Appendices

A Derivation of NPMLE of the retro-hazard

In this section, we briefly discuss some properties of the retro-hazard function H^* and derive the nonparametric maximum likelihood estimator (NPMLE) of H^* under the general condition of left-censored data, of which the CDR model's data structure constitutes a special case. Suppose we have $V_i = \max\{T_i, C_i\}, \Delta_i = \mathbf{1}(V_i = T_i), i = 1, \dots n$, where $T_i \sim e^{-H^*(t)}$. The likelihood for this data is

$$L(H^*) = \prod_{i=1}^{n} \left[-dH^*(V_i)\right]^{\Delta_i} e^{-H^*(V_i)}. \tag{A1}$$

From equation (8), we have $F(t) = e^{-H^*(t)}$, implying that the pdf of X under this formulation is $f(t) = -dH^*(t)\, e^{-H^*(t)}$. It is also apparent that $dH^*(t) \leq 0, t \in (0, \infty)$, so H^* must be nonincreasing. Furthermore, we may deduce that (for a proper distribution of T) since $F(0) = 0$ and $\lim_{x\to\infty} F(t) = 1$, $H^*(0) = \infty$ and $\lim_{t\to\infty} H^*(t) = 0$. The foregoing also implies that

$$H^*(t) = \int_t^{\infty} -dH^*(y). \tag{A2}$$

Apart from a sign change, dH^* is equivalent to the function ρ introduced by Lagakos et al. [21].

Define differentiation of a linear functional J with respect to H^* as [see 28, Section 3.2]

$$\delta_s J = \frac{\partial J}{\partial dH^*(s)}.$$

Now, differentiation of the log-likelihood proceeds using the chain rule and definition (A2):

$$\delta_s \log L(H^*) = \sum_{i=1}^{n} \left\{\Delta_i \delta_s \log\left[-dH^*(V_i)\right] - \delta_s H^*(V_i)\right\}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \Delta_i \frac{\partial \log\left[-dH^*(V_i)\right]}{\partial dH^*(s)} - \sum_{i=1}^{n} \frac{\partial H^*(V_i)}{\partial dH^*(s)} \\
&= \sum_{i=1}^{n} \frac{\Delta_i}{-dH^*(s)} \cdot \frac{\partial}{\partial dH^*(s)} \left[-dH^*(V_i)\right] - \sum_{i=1}^{n} \frac{\partial}{\partial dH^*(s)} \int_{V_i}^{\infty} -dH^*(t) \\
&= \sum_{i=1}^{n} \frac{\Delta_i}{-dH^*(s)} \cdot -\mathbf{1}(V_i = s) - \sum_{i=1}^{n} \int_{0}^{\infty} -\mathbf{1}(V_i \le t) \frac{\partial}{\partial dH^*(s)}\, dH^*(t) \\
&= \sum_{i=1}^{n} \frac{\Delta_i\, \mathbf{1}(V_i = s)}{dH^*(s)} - \sum_{i=1}^{n} \int_{0}^{\infty} -\mathbf{1}(V_i \le t)\, \mathbf{1}(t = s) \\
&= \sum_{i=1}^{n} \frac{\Delta_i\, \mathbf{1}(V_i = s)}{dH^*(s)} - \sum_{i=1}^{n} -\mathbf{1}(V_i \le s).
\end{aligned}
$$

The important identities established here are

$$
\delta_s \log\left[-dH^*(t)\right] = \frac{\mathbf{1}(t = s)}{dH^*(s)}, \quad \delta_s H^*(t) = -\mathbf{1}(t \le s). \tag{A3}
$$

Setting $\delta_s \log L(H^*) = 0$ implies a Nelson–Aalen estimator

$$
\widehat{dH^*}(s) = -\frac{\sum_{i=1}^{n} \Delta_i\, \mathbf{1}(V_i = s)}{\sum_{i=1}^{n} \mathbf{1}(V_i \le s)}.
$$

The negative sign of the estimator indicates that these will be decrements instead of the usual increments in the classical Nelson–Aalen estimator. Otherwise, the form of the estimator is identical, with the only difference being that the "risk set" at point s is composed of observations with $V_i \le s$. Recalling the identity in equation (A2), the estimate of H^* is

$$
\widehat{H^*}(t) = -\int_{t}^{\infty} \widehat{dH^*}(s).
$$

B Derivation of the profile likelihood

We confine ourselves to the observations for which $X_i > 0$ (that is, observations for which damage is observed), and consider the problem of estimating H^*. The log-likelihood for these observations may be written as

$$\ell_2(\boldsymbol{\beta}; H^*) = \sum_{i:X_i>0} \left\{ \int_0^\infty \log\left[-\eta_i \, dH^*(t)\right] \, dN_i^*(t) - \int_0^\infty (\eta_i + \mu_i) Y_i^*(t) \, dH^*(t) \right\} \tag{B1}$$

using the counting processes defined by (9) and (10). By functional differentiation of (11) with respect to H^*, we find

$$\begin{aligned}
\mathcal{U}(s) &= \delta_s \log \left\{ \prod_{i:X_i>0} \left[-\eta_i e^{-(\eta_i+\mu_i)H^*(X_i)} \, dH^*(X_i) \right] \right\} \\
&= \delta_s \sum_{i:X_i>0} \left\{ \log \eta_i + \log\left[-dH^*(X_i)\right] - (\eta_i + \mu_i) H^*(X_i) \right\} \\
&= \sum_{i:X_i>0} \left[\frac{dN_i^*(s)}{dH^*(s)} - (\eta_i + \mu_i) \cdot -Y_i^*(s) \right] \\
&= \sum_{i:X_i>0} \frac{dN_i^*(s)}{dH^*(s)} + \sum_{i:X_i>0} (\eta_i + \mu_i) Y_i^*(s).
\end{aligned}$$

Note that we have used the identities (A3) and the fact that $Y_i^*(s) = \mathbf{1}(X_i \leq s)$. Furthermore, since for this model all observations greater than 0 are uncensored, $dN_i^*(s) = \mathbf{1}(X_i = s)$ when $X_i > 0$. Setting $\mathcal{U}(s) = 0$ implies a Breslow estimator of

$$\widehat{dH^*}(s) = -\frac{\sum_{i:X_i>0} dN_i^*(s)}{\sum_{i:X_i>0} (\eta_i + \mu_i) Y_i^*(s)}. \tag{B2}$$

Substitution of (B2) into the log-likelihood (B1) yields

$$\begin{aligned}
\ell_2(\boldsymbol{\beta}; \widehat{H^*}) &= \int_0^\infty \sum_{i:X_i>0} \log \left[\eta_i \frac{\sum_{j:X_j>0} dN_j^*(t)}{\sum_{j:X_j>0} (\eta_j + \mu_j) Y_j^*(t)} \right] dN_i^*(t) \\
&\quad + \int_0^\infty \sum_{i:X_i>0} (\eta_i + \mu_i) Y_i^*(t) \, \frac{\sum_{j:X_j>0} dN_j^*(t)}{\sum_{j:X_j>0} (\eta_j + \mu_j) Y_j^*(t)} \\
&= \int_0^\infty \sum_{i:X_i>0} \log \left[\eta_i \frac{\sum_{j:X_j>0} dN_j^*(t)}{\sum_{j:X_j>0} (\eta_j + \mu_j) Y_j^*(t)} \right] dN_i^*(t) + \int_0^\infty \sum_{j:X_j>0} dN_j^*(t) \\
&= \text{const.} + \sum_{i:X_i>0} \int_0^\infty \left[\log \eta_i - \log \sum_{j:X_j>0} (\eta_j + \mu_j) Y_j^*(t) \right] dN_i^*(t),
\end{aligned}$$

where in the last line we have absorbed into the constant all terms not involving η or μ. Returning to (11), we see that

$$\begin{aligned} L(\boldsymbol{\beta};\widehat{H^*}) &= e^{\ell_1(\boldsymbol{\beta})+\ell_2(\boldsymbol{\beta};\widehat{H^*})} \\ &\propto \prod_{i:X_i=0} \frac{\mu_i}{\eta_i+\mu_i} \prod_{i:X_i>0} \frac{\eta_i}{\sum_{j:X_j>0}(\eta_j+\mu_j)Y_j^*(X_i)} \\ &= \prod_{i:X_i=0} \frac{\mu_i}{\eta_i+\mu_i} \prod_{i:X_i>0} \frac{\eta_i}{\sum_{j:0<X_j\le X_i}(\eta_j+\mu_j)}. \end{aligned}$$

C Score components and observed information matrix

The profile likelihood is given by equation (14).

C.1 Score components

In the interests of more compact notation, we hereafter adopt the convention that summations over j refer to the set $\{j : 0 < X_j \le X_i\}$. The score components are

$$\begin{aligned} U_0 \equiv \frac{\partial \ell_{\mathrm{pr}}(\boldsymbol{\beta})}{\partial \beta_0} &= \sum_{i:X_i=0} \frac{1}{1+e^{\beta_0+\mathbf{z}_i'\boldsymbol{\beta}_\theta}} - \sum_{i:X_i>0} \frac{\sum_j e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta+\mathbf{z}_j'\boldsymbol{\beta}_\eta}}{\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1+e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)} \\ \mathbf{U}_\theta \equiv \frac{\partial \ell_{\mathrm{pr}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_\theta} &= \sum_{i:X_i=0} \frac{\mathbf{z}_i}{1+e^{\beta_0+\mathbf{z}_i'\boldsymbol{\beta}_\theta}} - \sum_{i:X_i>0} \frac{\sum_j \mathbf{z}_j e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta+\mathbf{z}_j'\boldsymbol{\beta}_\eta}}{\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1+e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)} \\ \mathbf{U}_\eta \equiv \frac{\partial \ell_{\mathrm{pr}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_\eta} &= \sum_{i:X_i>0} \left[\mathbf{z}_i - \frac{\sum_j \mathbf{z}_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1+e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)}{\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1+e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)}\right]. \end{aligned}$$

The score vector is $\mathbf{U}(\boldsymbol{\beta}) = (U_0, \mathbf{U}_\theta', \mathbf{U}_\eta')'$.

C.2 Observed information

The observed information matrix will be

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}) = \begin{bmatrix} \mathcal{I}_{00} & \boldsymbol{\mathcal{I}}'_{\theta 0} & \boldsymbol{\mathcal{I}}'_{\eta 0} \\ \boldsymbol{\mathcal{I}}_{\theta 0} & \boldsymbol{\mathcal{I}}_{\theta\theta} & \boldsymbol{\mathcal{I}}'_{\eta\theta} \\ \boldsymbol{\mathcal{I}}_{\eta 0} & \boldsymbol{\mathcal{I}}_{\eta\theta} & \boldsymbol{\mathcal{I}}_{\eta\eta} \end{bmatrix},$$

with component matrices derived below. $\mathcal{I}_{00}$ is a scalar; $\boldsymbol{\mathcal{I}}_{\theta 0}$ and $\boldsymbol{\mathcal{I}}_{\eta 0}$ are $p \times 1$ vectors; and $\boldsymbol{\mathcal{I}}_{\theta\theta}$, $\boldsymbol{\mathcal{I}}_{\eta\theta}$, and $\boldsymbol{\mathcal{I}}_{\eta\eta}$ are $p \times p$ matrices. Clearly, then, $\boldsymbol{\mathcal{I}}(\boldsymbol{\beta})$ will be a $(2p+1) \times (2p+1)$ matrix. Below, we calculate the elements of this matrix.

- Derivatives of the score with respect to β_0:

$$\mathcal{I}_{00} \equiv -\frac{\partial U_0}{\partial \beta_0} = \sum_{i:X_i=0} \frac{e^{\beta_0+\mathbf{z}_i'\boldsymbol{\beta}_\theta}}{\left(1+e^{\beta_0+\mathbf{z}_i'\boldsymbol{\beta}_\theta}\right)^2} + \sum_{i:X_i>0} \frac{\left[\sum_j e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta+\mathbf{z}_j'\boldsymbol{\beta}_\eta}\right]\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\right]}{\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1+e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]^2}$$

$$\boldsymbol{\mathcal{I}}_{\theta 0} \equiv -\frac{\partial \mathbf{U}_\theta}{\partial \beta_0} = \sum_{i:X_i=0} \frac{\mathbf{z}_i e^{\beta_0+\mathbf{z}_i'\boldsymbol{\beta}_\theta}}{\left(1+e^{\beta_0+\mathbf{z}_i'\boldsymbol{\beta}_\theta}\right)^2} + \sum_{i:X_i>0} \frac{\left[\sum_j \mathbf{z}_j e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta+\mathbf{z}_j'\boldsymbol{\beta}_\eta}\right]\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\right]}{\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1+e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]^2}$$

$$\begin{aligned}\boldsymbol{\mathcal{I}}_{\eta 0} \equiv -\frac{\partial \mathbf{U}_\eta}{\partial \beta_0} = \sum_{i:X_i>0} \Bigg\{ & \frac{\left[\sum_j \mathbf{z}_j e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta+\mathbf{z}_j'\boldsymbol{\beta}_\eta}\right]\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1+e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]}{\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1+e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]^2} \\ & - \frac{\left[\sum_j \mathbf{z}_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1+e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]\left[\sum_j e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta+\mathbf{z}_j'\boldsymbol{\beta}_\eta}\right]}{\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1+e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]^2} \Bigg\}\end{aligned}$$

- Derivatives of the score with respect to $\boldsymbol{\beta}_\theta$:

$$\begin{aligned}\boldsymbol{\mathcal{I}}_{\theta\theta} \equiv -\frac{\partial \mathbf{U}_\theta}{\partial \boldsymbol{\beta}_\theta} = & \sum_{i:X_i=0} \frac{\mathbf{z}_i\mathbf{z}_i' e^{\beta_0+\mathbf{z}_i'\boldsymbol{\beta}_\theta}}{\left(1+e^{\beta_0+\mathbf{z}_i'\boldsymbol{\beta}_\theta}\right)^2} \\ & + \sum_{i:X_i>0} \Bigg\{ \frac{\left[\sum_j \mathbf{z}_j\mathbf{z}_j' e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta+\mathbf{z}_j'\boldsymbol{\beta}_\eta}\right]\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1+e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]}{\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1+e^{\beta_0+\mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]^2}\end{aligned}$$

$$- \frac{\left[\sum_j \mathbf{z}_j e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta + \mathbf{z}_j'\boldsymbol{\beta}_\eta}\right]^{\otimes 2}}{\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1 + e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]^2} \Bigg\}$$

$$\boldsymbol{\mathcal{I}}_{\eta\theta} \equiv -\frac{\partial \mathbf{U}_\eta}{\partial \boldsymbol{\beta}_\theta} = \sum_{i: X_i > 0} \Bigg\{ \frac{\left[\sum_j \mathbf{z}_j \mathbf{z}_j' e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta + \mathbf{z}_j'\boldsymbol{\beta}_\eta}\right]\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1 + e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]}{\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1 + e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]^2}$$

$$- \frac{\left[\sum_j \mathbf{z}_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1 + e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]\left[\sum_j \mathbf{z}_j e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta + \mathbf{z}_j'\boldsymbol{\beta}_\eta}\right]'}{\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1 + e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]^2} \Bigg\}$$

- Derivatives of the score with respect to $\boldsymbol{\beta}_\eta$:

$$\boldsymbol{\mathcal{I}}_{\eta\eta} \equiv -\frac{\partial \mathbf{U}_\eta}{\partial \boldsymbol{\beta}_\eta} = \sum_{i: X_i > 0} \Bigg\{ \frac{\left[\sum_j \mathbf{z}_j \mathbf{z}_j' e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1 + e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1 + e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]}{\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1 + e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]^2}$$

$$- \frac{\left[\sum_j \mathbf{z}_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1 + e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]^{\otimes 2}}{\left[\sum_j e^{\mathbf{z}_j'\boldsymbol{\beta}_\eta}\left(1 + e^{\beta_0 + \mathbf{z}_j'\boldsymbol{\beta}_\theta}\right)\right]^2} \Bigg\}$$

D Further simulation results

D.1 Correct specification

For these simulations, the intercept β_0 in the θ part of the model was allowed to take values -2, 0, and 2, corresponding to, respectively, approximately 18%, 43%, and 71% of observations equal to zero.

The results of the simulation study for the scenario without misspecification are displayed in Table 3. This table shows that bias and variance decrease with increasing sample size, as we would expect. Bias of all parameter estimates also seems to be adversely affected by intercept values differing from zero, however, as is the case for the large values of ESD and ASE. We also see good agreement between the ESD and ASE for moderate to large samples.

In contrast to results for the logistic part of the model, it is clear that bias and variance of the parameter estimates in the continuous part of the model monotonically increase with increasing proportions of zero observations, which is due to effectively decreasing the sample size available for estimation of the η part of the model. We observe good agreement between the ESD and ASE for moderate to large samples, indicating the adequacy of the asymptotic approximations for the covariance matrix of the parameter estimates.

We observe some interesting bias patterns in these results. When the true value of the parameter is negative (corresponding to approximately 18% of responses equal to zero), the bias is also negative, but decreases quickly with increasing sample size. The reverse is true for the simulations with approximately 71% of responses equal to zero. This shows a consistent pattern of bias away from the null in small samples. Bias tends to be larger for the negative and positive intercept scenarios than for the zero intercept scenario, but this is to be expected, because for a binary variable, maximal information is gained from responses with probability roughly equal to 1/2.

There seems to be little to no bias in some cases for the intercept in the logistic part of the model. To explain this, recall that we are varying this parameter in order to examine the effect of different levels of censoring on the model performance: minimal bias occurs when we set the true value of the intercept equal to zero. Usually, we might report relative bias, where the numbers in this table would be divided by the true values of the parameters in order to facilitate comparisons, but of course this is not sensible for a parameter with a true value of zero.

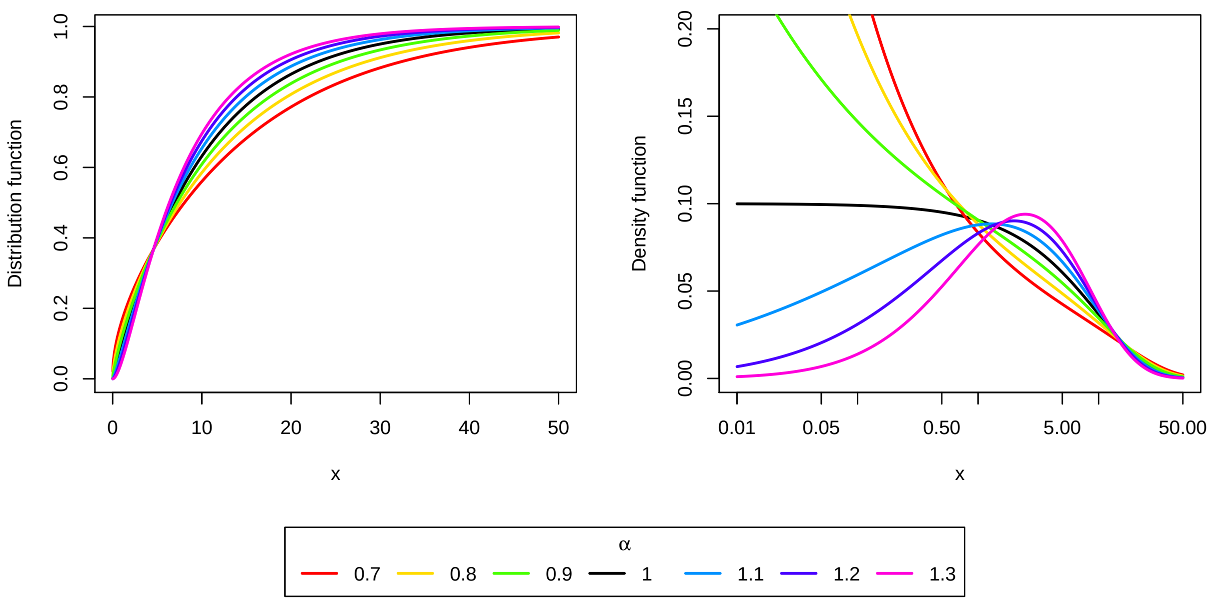
D.2 Non-proportional retro-hazards

For the expected outcome under the true model, we compute $\mathbb{E}\left[D_i\,\mathbf{1}(D_i > R_i)\right]$ for each observation as

$$\int_0^\infty t e^{-\mu_i H^*(t)}\, d\left(e^{-\eta_i [H^*(t)]^\alpha}\right).$$

Because the scale of the outcome may vary with α under this form of misspecification and we would like to be able to compare performance for different degrees of misspecification,

Figure 4: Baseline cdf and pdf plots for simulations in which the baseline retro-hazard is misspecified in the form of equations (18) and (19). The curve in black is the baseline cdf that would be shared between the resistance and damage processes under correct model specification, while the lines in color represent departures from that. Note that while the left-hand panel has an untransformed x-axis, the right-hand panel's x-axis is on the log scale in order to give a clearer idea of the behavior of the density curves.



we evaluated predictive ability of the models using a modified mean-squared error:

$$\text{MSEP}_1 = \frac{1}{n}\sum_{i=1}^{n}\left[1 - \frac{\hat{X}_i}{\mathbb{E}(X_i|\mathbf{z}_i)}\right]^2. \tag{D1}$$

Average values of this quantity across 1000 simulated data sets are shown Table 4.

We see from this table that our method outperforms the standard method uniformly and by a large margin, generally 40-50% regardless of other model parameters. The increased efficiency of our method is most pronounced when there is a lower proportion of zero values in the outcome ($\beta_0 = -2$). The predictive errors increase with increasing proportion of zero values, which is to be expected, as this reduces the amount of information contained in the observed outcome data.

There is an apparent U-shaped relationship between α and the predictive errors under our method (it seems this is also the case for the standard method, although the effect is less obvious). This is sensible, as the model should perform best when it is correctly specified, and indeed $\alpha = 1$ is where we find the minimal average prediction errors under our method. However, this does not seem to be the case for $\beta_0 = 0$: the average predictive error seems relatively flat for $\alpha \leq 1$, while increasing thereafter.

Overall, however, the effect of this kind of misspecification seems to be quite limited, both on our proposed method as well as the standard method. We would not expect the standard method to be affected in any particular way by this form of model misspecification, as it does not rely on our specific model assumptions. The proposed method, on the other hand, can be said to be quite robust to moderate violations of its primary assumption, the proportionality of retro-hazards between the damage and resistance processes.

D.3 Unlinked models

The results for the Cauchy transformation simulations, shown in Table 5, are more favorable for the standard method than our proposed method. For smaller values of the scale parameter

σ, the standard method is substantially more efficient than our proposed method. However, as the curve becomes more linear (i.e., with increasing σ), our proposed method becomes more competitive. Both methods display decreasing trends in predictive error as σ increases, but this effect is much stronger for our proposed method.

Indeed, comparing the results of this table for $\sigma = 4$ with those in Table 1 for $\omega = 1$, we see that predictive errors are quite similar, with $\sqrt{\text{MSEP}_2}$ slightly less than 3 for the standard method and slightly greater than 3 for our proposed method. This reflects the near-linearity of the curve for $\sigma = 4$ and the perfect linearity of the curve for $\omega = 1$ (see Figure 2). Otherwise, it is true that the Cauchy transformations lead to smaller predictive errors than the family of power transformations.

Table 3: Simulation results under correct model specification (bias and standard errors). This table shows the bias, empirical standard deviation (ESD), and average standard error (ASE) of the parameter estimates across all simulated data sets for the part of the model pertaining to the probability of positive damage being observed ("logistic part") and for the part of the model pertaining to amount of damage ("continuous part"). The intercept parameter β_0 was allowed to take values -2, 0, and 2 (shown in the first column), corresponding to, respectively, approximately 18%, 43%, and 71% of observations equal to zero.

Logistic part of model

		β_0			$\beta_{\theta 1} = 2$			$\beta_{\theta 2} = -1$		
β_0	n	Bias	ESD	ASE	Bias	ESD	ASE	Bias	ESD	ASE
-2	100	-0.177	0.620	0.556	0.248	0.598	0.519	-0.137	0.784	0.693
	200	-0.077	0.397	0.367	0.101	0.358	0.334	-0.040	0.468	0.458
	500	-0.027	0.224	0.224	0.029	0.210	0.200	-0.004	0.279	0.281
0	100	-0.007	0.380	0.364	0.154	0.428	0.406	-0.070	0.528	0.522
	200	0.008	0.253	0.250	0.073	0.294	0.272	-0.026	0.345	0.357
	500	-0.006	0.159	0.156	0.022	0.173	0.165	-0.002	0.220	0.222
2	100	0.167	0.576	0.512	0.184	0.501	0.448	-0.091	0.646	0.586
	200	0.075	0.351	0.342	0.066	0.306	0.292	-0.058	0.392	0.395
	500	0.027	0.212	0.210	0.036	0.178	0.179	-0.011	0.243	0.244

Continuous part of model

		$\beta_{\eta 1} = -1$			$\beta_{\eta 2} = 2$		
β_0	n	Bias	ESD	ASE	Bias	ESD	ASE
-2	100	-0.020	0.162	0.164	0.049	0.318	0.310
	200	-0.007	0.113	0.112	0.021	0.220	0.214
	500	-0.002	0.071	0.069	0.012	0.132	0.133
0	100	-0.016	0.225	0.210	0.073	0.394	0.389
	200	-0.013	0.146	0.143	0.046	0.274	0.266
	500	-0.001	0.089	0.088	0.013	0.166	0.164
2	100	-0.013	0.366	0.339	0.133	0.658	0.623
	200	-0.015	0.227	0.220	0.090	0.422	0.408
	500	-0.010	0.136	0.133	0.026	0.251	0.246

Table 4: Simulation results under misspecified model, with non-proportional retro-hazards: predictive errors, $n = 500$. This table shows the root mean-square error of the predictions ($\sqrt{\text{MSEP}_1}$) for both the standard method (LSSIM) and our proposed method (CDRM); the final column is a measure of relative efficiency, calculated as the ratio of $\sqrt{\text{MSEP}_1}$ for the CDRM method to that of the LSSIM method. This is averaged over 1000 simulated data sets at each distinct combination of intercept value β_0 and misspecification parameter α (where $\alpha = 1$ corresponds to proportional retro-hazards, i.e., correct model specification). The intercept parameter β_0 was allowed to take values -2, -1, and 0 (shown in the first column), corresponding to, respectively, approximately 18%, 29%, and 43% of observations equal to zero.

		Method		
β_0	α	LSSIM	CDRM	Ratio
-2	0.7	0.157	0.077	0.489
	0.8	0.155	0.073	0.467
	0.9	0.151	0.069	0.455
	1	0.152	0.068	0.450
	1.1	0.150	0.070	0.468
	1.2	0.153	0.073	0.478
	1.3	0.150	0.076	0.508
-1	0.7	0.166	0.092	0.558
	0.8	0.160	0.083	0.520
	0.9	0.155	0.082	0.526
	1	0.154	0.079	0.513
	1.1	0.156	0.087	0.555
	1.2	0.160	0.094	0.588
	1.3	0.166	0.107	0.641
0	0.7	0.175	0.107	0.611
	0.8	0.172	0.104	0.604
	0.9	0.169	0.103	0.607
	1	0.169	0.104	0.617
	1.1	0.173	0.111	0.642
	1.2	0.178	0.124	0.698
	1.3	0.191	0.145	0.760

Table 5: Simulation results under misspecified model, Cauchy transformation as unknown function of the index: predictive errors, $n = 500$. This table shows the root mean-square error of the predictions ($\sqrt{\text{MSEP}_2}$, RMSEP) for both the standard method (LSSIM) and our proposed method (CDRM); the final column is a measure of relative efficiency, calculated as the ratio of $\sqrt{\text{MSEP}_2}$ for the CDRM method to that of the LSSIM method. This is averaged over 1000 simulated data sets at each distinct combination of intercept value β_0 and misspecification parameter σ. Also displayed in this table is the average outcome across all subjects and simulated data sets, intended to give an idea of the relative size of the $\sqrt{\text{MSEP}_2}$ values (which are not normalized as they are for MSEP_1). The intercept parameter β_0 was allowed to take values -2, -1, and 0 (shown in the first column), corresponding to, respectively, approximately 18%, 29%, and 43% of observations equal to zero.

β_0	σ	$\mathbb{E}\,X$	LSSIM	CDRM	Ratio
-2	1	72.4	3.166	7.242	2.287
	2	75.7	3.017	4.861	1.611
	3	77.5	2.983	3.878	1.300
	4	78.5	2.912	3.414	1.172
-1	1	61.0	3.108	7.181	2.310
	2	64.4	2.959	4.522	1.528
	3	66.2	2.953	3.541	1.199
	4	67.3	2.964	3.362	1.134
0	1	47.7	2.840	6.207	2.185
	2	50.9	2.881	4.043	1.403
	3	52.7	2.814	3.263	1.160
	4	53.6	2.973	3.259	1.096