j^\dagger$, while under-represented components have deflated estimates.

Proof. The misspecified score for exponential component j is

$$\frac{\partial \ell_i^{\text{wrong}}}{\partial \theta_j} = -s_i + \delta_i \frac{\mathbf{1}_{j \in c_i}}{\sum_{k \in c_i} \theta_k}. \tag{15}$$

The true score under C1-C3 (with masking weights $\pi_{k,c}$) is

$$\frac{\partial \ell_i^{\text{true}}}{\partial \theta_j} = -s_i + \delta_i \frac{\pi_{j,c_i}\,\mathbf{1}_{j \in c_i}}{\sum_{k \in c_i} \theta_k\,\pi_{k,c_i}}. \tag{16}$$

At $\boldsymbol{\theta}^*$, the true score has expectation zero. The misspecified score replaces $\pi_{k,c}$ with uniform weights $1/|c|$. Setting $\mathbb{E}[\partial \ell^{\text{wrong}}/\partial \theta_j] = 0$ at $\boldsymbol{\theta}^\dagger$ yields a system whose solution differs from $\boldsymbol{\theta}^*$ whenever the masking weights $\pi_{k,c}$ vary with k.

For part (a), summing the misspecified score over all j yields the total hazard score

$$\sum_j \frac{\partial \ell_i^{\text{wrong}}}{\partial \theta_j} = -m s_i + \delta_i \frac{|c_i|}{\sum_{k \in c_i} \theta_k}. \tag{17}$$

The analogous sum for the true score is

$$\sum_j \frac{\partial \ell_i^{\text{true}}}{\partial \theta_j} = -ms_i + \delta_i \frac{\sum_{k\in c_i} \pi_{k,c_i}}{\sum_{k\in c_i} \theta_k\, \pi_{k,c_i}}. \tag{18}$$

When the masking weights satisfy $\sum_{k\in c} \pi_{k,c} / \sum_{k\in c} \theta_k \pi_{k,c} \approx |c| / \sum_{k\in c} \theta_k$, the total hazard score equations agree and $\sum_j \theta_j^\dagger \approx \sum_j \theta_j^*$. This holds exactly when C2 is satisfied and approximately when the masking asymmetry is moderate. □

Remark 3.2 (Interpretation). The misspecified estimator acts as if each candidate in c_i contributes equally to the observed failure, regardless of the actual masking weights. When a component is over-included in candidate sets (its column of $\mathbf{P}$ has higher off-diagonal values), it "claims credit" for more failures than it actually caused, inflating its estimated rate.

3.4 The Identifiability Trap

A natural response to C2 uncertainty would be to estimate $\mathbf{P}$ jointly with $\boldsymbol{\theta}$ from the data. This fails.

Theorem 3.5 (Non-Identifiability of $(\boldsymbol{\theta}, \mathbf{P})$)**.** *For exponential series systems with masked data, the joint parameter $(\boldsymbol{\theta}, \mathbf{P})$ is not identifiable from the observed data $(s_i, c_i, \delta_i)_{i=1}^n$. Different combinations of $(\boldsymbol{\theta}, \mathbf{P})$ can yield equivalent likelihoods.*

However, the total system hazard $\sum_{j=1}^m \theta_j$ remains identifiable from the marginal system lifetime data.

This result is supported by both theoretical argument and simulation evidence. Even with $n = 2000$ observations, the joint MLE of $(\boldsymbol{\theta}, \mathbf{P})$ exhibits persistent bias with individual component rates confounded with the off-diagonal entries of $\mathbf{P}$, while the total hazard $\sum_j \hat{\theta}_j$ converges to the true value (see Appendix B).

Remark 3.3 (Implications). Since $\boldsymbol{\theta}$ and $\mathbf{P}$ cannot be disentangled from the data, we cannot "just estimate" the masking mechanism. This makes sensitivity analysis the appropriate tool: fit the standard C1-C2-C3 model, then assess how conclusions change under plausible perturbations of the masking structure via different $\mathbf{P}$ matrices. If the substantive conclusions are stable, C2 violations are not a concern for the application at hand.

4 Simulation Study

We conduct a systematic sweep of C2 violation severity for a 5-component Weibull series system, fitting the standard C1-C2-C3 model throughout. This quantifies the practical impact of C2 misspecification as a function of the departure from non-informative masking.

4.1 Design

System configuration. We consider an $m = 5$ component Weibull series system with shapes $\mathbf{k}^* = (2.0, 1.5, 1.2, 1.8, 1.0)$ and scales $\boldsymbol{\lambda}^* = (3.0, 4.0, 5.0, 3.5, 4.5)$. The fifth component ($k_5 = 1.0$) is exponential, providing a natural point of comparison within the system. Right-censoring time is $\tau = 5$, sample size $n = 500$, and $B = 200$ replications per severity level. The true system hazard at reference time $t_0 = 2.0$ is $h_T^*(2.0) = 1.460$.

Severity sweep. We sweep the severity parameter s over $\{0, 0.1, 0.2, \dots, 1.0\}$ (11 levels). At each level, data is generated from the Bernoulli model with $\mathbf{P}(s)$ as defined in (13), using base probability $p_0 = 0.5$. The direction matrix has a cyclic structure with rows summing to zero:

$$\mathbf{D} = \begin{pmatrix} 0 & 0.3 & -0.2 & 0.1 & -0.2 \\ -0.2 & 0 & 0.3 & -0.2 & 0.1 \\ 0.1 & -0.2 & 0 & 0.3 & -0.2 \\ -0.2 & 0.1 & -0.2 & 0 & 0.3 \\ 0.3 & -0.2 & 0.1 & -0.2 & 0 \end{pmatrix}. \tag{19}$$

The cyclic structure ensures that each component is equally affected in aggregate, so the bias pattern is driven by the interaction between masking asymmetry and the heterogeneous Weibull parameters rather than the direction matrix itself.

Estimation. At each severity level, all B datasets are fitted with the standard *C1-C2-C3* Weibull series MLE from (8), ignoring the informative masking. Optimization uses L-BFGS-B with starting values at 90% of true parameters. Convergence was 100% across all conditions.

Metrics. For each parameter and severity level, we report:

- **Bias**: $\bar{\hat{\theta}}_j - \theta_j^*$,
- **RMSE**: $\sqrt{B^{-1} \sum_b (\hat{\theta}_j^{(b)} - \theta_j^*)^2}$,
- **Coverage**: proportion of 95% Wald confidence intervals $\hat{\theta}_j \pm 1.96\, \hat{\mathrm{se}}_j$ that contain θ_j^*.

We additionally track the system hazard $\hat{h}_T(t_0)$ evaluated at $t_0 = 2.0$.

4.2 Results

Figures 1 to 3 display the three metrics as a function of severity.

Table 1 provides numerical summaries at selected severity levels for a representative subset of parameters. The full results for all 10 component parameters are shown in the figures.

4.3 Interpretation

System-level robustness. The most striking finding is the stability of the system hazard. Across all severity levels from $s = 0$ to $s = 1.0$, the system hazard bias remains between $+0.012$ and $+0.038$ (at most 2.6% relative error), and the RMSE is virtually constant near 0.10. This confirms part (a) of Theorem 3.4: the misspecified estimator preserves the system hazard even when individual component parameters are severely biased. The robustness holds despite the system having 10 free parameters (five shape-scale pairs) and substantial heterogeneity across components.

Component-level degradation. Individual parameter estimates degrade with severity. The shape parameter k_2 (true value 1.5) shows bias growing from $+0.03$ at $s = 0$ to -0.34 at $s = 1.0$ (23% relative), with coverage dropping from 0.96 to below nominal. The shape parameters k_3 and k_4 show similar degradation, with coverage falling to 0.70 and 0.65 respectively at full severity.

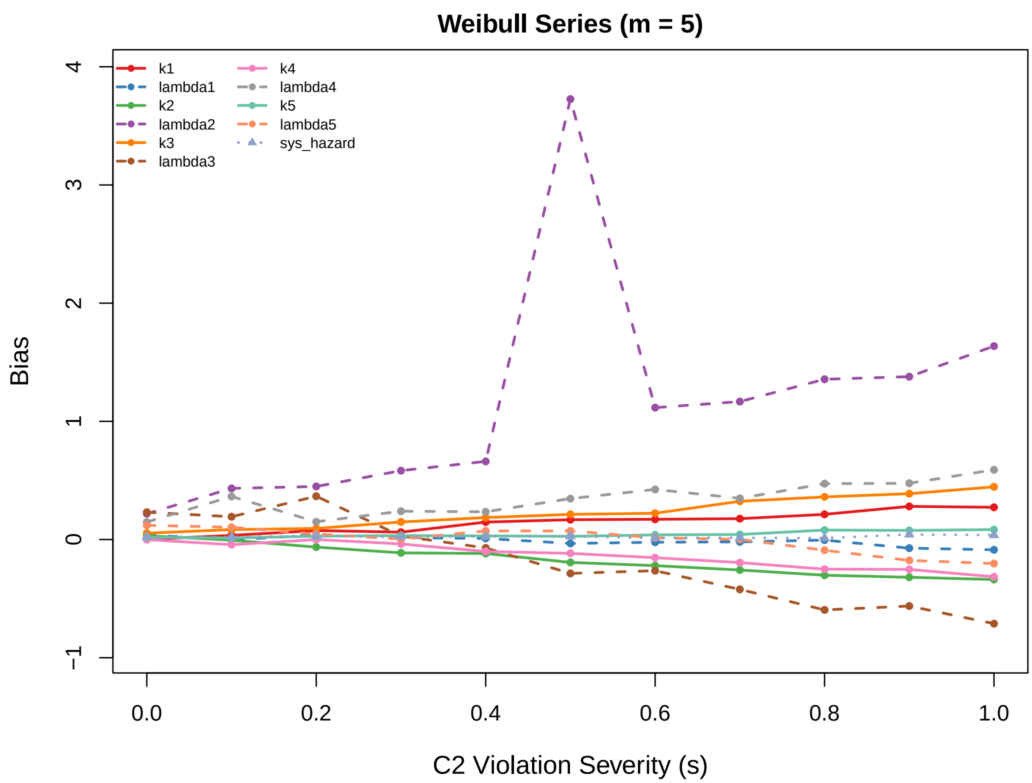

Figure 1: Bias as a function of C2 violation severity for a 5-component Weibull series system. Individual component parameters show increasing bias with severity, while the system hazard (triangle markers) remains approximately unbiased.

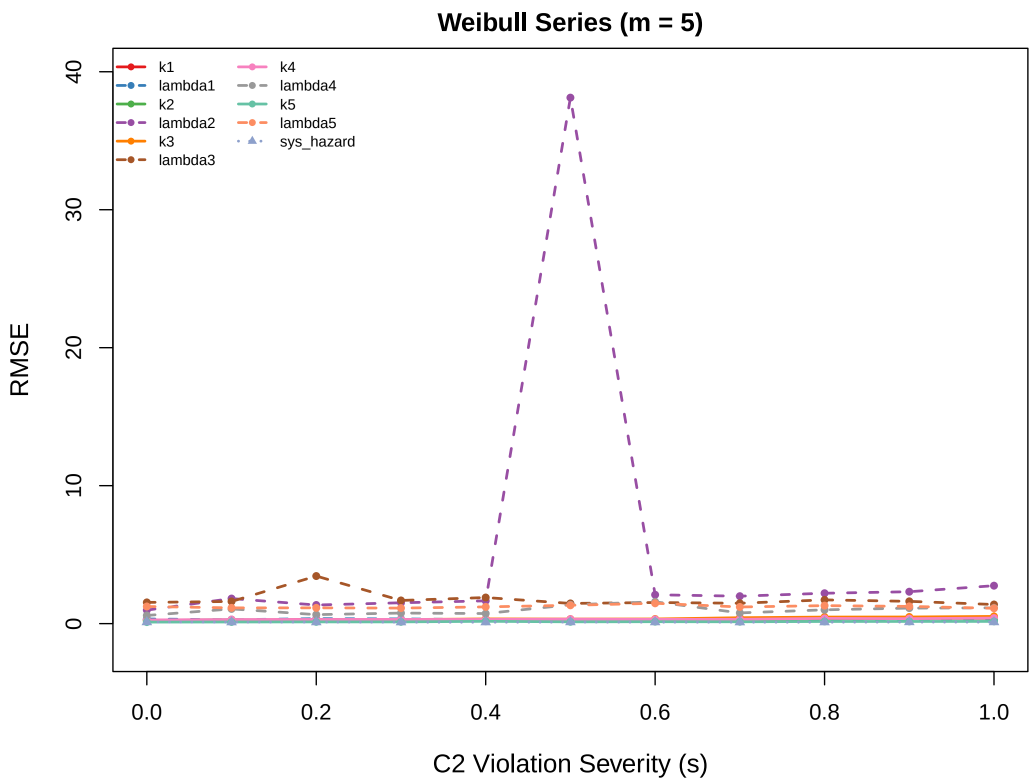

Figure 2: Root mean squared error as a function of C2 violation severity. The RMSE for individual parameters grows with severity; the system hazard RMSE remains stable near 0.10 throughout.

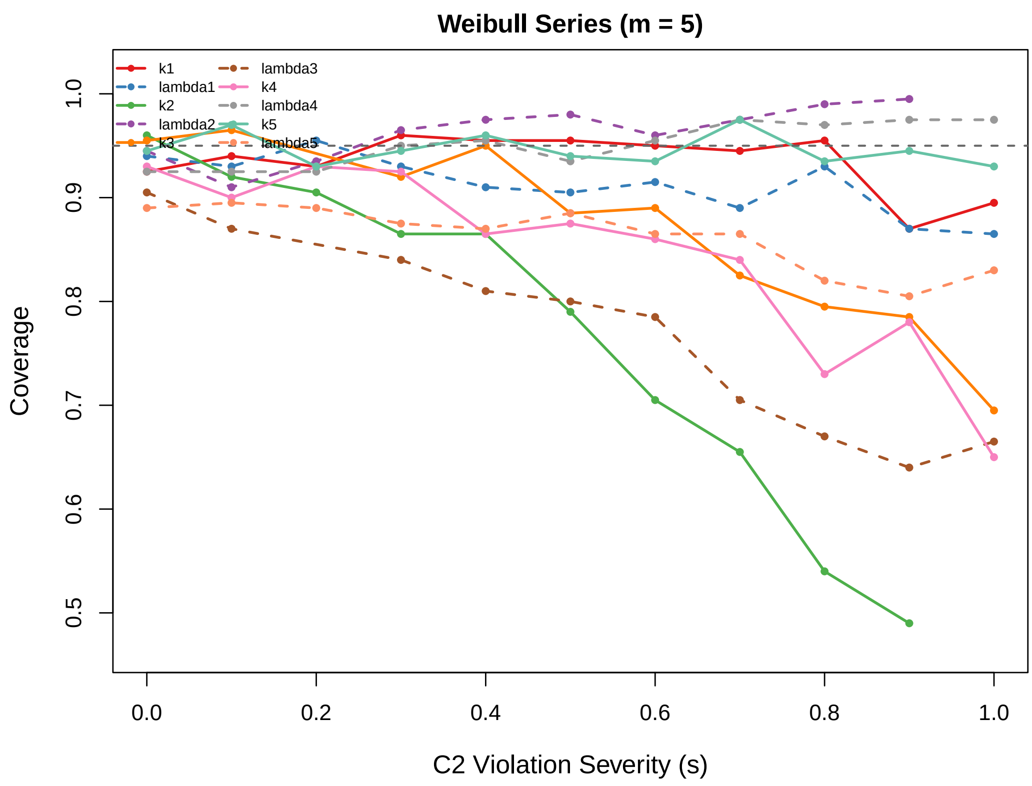

Figure 3: 95% confidence interval coverage as a function of C2 violation severity. The dashed line marks the nominal 95% level. Coverage for the most affected shape parameters (k_2, k_3, k_4) drops below 70% at $s = 1.0$.

Table 1: Weibull sensitivity sweep ($m = 5$, $n = 500$, $B = 200$). Selected parameters illustrating the range of misspecification effects. True system hazard $h_T^*(2.0) = 1.460$.

s	Parameter	Bias	RMSE	Coverage
0.0	k_2 (1.5)	+0.032	0.259	0.960
	λ_3 (5.0)	+0.231	1.538	0.905
	k_4 (1.8)	−0.001	0.268	0.930
	k_5 (1.0)	+0.018	0.111	0.945
	$h_T(2.0)$	+0.024	0.108	—
0.2	k_2 (1.5)	−0.064	0.251	0.905
	λ_3 (5.0)	+0.366	3.449	—
	k_4 (1.8)	+0.000	0.303	0.930
	k_5 (1.0)	+0.028	0.125	0.930
	$h_T(2.0)$	+0.031	0.105	—
0.5	k_2 (1.5)	−0.194	0.305	0.790
	λ_3 (5.0)	−0.287	1.465	0.800
	k_4 (1.8)	−0.117	0.322	0.875
	k_5 (1.0)	+0.027	0.129	0.940
	$h_T(2.0)$	+0.028	0.108	—
0.8	k_2 (1.5)	−0.302	0.346	0.540
	λ_3 (5.0)	−0.596	1.720	0.670
	k_4 (1.8)	−0.251	0.378	0.730
	k_5 (1.0)	+0.080	0.155	0.935
	$h_T(2.0)$	+0.012	0.103	—
1.0	k_2 (1.5)	−0.338	0.384	—
	λ_3 (5.0)	−0.712	1.385	0.665
	k_4 (1.8)	−0.315	0.401	0.650
	k_5 (1.0)	+0.084	0.160	0.930
	$h_T(2.0)$	+0.038	0.109	—

The exponential component. Component 5 ($k_5 = 1.0$, exponential) is notably resistant to misspecification: its shape bias reaches only $+0.08$ at $s = 1.0$ (8% relative) with coverage of 0.93, near the nominal level. This suggests that the exponential case is a natural "robustness baseline" within the Weibull family, consistent with the theoretical analysis in Section 3.3 where the exponential score is simpler and less susceptible to masking weight distortion.

The breakdown boundary. For practitioners concerned with individual component estimates, the simulation results suggest a rough breakdown boundary around $s \approx 0.3$: below this threshold, component-level biases remain within typical estimation uncertainty and coverage stays above 90%. Above this threshold, the standard model increasingly misattributes masking effects to component failure rates. For system-level inference (system hazard, mean time to failure), the standard model remains reliable across the full severity range.

5 Discussion

5.1 When Does C2 Matter?

Our results delineate two distinct regimes for masked series system inference:

System-level inference: C2 rarely matters. If the goal is to estimate the system hazard rate, the mean time to failure, or the system reliability function, the standard C1-C2-C3 model is robust to C2 violations. The system hazard remained approximately unbiased across all severity levels in our 5-component Weibull system. This robustness arises from the structure of the misspecified score: summing over components eliminates the masking weights, so the total hazard estimating equation is approximately correct regardless of the masking mechanism (Theorem 3.4).

Component-level inference: C2 matters. If the goal is to estimate individual component parameters—for example, to identify the most unreliable component or to plan component-specific maintenance—the standard model can produce substantially biased estimates when C2 is violated. The bias grows with violation severity, and coverage degrades to the point where confidence intervals are misleading. Scale parameters are particularly affected, with the Weibull shape-scale interaction amplifying misspecification bias nonlinearly.

5.2 Practical Guidance

Since the masking mechanism is a diagnostic process that is rarely known in detail, the practitioner typically cannot determine whether C2 holds. We recommend the following approach:

1. **Fit the standard model.** The C1-C2-C3 MLE is well-understood, computationally straightforward, and yields reliable system-level estimates regardless of C2 status.

2. **Assess the inferential target.** If only system-level quantities are needed, the standard model suffices and no further action is required.

3. **Run sensitivity analysis for component-level inference.** Construct a family of plausible $\mathbf{P}$ matrices representing different degrees and directions of masking asymmetry. Refit the model with known $\mathbf{P}$ under each scenario. If the component-level conclusions are stable across the family, C2 violations are not a concern.

4. **Report sensitivity bounds.** When conclusions are sensitive to the assumed masking structure, report the range of estimates across the $\mathbf{P}$ family as sensitivity bounds, analogous to sensitivity analysis in causal inference.

5.3 Limitations

Several limitations should be acknowledged:

- **Perturbation family.** Our simulation sweep uses the Bernoulli $\mathbf{P}(s)$ model as the sole perturbation device. Other violation structures—for example, time-dependent masking where $\mathbf{P}$ varies with t—could behave differently.
- **Lifetime distributions.** We study Weibull components, which subsume exponential as a special case ($k = 1$). Other parametric families (e.g., log-normal, Gompertz) may exhibit different sensitivity profiles.
- **Direction dependence.** The specific direction matrix $\mathbf{D}$ determines which parameters are over- or under-estimated. Different $\mathbf{D}$ matrices would produce different bias patterns, though the system-hazard robustness should persist by Theorem 3.4.
- **Sample size.** Our simulations use $n = 500$. The relative bias would be similar at larger sample sizes (misspecification bias does not shrink with n), but coverage degradation would be more pronounced as confidence intervals narrow.

6 Conclusion

We have studied the sensitivity of maximum likelihood estimation in masked series systems to violations of the non-informative masking assumption (C2). Three contributions emerge:

1. **Misspecification characterization.** When C2 is violated but assumed to hold, the MLE converges to a pseudo-true parameter whose individual component parameters absorb the masking asymmetry, while the system hazard remains consistently estimated (Theorem 3.4). Joint estimation of the masking mechanism and the reliability parameters is not possible due to fundamental non-identifiability (Theorem 3.5).
2. **Quantitative sensitivity map.** A systematic simulation sweep across 11 severity levels for a 5-component Weibull series system reveals that system-level quantities (system hazard) are robust to C2 violations across the full range (bias $< 3\%$), while individual component estimates degrade with severity. The exponential component within the system shows notable resistance to misspecification.
3. **Practical guidance.** Since the masking mechanism is unknowable from the data alone, sensitivity analysis over plausible masking structures is the appropriate response. For system-level inference, the standard C1-C2-C3 model suffices. For component-level inference, we recommend reporting results under multiple masking scenarios.

All methods are implemented in the `mdrelax` R package, available at `https://github.com/queelius/mdrelax`.

References

[1] R. V. Craiu and B. Reiser. Inference for the dependent competing risks model with masked causes of failure. *Lifetime Data Analysis*, 12(1):21–33, 2006. doi: 10.1007/s10985-005-7217-9.

[2] Huairui Guo, Pengying Niu, and F. Szidarovszky. Estimating component reliabilities from incomplete system failure data. In *Proceedings of the Annual Reliability and Maintainability Symposium (RAMS)*, pages 1–6. IEEE, 2013. doi: 10.1109/rams.2013.6517765.

[3] I. Guttman, D. K. J. Lin, B. Reiser, and J. S. Usher. Dependent masking and system life data analysis: Bayesian inference for two-component systems. *Lifetime Data Analysis*, 1(1):87–100, 1995. doi: 10.1007/BF00985260.

[4] D. K. J. Lin and F. M. Guess. System life data analysis with dependent partial knowledge on the exact cause of system failure. *Microelectronics and Reliability*, 34(3):535–544, 1994. doi: 10.1016/0026-2714(94)90090-6.

[5] C. Mukhopadhyay and A. P. Basu. Bayesian analysis of incomplete time and cause of failure data. *Journal of Statistical Planning and Inference*, 136(3):803–838, 2006. doi: 10.1016/j.jspi.2004.08.005.

[6] Alex Towell. Reliability estimation in series systems: Maximum likelihood techniques for right-censored and masked failure data, 2023. URL `https://github.com/queelius/reliability-estimation-in-series-systems`. Master's thesis. Available: `https://github.com/queelius/reliability-estimation-in-series-systems`.

[7] J. S. Usher and T. J. Hodgson. Maximum likelihood analysis of component reliability using masked system life-test data. *IEEE Transactions on Reliability*, 37(5):550–555, 1988. doi: 10.1109/24.9880.

[8] J. S. Usher, D. K. J. Lin, and F. M. Guess. Exact maximum likelihood estimation using masked system data. *IEEE Transactions on Reliability*, 42(4):631–635, 1993. doi: 10.1109/24.273596.

A Score Functions

For reference, we provide the score functions under the misspecified and true models for series systems with general component hazards.

Misspecified score (C1-C2-C3). The score under the standard model for parameter θ_j is:

$$\frac{\partial \ell^{\text{wrong}}}{\partial \theta_j} = \sum_{i=1}^{n} \frac{\partial \log R_j(s_i;\theta_j)}{\partial \theta_j} + \sum_{i:\delta_i=1} \frac{\mathbf{1}_{j\in c_i}\, h_j'(s_i;\theta_j)}{\sum_{k\in c_i} h_k(s_i;\theta_k)}, \tag{20}$$

where $h_j' = \partial h_j / \partial \theta_j$.

True score (C1-C3 with known P). Under the Bernoulli masking model with masking weights $\pi_{k,c}$:

$$\frac{\partial \ell^{\text{true}}}{\partial \theta_j} = \sum_{i=1}^{n} \frac{\partial \log R_j(s_i;\theta_j)}{\partial \theta_j} + \sum_{i:\delta_i=1} \frac{\pi_{j,c_i}\, \mathbf{1}_{j\in c_i}\, h_j'(s_i;\theta_j)}{\sum_{k\in c_i} h_k(s_i;\theta_k)\, \pi_{k,c_i}}. \tag{21}$$

Exponential specialization. For exponential components ($h_j = \theta_j$, $R_j = e^{-\theta_j t}$), these simplify to:

$$\frac{\partial \ell^{\text{wrong}}}{\partial \theta_j} = -\sum_{i=1}^{n} s_i + \sum_{i:\delta_i=1} \frac{\mathbf{1}_{j\in c_i}}{\sum_{k\in c_i} \theta_k}, \tag{22}$$

$$\frac{\partial \ell^{\text{true}}}{\partial \theta_j} = -\sum_{i=1}^{n} s_i + \sum_{i:\delta_i=1} \frac{\pi_{j,c_i}\, \mathbf{1}_{j\in c_i}}{\sum_{k\in c_i} \theta_k\, \pi_{k,c_i}}. \tag{23}$$

B Non-Identifiability Evidence

To demonstrate the non-identifiability of $(\boldsymbol{\theta}, \mathbf{P})$, we attempted joint estimation with $n = 2000$ observations from a 2-component exponential system with $\boldsymbol{\theta}^* = (1.0, 2.0)$ and asymmetric $\mathbf{P}$ ($p_{21} = 0.10$, $p_{12} = 0.90$).

Method	$\hat{\theta}_1$	$\hat{\theta}_2$
Joint $(\boldsymbol{\theta}, \mathbf{P})$	1.62	1.44
Known $\mathbf{P}$	0.97	2.08
True values	1.00	2.00

The joint estimator consistently converges to $\hat{\boldsymbol{\theta}} \approx (1.6, 1.4)$ with $\hat{P}_{21} \approx 0.45$ (true: 0.10), regardless of sample size. The total hazard is well-identified: $\sum_j \hat{\theta}_j = 3.01$ vs. true 3.00. This confirms Theorem 3.5: individual rates and masking probabilities are confounded, but the total system hazard is identifiable.

C Software

All methods are implemented in the `mdrelax` R package (version 1.1.0), available at `https://github.com/queelius/mdrelax`. The package provides:

- Data generation under C1-C2-C3 and relaxed C2 for both exponential and Weibull series systems.
- MLE with analytical score and Fisher information for all model tiers.
- The $\mathbf{P}$ matrix construction and masking probability computation functions used in this paper.
- 1276 unit tests verifying likelihood correctness, score–gradient agreement, FIM consistency, and MLE convergence properties.

The simulation sweep script (`paper/run_sensitivity_sweep.R`) reproduces all figures and tables in Section 4.