APPENDIX A. Supplementary Material

APPENDIX A.1. Regularity conditions, lemmas, and proofs of the theorems

The following regularity conditions will be needed for the asymptotic properties of the proposed prediction accuracy measures R^2 and L^2.

(C1) The censoring time C is independent of Y and X.

(C2) $\hat{\theta}$ converges in probability to a limit θ^* as $n \to \infty$.

(C3) $m_{\theta^*}(x)$ is a bounded function and $E(Y^4) < \infty$.

(C4) As $n \to \infty$, $m_{\hat{\theta}}(x) - m_{\theta^*}(x) = K(x)\frac{1}{n}\sum_{i=1}^{n}\xi_i + o_p(\frac{1}{\sqrt{n}})$, uniformaly in x, for some bounded function $K(x)$ and some sequence of independent and identically distributed random variables ξ_i's with mean 0 and finite variance.

(C5) $F(\tau_H-) < 1$ or $\Delta G(\tau_H) = 0$, where F is the marginal distribution of Y, $H = 1-(1-F)(1-G)$, and $\tau_H = \sup\{t : H(t) < 1\}$

Condition (C1) assumes that the censoring time is independent of both the survival time Y and the covariate X, which is used to prove the consistency of the proposed censored accuracy measures. Condition (C2) is satisfied by a consistent estimator under a correctly specified model. For common parametric and semiparametric models, the maximum likelihood estimate typically converges to a well defined limit even if the model is mis-specified (see, e.g., Huber (1967)) and, in which case, θ^* is usually the parameter value that minimizes the Kullback-Leibler Information Criterion (Akaike, 1998). (C3)-C4) are technical conditions for the asymptotic properties in Theorem 2.2, which usually holds for common used parametric and semiparametric models under mild regularity conditions . For example, if $\hat{\theta}$ is the maximum likelihood estimate for a correctly specified parametric model, then by the Taylor series expansion with respect to θ, (C4) is trivially satisfied provided that $m_\theta(x)$ has bounded first and second derivatives with respect to θ. (C5) is required for the uniform consistency of $\hat{G}$, which is needed by Lemma A.4 and Theorem 3.1.

The following lemma establishes a variance decomposition and a prediction error decomposition, which provide the rationale for the proposed population prediction accuracy measures $\rho^2_{m_{\theta^*}}$ and $\lambda^2_{m_{\theta^*}}$ defined in (6) and (7).

Lemma A.1 *Let* $m^{(c)}_{\theta^*}(X)$ *be the corrected prediction function of* $m_{\theta^*}(X)$ *defined by (3). Then,*

(a) (Variance decomposition)

$$var(Y) = E\{m^{(c)}_{\theta^*}(X) - \mu_Y\}^2 + E\{Y - m^{(c)}_{\theta^*}(X)\}^2, \quad (A.1)$$

where the first and second terms on the right hand side represent respectively the explained variance and the unexplained variance of Y *by* $m^{(c)}_{\theta^*}(X)$.

(b) (Prediction Error Decomposition)

$$MSPE(m_{\theta^*}(X)) = E\{Y - m^{(c)}_{\theta^*}(X)\}^2 + E\{m^{(c)}_{\theta^*}(X) - m_{\theta^*}(X)\}^2 \quad (A.2)$$

where the first and second terms on the right hand side can be interpreted as the explained prediction error and unexplained prediction error of $m_{\theta^*}(X)$ *by* $m^{(c)}_{\theta^*}(X)$.

PROOF OF LEMMA A.1. (a) Note that

$$\begin{aligned} var(Y) &= E(Y - \mu_Y)^2 \\ &= E\{Y - m^{(c)}_{\theta^*}(X)\}^2 + 2E\{m^{(c)}_{\theta^*}(X) - \mu_Y\}\{Y - m^{(c)}_{\theta^*}(X)\} + E\{m^{(c)}_{\theta^*}(X) - \mu_Y\}^2. \end{aligned}$$

So it suffices to show that

$$E\{m^{(c)}_{\theta^*}(X) - \mu_Y\}\{Y - m^{(c)}_{\theta^*}(X)\} = 0. \quad (A.3)$$

Recall that $m^{(c)}_{\theta^*}(X) = \tilde{a} + \tilde{b} m_{\theta^*}(X)$, where $(\tilde{a}, \tilde{b}) = \arg\min_{\alpha,\beta} E\{Y - (\alpha + \beta m_{\theta^*}(X))\}^2$. Thus,

$$\left.\frac{\partial E\{Y - (\alpha + \beta m_{\theta^*}(X))\}^2}{\partial \alpha}\right|_{(\alpha,\beta)=(\tilde{a},\tilde{b})} = -2E\{Y - (\tilde{a} + \tilde{b} m_{\theta^*}(X))\} = 0,$$

and

$$\left.\frac{\partial E\{Y - (\alpha + \beta m_{\theta^*}(X))\}^2}{\partial \beta}\right|_{(\alpha,\beta)=(\tilde{a},\tilde{b})} = -2E[\{Y - (\tilde{a} + \tilde{b} m_{\theta^*}(X))\} m_{\theta^*}(X)] = 0,$$

which imply that

$$E\{Y - m^{(c)}_{\theta^*}(X)\} = 0, \quad (A.4)$$

and

$$E[\{Y - m_{\theta^*}^{(c)}(X)\}m_{\theta^*}(X)] = 0. \tag{A.5}$$

Finally, (A.3) follows from (A.2) and (A.5). This proves (A.1).

(b). Note that

$$\begin{aligned}
& E\{Y - m_{\theta^*}^{(c)}(X)\}\{m_{\theta^*}^{(c)}(X) - m_{\theta^*}(X)\} \\
= \ & E\{Y - m_{\theta^*}^{(c)}(X)\}\{\tilde{a} + \tilde{b}m_{\theta^*}(X) - m_{\theta^*}(X)\} \\
= \ & \tilde{a}E\{Y - m_{\theta^*}^{(c)}(X)\} + (\tilde{b} - 1)E[\{Y - m_{\theta^*}^{(c)}(X)\}m_{\theta^*}(X)] \\
= \ & 0,
\end{aligned}$$

where the last equality follows from (A.2) and (A.5). This implies that (A.2) holds. □

PROOF OF THEOREM 2.1. The proofs for parts (a)-(c) are straightforward. Part (d) follows directly from the fact that $\mu(X) = E(Y|X)$ is the best prediction function for Y among all functions of X in a sense that $E\{Y - \mu(X)\}^2 \leq E\{Y - Q(X)\}^2$ for any p-variate function Q, and that the equality holds when $Q(X) = \mu(X)$. □

The following lemma establishes a sample variance decomposition and a sample prediction error decomposition, which provide the rationale for the proposed sample prediction accuracy measures $R^2_{m_{\hat{\theta}}}$ and $L^2_{m_{\hat{\theta}}}$ defined in (11) and (12).

Lemma A.2 *Define*

$$m_{\hat{\theta}}^{(c)}(x) = \hat{a} + \hat{b}m_{\hat{\theta}}(x), \tag{A.6}$$

to be the linearly corrected function for $m_{\hat{\theta}}(x)$*, where* $\hat{a} = \bar{Y} - \hat{b}\bar{m}_{\hat{\theta}}$, $\hat{b} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})\{m_{\hat{\theta}}(X_i) - \bar{m}_{\hat{\theta}}\}}{\sum_{i=1}^n \{m_{\hat{\theta}}(X_i) - \bar{m}_{\hat{\theta}}\}^2}$, $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i$*, and* $\bar{m}_{\hat{\theta}} = n^{-1}\sum_{i=1}^n m_{\hat{\theta}}(X_i)$*. In other words,* $m_{\hat{\theta}}^{(c)}(x)$ *is the ordinary least squares regression function obtained by linearly regressing* $Y_1, \ldots, Y_n$ *on* $m_{\hat{\theta}}(X_1), \ldots, m_{\hat{\theta}}(X_n)$*. Then*

(a) (Variance Decomposition)

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (m_{\hat{\theta}}^{(c)}(X_i) - \bar{Y})^2 + \sum_{i=1}^n (Y_i - m_{\hat{\theta}}^{(c)}(X_i))^2; \tag{A.7}$$

(b) (Prediction Error Decomposition)

$$\sum_{i=1}^{n}(Y_i - m_{\hat{\theta}}(X_i))^2 = \sum_{i=1}^{n}(Y_i - m_{\hat{\theta}}^{(c)}(X_i))^2 + \sum_{i=1}^{n}(m_{\hat{\theta}}^{(c)}(X_i) - m_{\hat{\theta}}(X_i))^2. \qquad \text{(A.8)}$$

PROOF OF LEMMA A.2. (a). The variance decomposition (A.7) is a trivial consequence of the fact that $m_{\hat{\theta}}^{(c)}(X)$ is the fitted value from the simple linear regression of Y on $m_{\hat{\theta}}(X)$.

(b) Now we prove the prediction error decomposition (A.8). For the simple linear regression of Y on a covariate Z, it is well known that

$$\sum_{i=1}^{n} e_i Z_i = 0 \quad \text{and} \quad \sum_{i=1}^{n} e_i \hat{y}_i = 0, \qquad \text{(A.9)}$$

where $\hat{y}_i$ is the fitted value and $e_i = Y_i - \hat{y}_i$ is the residual at Z_i, $i = 1, \ldots, n$. In our context, $Z_i = m_{\hat{\theta}}(X_i)$ and $\hat{y}_i = m_{\hat{\theta}}^{(c)}(X_i)$, and thus (A.9) implies that

$$\sum_{i=1}^{n}\{Y_i - m_{\hat{\theta}}^{(c)}(X_i)\} m_{\theta^*}(X_i) = 0 \quad \text{and} \quad \sum_{i=1}^{n}\{Y_i - m_{\hat{\theta}}^{(c)}(X_i)\} m_{\hat{\theta}}^{(c)}(X_i) = 0.$$

Consequently,

$$\begin{aligned}
\sum_{i=1}^{n}\{Y_i - m_{\hat{\theta}}(X_i)\}^2 &= \sum_{i=1}^{n}\{Y_i - m_{\hat{\theta}}^{(c)}(X_i)\}^2 + \sum_{i=1}^{n}\{m_{\hat{\theta}}^{(c)}(X_i) - m_{\hat{\theta}}(X_i)\}^2 \\
&\quad +2\sum_{i=1}^{n}\{Y_i - m_{\hat{\theta}}^{(c)}(X_i)\}\{m_{\hat{\theta}}^{(c)}(X_i) - m_{\hat{\theta}}(X_i)\}^2 \\
&= \sum_{i=1}^{n}\{Y_i - m_{\hat{\theta}}^{(c)}(X_i)\}^2 + \sum_{i=1}^{n}\{m_{\hat{\theta}}^{(c)}(X_i) - m_{\hat{\theta}}(X_i)\}^2.
\end{aligned}$$

This proves (A.8). $\square$

PROOF OF THEOREM 2.2. (a) It suffices to show that

$$\frac{1}{n}\sum_{i=1}^{n} Y_i m_{\hat\theta}(X_i) \xrightarrow{P} E\{Y m_{\theta^*}(X)\}, \tag{A.10}$$

$$\frac{1}{n}\sum_{i=1}^{n} m_{\hat\theta}(X_i) \xrightarrow{P} E\{m_{\theta^*}(X)\}, \tag{A.11}$$

$$\frac{1}{n}\sum_{i=1}^{n} m_{\hat\theta}^2(X_i) \xrightarrow{P} E\{m_{\theta^*}^2(X)\}. \tag{A.12}$$

We only prove (A.10) here because the proof of the other two results are similar. Note that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} Y_i m_{\hat\theta}(X_i) &= \frac{1}{n}\sum_{i=1}^{n} Y_i m_{\theta^*}(X_i) + \frac{1}{n}\sum_{i=1}^{n} Y_i\{m_{\hat\theta}(X_i) - m_{\theta^*}(X_i)\} \\ &= I_1 + I_2.\end{aligned}$$

By the law of large numbers, $I_1 \xrightarrow{P} E\{Y m_{\theta^*}(X)\}$. Moreover, by condition (C4) and the law of large numbers,

$$I_2 = \left\{\frac{1}{n}\sum_{i=1}^{n} Y_i K(X_i)\right\}\left(\frac{1}{n}\sum_{j=1}^{n}\xi_j\right) + o_p(\frac{1}{\sqrt{n}})\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \xrightarrow{P} 0.$$

This proves (A.10).

(b). Note that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[Y_i m_{\hat\theta}(X_i) - E\{Y m_{\theta^*}(X)\}] &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[Y_i m_{\theta^*}(X_i) - E\{Y m_{\theta^*}(X)\}] \\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i\{m_{\hat\theta}(X_i) - m_{\theta^*}(X_i)\} \\ &= J_1 + J_2.\end{aligned}$$

Furthermore,

$$\begin{aligned}
J_2 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i\{m_{\hat\theta}(X_i) - m_{\theta^*}(X_i)\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i\{K(X_i)\frac{1}{n}\sum_{j=1}^{n}\xi_j + o_p(\frac{1}{\sqrt{n}})\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i E[YK(X)] + \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\right)\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i K(X_i) - E[YK(X)]\right\} + o_p(\frac{1}{\sqrt{n}})\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i \\
&\equiv J_{21} + J_{22} + J_{23},
\end{aligned}$$

where the second equality is from condition (C4). Then, by the central limit theorem, $J_1 + J_{21}$ is asymptotically normal with mean 0. Moreover, applying the central limit theorem and the law of large numbers, $J_{22} = o_p(1)$ and $J_{23} = o_p(1)$ as $n \to \infty$. Therefore, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[Y_i m_{\hat\theta}(X_i) - E\{Y m_{\theta^*}(X)\}]$ is asymptotically normal with mean 0.

Part (b) can be proved by first establishing the joint convergence of multiple quantities in the expression of $R^2_{m_{\hat\theta}}$ and $L^2_{m_{\hat\theta}}$ to a multivariate normal limit along similar lines to the above and then applying the delta method. $\square$

The following lemma establishes a weighted sample version of the variance decomposition and prediction error decompositions, which together with Lemma A.4 stated later, provides the rationale for the proposed right-censored sample prediction accuracy measures $R^2_{m_{\hat\theta}}$ and $L^2_{m_{\hat\theta}}$ defined in (16) and (17).

Lemma A.3 *Let $w_1, \ldots, w_n$ be a set of nonnegative real numbers satisfying $\sum_{i=1}^n w_i = 1$ Define*

$$m_{\hat\theta}^{(wc)}(x) = \hat a^{(w)} + \hat b^{(w)} m_{\hat\theta}(x), \tag{A.13}$$

to be a linearly corrected function for $m_{\hat\theta}(x)$, where $\hat a^{(w)} = \bar T^{(w)} - \hat b^{(w)}\bar m_{\hat\theta}^{(w)}$, $\bar T^{(w)} = \sum_{i=1}^n w_i T_i$, $\hat b^{(w)} = \frac{\sum_{i=1}^n w_i (T_i - \bar T^{(w)})\{m_{\hat\theta}(X_i) - \bar m_{\hat\theta}^{(w)}\}}{\sum_{i=1}^n w_i\{m_{\hat\theta}(X_i) - \bar m_{\hat\theta}^{(w)}\}^2}$, and $\bar m_{\hat\theta}^{(w)} = \sum_{i=1}^n w_i m_{\hat\theta}(X_i)$. In other words, $m_{\hat\theta}^{(wc)}(x)$ is the fitted regression function from the weighted least squares linear regression of $Y_1, \ldots, Y_n$ on $m_{\hat\theta}(X_1), \ldots, m_{\hat\theta}(X_n)$ with weight $W = diag\{w_1, \ldots, w_n\}$. Then

(a) (Weighted Variance Decomposition for T)

$$\sum_{i=1}^n w_i\{T_i - \bar T^{(w)}\}^2 = \sum_{i=1}^n w_i\{m_{\hat\theta}^{(wc)}(X_i) - \bar T^{(w)}\}^2 + \sum_{i=1}^n w_i\{T_i - m_{\hat\theta}^{(wc)}(X_i)\}^2; \tag{A.14}$$

(b) (Weighted Prediction Error Decomposition for T)

$$\sum_{i=1}^{n} w_i\{T_i - m_{\hat{\theta}}(X_i)\}^2 = \sum_{i=1}^{n} w_i\{T_i - m_{\hat{\theta}}^{(wc)}(X_i)\}^2 + \sum_{i=1}^{n} w_i\{m_{\hat{\theta}}^{(wc)}(X_i) - m_{\hat{\theta}}(X_i)\}^2. \tag{A.15}$$

PROOF OF LEMMA A.3. (a) Recall that $W = diag(w_1, \ldots, w_n)$. Define $\boldsymbol{t} = (T_1, \ldots, T_n)'$, $\hat{\boldsymbol{t}} = (m_{\hat{\theta}}^{(wc)}(X_1), \ldots, m_{\hat{\theta}}^{(wc)}(X_n))'$, $\boldsymbol{z} = (m_{\hat{\theta}}(X_1), \ldots, m_{\hat{\theta}}(X_n))'$, and $\boldsymbol{Z} = (\mathbf{1}, \boldsymbol{z})$. where $\mathbf{1} = (1, \ldots, 1)'$ is a n dimensional column vector of 1's. Then, by the definition of $m_{\hat{\theta}}^{(wc)}$, we have

$$\hat{\boldsymbol{t}} = \boldsymbol{Z}(\boldsymbol{Z}'W\boldsymbol{Z})^{-1}\boldsymbol{Z}'W\boldsymbol{t}.$$

Note that

$$(\boldsymbol{t}-\hat{\boldsymbol{t}})'W(\mathbf{1}\ \boldsymbol{z}) = (\boldsymbol{t}-\hat{\boldsymbol{t}})'W\boldsymbol{Z} = \boldsymbol{t}'\{I - W\boldsymbol{Z}(\boldsymbol{Z}'W\boldsymbol{Z})^{-1}\boldsymbol{Z}'\}W\boldsymbol{Z} = 0,$$

which implies that

$$(\boldsymbol{t} - \hat{\boldsymbol{t}})'W\mathbf{1} = 0, (\boldsymbol{t} - \hat{\boldsymbol{t}})'W\boldsymbol{z} = 0, \text{ and} (\boldsymbol{t}-\hat{\boldsymbol{t}})'W\hat{\boldsymbol{t}} = (\boldsymbol{t}-\hat{\boldsymbol{t}})'W\boldsymbol{Z}(\boldsymbol{Z}'W\boldsymbol{Z})^{-1}\boldsymbol{Z}'W\boldsymbol{t} = 0 \tag{A.16}$$

Therefore,

$$\begin{aligned} \sum_{i=1}^{n} w_i\{T_i - \bar{T}^{(w)}\}^2 &= (\boldsymbol{t} - \mathbf{1}\mathbf{1}'W\boldsymbol{t})'W(\boldsymbol{t} - \mathbf{1}\mathbf{1}'W\boldsymbol{t}) \\ &= (\boldsymbol{t} - \hat{\boldsymbol{t}})'W(\boldsymbol{t} - \hat{\boldsymbol{t}}) + (\hat{\boldsymbol{t}} - \mathbf{1}\mathbf{1}'W\boldsymbol{t})'W(\hat{\boldsymbol{t}} - \mathbf{1}\mathbf{1}'W\boldsymbol{t}) \\ &\quad +2(\boldsymbol{t} - \hat{\boldsymbol{t}})'W(\hat{\boldsymbol{t}} - \mathbf{1}\mathbf{1}'W\boldsymbol{t}) \\ &= (\boldsymbol{t} - \hat{\boldsymbol{t}})'W(\boldsymbol{t} - \hat{\boldsymbol{t}}) + (\hat{\boldsymbol{t}} - \mathbf{1}\mathbf{1}'W\boldsymbol{t})'W(\hat{\boldsymbol{t}} - \mathbf{1}\mathbf{1}'W\boldsymbol{t}) \\ &= \sum_{i=1}^{n} w_i\{m_{\hat{\theta}}^{(wc)}(X_i) - \bar{T}^{(w)}\}^2 + \sum_{i=1}^{n} w_i\{T_i - m_{\hat{\theta}}^{(wc)}(X_i)\}^2, \end{aligned}$$

where the third equality follows from (A.16). This proves part (a).

(b).

$$\begin{aligned}
\sum_{i=1}^{n} w_i\{T_i - m_{\hat\theta}(X_i)\}^2 &= \sum_{i=1}^{n} w_i\{T_i - m_{\hat\theta}^{(wc)}(X_i)\}^2 + \sum_{i=1}^{n} w_i\{m_{\hat\theta}^{(wc)}(X_i) - m_{\hat\theta}(X_i)\}^2 \\
&\quad +2\sum_{i=1}^{n} w_i\{T_i - m_{\hat\theta}^{(wc)}(X_i)\}\{m_{\hat\theta}^{(wc)}(X_i) - m_{\hat\theta}(X_i)\} \\
&= \sum_{i=1}^{n} w_i\{T_i - m_{\hat\theta}^{(wc)}(X_i)\}^2 + \sum_{i=1}^{n} w_i\{m_{\hat\theta}^{(wc)}(X_i) - m_{\hat\theta}(X_i)\}^2 \\
&\quad +2(\boldsymbol{t} - \hat{\boldsymbol{t}})' W(\hat{\boldsymbol{t}} - \boldsymbol{z}) \\
&= \sum_{i=1}^{n} w_i\{T_i - m_{\hat\theta}^{(wc)}(X_i)\}^2 + \sum_{i=1}^{n} w_i\{m_{\hat\theta}^{(wc)}(X_i) - m_{\hat\theta}(X_i)\}^2,
\end{aligned}$$

where the last equality follows from (A.16). This proves part (b). □

The following lemma, together with Lemma A.3, provides the rationale for the right-censored sample prediction accuracy measures $R^2_{m_{\hat\theta}}$ and $L^2_{m_{\hat\theta}}$ defined in (16) and (17).

Lemma A.4 *Let*

$$w_i = \frac{\frac{\delta_i}{\hat G(T_i-)}}{\sum_{j=1}^{n} \frac{\delta_j}{\hat G(T_j-)}}, \quad i = 1, ..., n, \tag{A.17}$$

where $\hat G$ is the Kaplan-Meier (Kaplan and Meier, 1958) estimate of $G(c) = P(C > c)$. Assume conditions (C1)-(C5) hold. Then,

$$\begin{aligned}
&\sum_{i=1}^{n} w_i\{T_i - \bar T^{(w)}\}^2 \xrightarrow{P} var(Y); \\
&\sum_{i=1}^{n} w_i\{m_{\hat\theta}^{(wc)}(X_i) - \bar T^{(w)}\}^2 \xrightarrow{P} E\{m_{\theta^*}^{(c)}(X) - \mu_Y\}^2; \\
&\sum_{i=1}^{n} w_i\{T_i - m_{\hat\theta}^{(wc)}(X_i)\}^2 \xrightarrow{P} E\{Y - m_{\theta^*}^{(c)}(X)\}^2; \\
&\sum_{i=1}^{n} w_i\{T_i - m_{\hat\theta}(X_i)\}^2 \xrightarrow{P} E\{Y - m_{\theta^*}(X)\}^2; \\
&\sum_{i=1}^{n} w_i\{m_{\hat\theta}^{(wc)}(X_i) - m_{\hat\theta}(X_i)\}^2 \xrightarrow{P} E\{m_{\theta^*}^{(c)}(X) - m_{\theta^*}(X)\}^2.
\end{aligned}$$

PROOF OF LEMMA A.4. We first prove the first result of Lemma A.4. Note that for any function $h(T,X)$ of (T,X), we have

$$\begin{aligned} E\left\{\frac{\delta h(T,X)}{1-G(T|X)}\right\} &= E\left[E\left\{\frac{\delta h(T,X)}{1-G(T|X)}\Big|X,Y\right\}\right] \\ &= E\left[E\left\{\frac{\delta h(Y,X)}{1-G(Y|X)}\Big|X,Y\right\}\right] \\ &= E\left\{\frac{h(Y,X)}{1-G(Y|X)}E(\delta|X,Y)\right\} \\ &= E\left\{\frac{h(Y,X)}{1-G(Y|X)}P(C>Y|X,Y)\right\} \\ &= E\left\{\frac{h(Y,X)}{1-G(Y|X)}\{1-G(Y|X)\}\right\} \\ &= E\left\{h(Y,X)\right\}. \end{aligned}$$

In particular, $h(T,X)=1$, $h(T,X)=T$ and $h(T,X)=T^2$, correspond to

$$E\left\{\frac{\delta}{1-G(T|X)}\right\}=1, \quad E\left\{\frac{\delta T}{1-G(T|X)}\right\}=E(Y), \quad \text{and} \quad E\left\{\frac{\delta T^2}{1-G(T|X)}\right\}=E(Y^2),$$

which, combined with the uniform consistency of $\hat{G}$ (Wang, 1987), imply that $\bar{T}^{(w)} = \sum_{i=1}^n w_i T_i = \frac{\sum_{i=1}^n \frac{\delta_i T_i}{\hat{G}(T_i-)}}{\sum_{i=1}^n \frac{\delta_i}{\hat{G}(T_i-)}} \xrightarrow{P} E(Y)$, and $\sum_{i=1}^n w_i T_i^2 \xrightarrow{P} E(Y^2)$. Thus,

$$\sum_{i=1}^n w_i\{T_i-\bar{T}^{(w)}\}^2 = \sum_{i=1}^n w_i T_i^2 - \{\bar{T}^{(w)}\}^2 \xrightarrow{P} E(Y^2)-\{E(Y)\}^2 = var(Y).$$

The proof for the other results of the lemma are similar and omitted. □

PROOF OF THEOREM 3.1. (a). If there is no censoring, or $\delta_i = 1$ for all i, then the Kaplan-Meier estimate of the survival function of the censoring time is identical to 1. Thus $w_i = 1/n$ for all i. The conclusion of (a) follows immediately.

The proof of parts (b) and (c) is essentially the same as that of Theorem 2.2. and thus we omit the details. □

APPENDIX A.2. Additional Simulation Results

APPENDIX A.2.1. Additional Results for Simulation 1

Figure A.1 depicts the plots of the population R^2_{NP}, R^2_{SPH}, and R^2_{SH} measures versus β for Cox's models under the Simulation 1 setting described in Section 4, with the Weibull baseline shape parameter fixed at different values (top panel: $\nu = 0.5$; middle panel: $\nu = 1$; and bottom panel: $\nu = 1$). For each pair of (β, ν), the population measures are approximated by the average over 10 Monte-Carlo samples of size $n = 5,000$. 95% confidence intervals are also provided at selected β values. A snapshot of the results is given in Table 1 of Section 4 to illustrate some weaknesses of R^2_{NP}, R^2_{SPH} revealed by this simulation.

APPENDIX A.2.2. Simulation results for Cox's model ($\rho^2 = 0.20$)

Similar to Figure 3 in which $\rho^2 = 0.50$ for the Cox model, Figure A.2 presents box plots of simulated R^2 and L^2 for the Cox model when $\rho^2 = 0.20$, based on 1,000 replications. Here the parameters under each data setting are adjusted to produce $\rho^2 = 0.20$. Specifically, for the Weibull setting, data is generated from a Weibull model $\log(Y) = \beta^T X + \sigma W$, where $\beta = 1$, $\sigma = 0.52$, $X \sim U(0, 1)$, $W \sim$ standard extreme value distribution. For the log-normal setting, data is generated from $\log(Y) = \beta^T X + \sigma W$, where $\beta = 1$, $\sigma = 0.52$, $X \sim U(0, 1)$, $W \sim$ N(0,1), and $C \sim Weibull(shape = 1, scale = b)$ with b adjusted to produce a given censoring rate. For the inverse Gaussian setting, data is generated from $Y \sim Inverse\,Gaussian(mean = -\frac{e^{\alpha_0+\alpha_1 X}}{\beta_0+\beta_1 X}, shape = e^{2*(\alpha_0+\alpha_1 X)})$,where $\alpha_0 = 3$, $\alpha_1 = -1.55$, $\beta_0 = -1$, $\beta_1 = 0.6$, $X \sim U(0, 1)$. For all three data generation settings, censoring time is generated from $C \sim Weibull(shape = 1, scale = b)$ with b adjusted to produce a given censoring rate.

Cox-Snell residual plots for the Cox model under the nine scenarios of Figure A.2 with no censoring are provided in Figure A.3, which indicate that the Cox's model fits the data well under the Weibull setting (first row), shows almost unnoticeable mild mis-specification under the log-normal setting (second row), and has a little more noticeable misspecification under the inverse Gaussian setting.

It is seen from Figure A.2 that R^2 and L^2 estimate their population values well under all the three data settings under which the Cox model is either correctly specified or only mildly mis-specified as indicated by the Cox-Snell residual plots in Figure A.3. More noticeable bias is only observed when there is more evidence of model misspecification

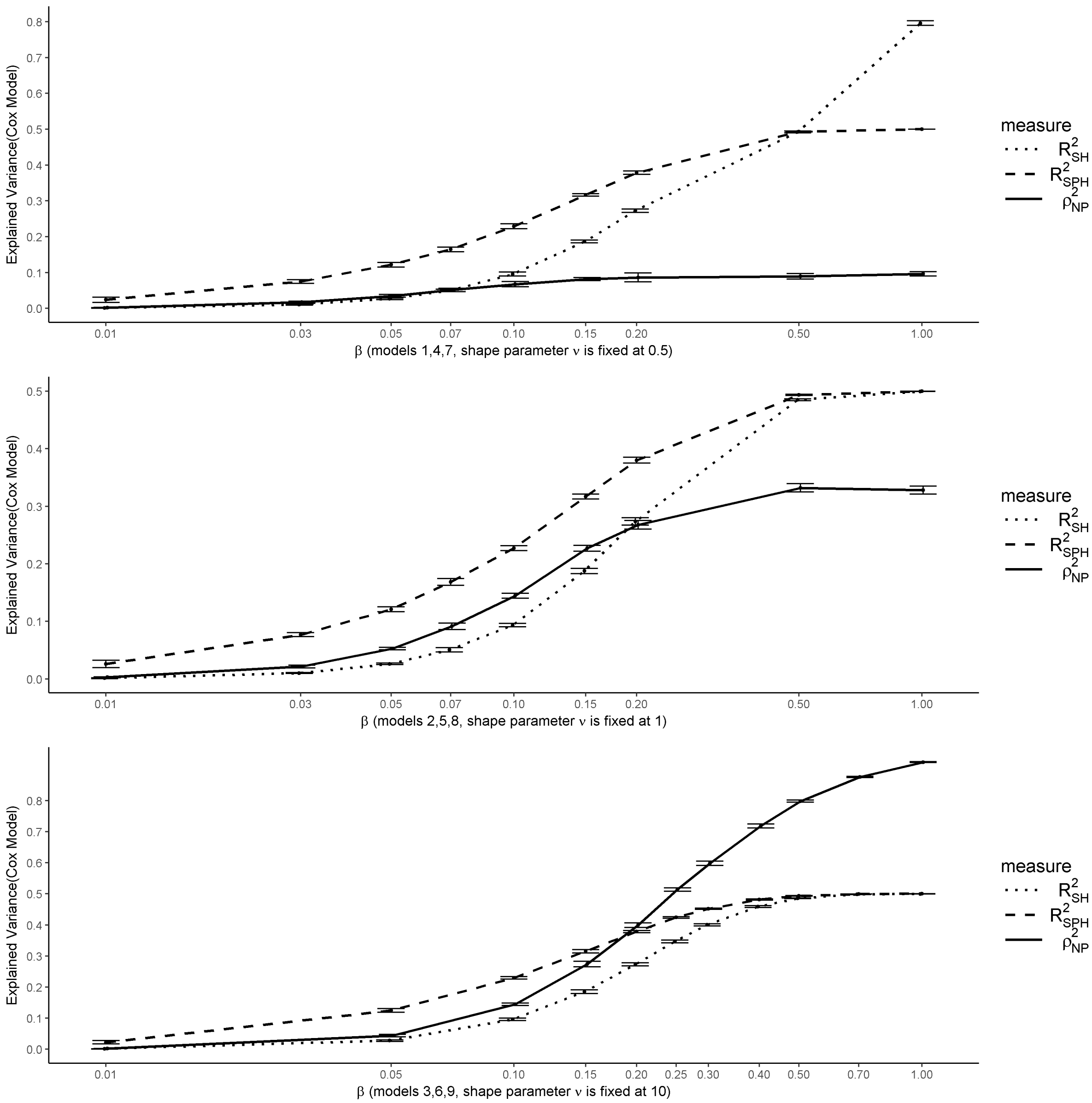

Figure A.1: Population R^2_{NP}, R^2_{SPH}, and R^2_{SH} for Cox's models as the regression coefficient β varies, with the Weibull baseline shape parameter fixed at different values (top panel: $\nu = 0.5$; middle panel: $\nu = 1$; and bottom panel: $\nu = 1$). For each pair of (β, ν), the population measures are approximated by the average over 10 Monte-Carlo samples of size $n = 5,000$. 95% confidence intervals are also provided at selected β values.

(inverse Gaussian setting), smaller sample size(e.g. $n = 100$) and higher censoring rate (e.g. CR=50%).

APPENDIX A.2.3. Simulation results for the threshold regression model

Figure A.4 presents the box plots of simulated R^2 and L^2 for the threshold regression model (Lee and Whitmore, 2006) based on 1,000 replications under the same nine scenarios as in Figure 3. Cox-Snell residual plots for the threshold regression model under the nine scenarios of Figure A.4 with censoring rate CR = 0% are provided in Figure A.5.

Cox-Snell residual plots for the threshold regression model in Figure A.5 indicate that the threshold regression model fits the data well under the inverse Gaussian setting (third row), shows almost unnoticeable mild mis-specification under the log-normal setting (second row), and shows severe lack-of-fit under the Weibull setting (first row).

It is seen from Figure A.4 that R^2 and L^2 estimate their population values well under all the three data settings regardless of whether the threshold regression model is correctly or incorrectly specified.

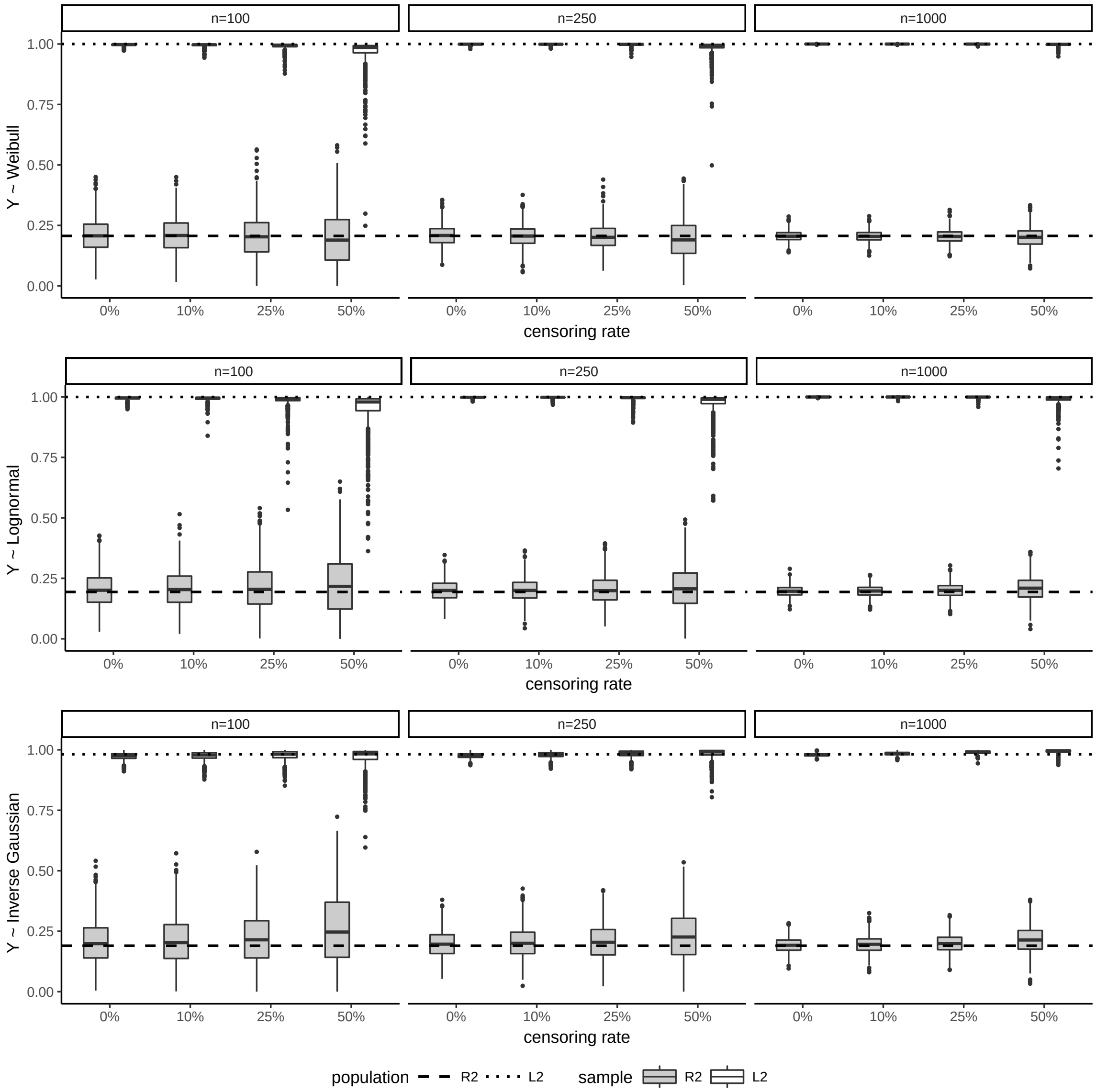

Figure A.2: (Cox's Model with Independent Censoring; $\rho^2 = 0.20$) Box plots of simulated R^2 (shaded box) and L^2 (unshaded box) for the Cox model by censoring rate (0%, 10%, 25%, 50%), sample size (100, 250, 1,000), and data generation setting (upper panel: Weibull; middle panel: log-normal AFT; bottom panel: inverse Gaussian)

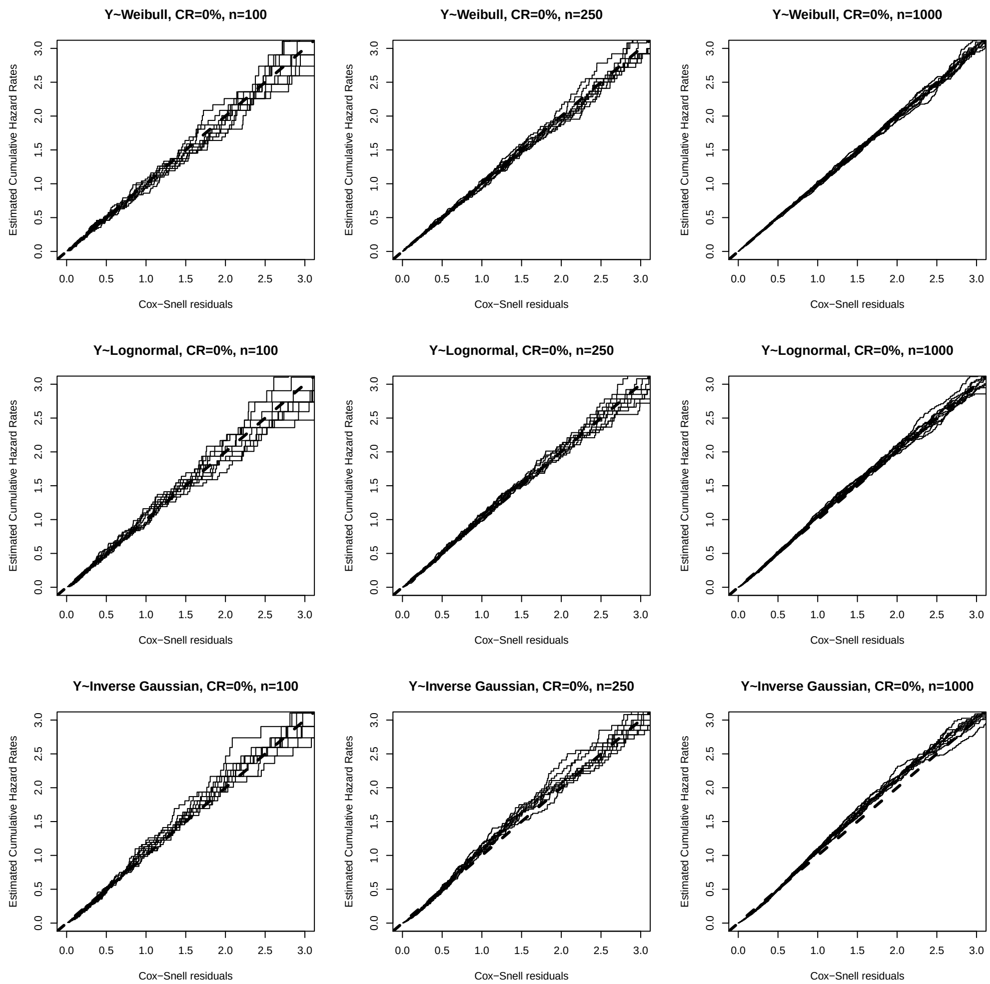

Figure A.3: (Cox Model with Independent Censoring; $\rho^2 = 0.20$; Censoring Rate CR=0%) Cox-Snell residual plot for the Cox model for each of the nine scenarios of Figure A.2 based on the first 10 Monte Carlo replications with censoring rate equal to 0%, varying sample size (first column: $n = 100$; second column: $n = 250$; third column: $n = 1,000$), and varying data generation setting (first row: Weibull; second row: log-normal; third row: inverse Gaussian)

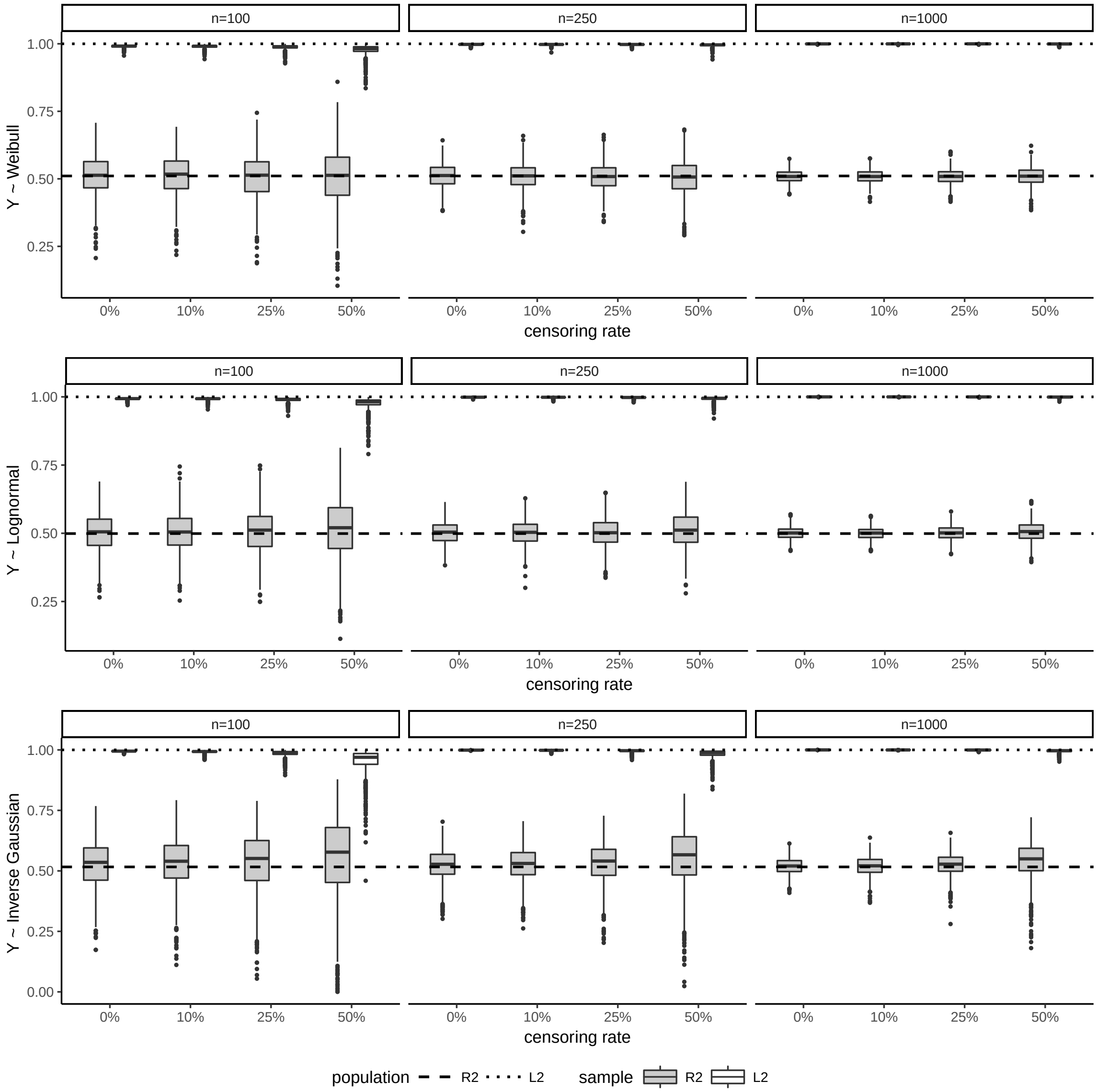

Figure A.4: (Threshold Regression Model with Independent Censoring - $\rho^2 = 0.50$) Box plots of simulated R^2 (shaded box) and L^2 (unshaded box) for the threshold regression model by censoring rate (0%, 10%, 25%, 50%), sample size (100, 250, 1,000), and data generation setting (upper panel: Weibull; middle panel: log-normal; bottom panel: inverse Gaussian)

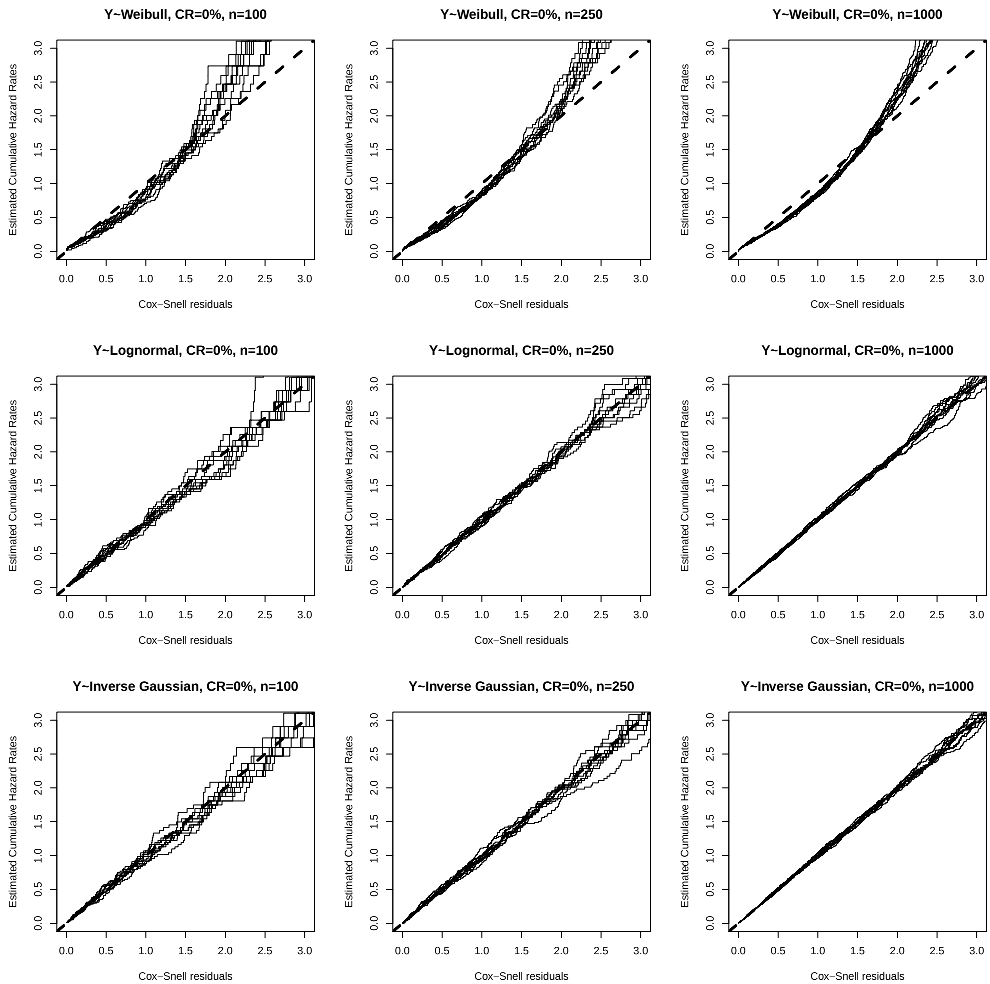

Figure A.5: (Threshold Regression Model with Independent Censoring; $\rho^2 = 0.50$; Censoring Rate CR=0%) Cox-Snell residual plots based on the first 10 Monte Carlo replications for the fitted threshold regression model for each of the nine scenarios of Figure A.4 with censoring rate equal to 0%, varying sample size (first column: $n = 100$; second column: $n = 250$; third column: $n = 1,000$), and varying data generation setting (first row: Weibull; second row: log-normal; third row: inverse Gaussian)