2. Theoretical Background

2.1. Cox Model and Measurement Error

In biometric studies the Cox's *proportional hazards model* (Cox, 1972) has been established as the most popular regression model for analysing survival data. This kind of data is characterized by two parts. On the one hand by *failure*, which describes the occurrence of an event such as death or a specific disease, and on the other hand by *failure time*, which is the time period until the event occurs. A special property of survival data is the presence of *censoring*. A subject is called as censored, when it drops out early without having an event or is event free by the end of the study. This type of censorship is also called *right censoring*. Here, only this type of censorship is considered. For a detailed overview of the different censoring mechanisms see e. g. Kalbfleisch and Prentice (2002).
More formally the survival data for a subject i, $i = 1, \ldots, n$, is defined as follows: T_i denotes the failure time and δ_i is an indicator function, which is 1 if the subject has an event and 0 if the subject is censored. The pair (Y_i, δ_i) characterizes the survival data, where $Y_i = min(T_i, C_i)$ and C_i is the censoring time. The censoring process is assumed to be stochastically independent of T_i given the covariates V_i. The Cox model is used to explore the effect of the covariables on the censored failure times. For example, a question of interest could be how smoking affects the appearance of lung cancer. Before the definition of the Cox model is specified, a few fundamental definitions have to be outlined. A central concept to interpret the Cox model is the *hazard rate* $h(t)$, which is based on the conditional probability that an event occurs in the interval $[t, t + \Delta t)$, given the subject has reached the interval. It is defined by

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}, \tag{2.1}$$

where t is an arbitrary point in time and T the failure time. The hazard rate can also be described by the ratio of the density function $f(t)$ and the survival

H. Birke, *Model-Based Recursive Partitioning with Adjustment for Measurement Error*,
BestMasters, DOI 10.1007/978-3-658-08505-6_2,

function $S(t) = P(T \geq t)$ of T. The density function $f(t)$ is equal to the derivative of the distribution function $F(t) = P(T < t)$. $F(t)$ denotes the probability for the occurrence of an event (before t). The *survival function* is the inverse of $F(t)$, thus $S(t)$ is also given by $S(t) = 1 - F(t)$ and describes the probability that no event has occurred before t or in other words the probability to survive the time point t. From the hazard rate the *cumulative hazard rate* is derived:

$$H(t) = \int_0^t h(u)du = \int_0^t \frac{f(u)}{S(u)}du = -ln(S(t)), \tag{2.2}$$

which is the cumulative risk for an event over the time. Therefore, the survival function is also characterized via the cumulative hazard rate, $S(t) = exp(-H(t))$. In the Cox model the hazard rate is considered as a function of the covariates. The model is given for subject i, $i = 1, \ldots, n$, with a vector of covariates V_i by

$$h(t, V_i) = h_0(t) \cdot exp(V_i' \beta), \tag{2.3}$$

where β is the unknown coefficient vector for V_i. $h_0(t)$ denotes the *baseline hazard* and specifies the risk for an event in t given all covariates are set to zero. The baseline hazard can be left unspecified. Thus, the model is a semi-parametric approach, i. e. no assumption about the distribution of the hazard rate is made, and therefore quite flexible in practice. The Cox model underlies the so-called *proportional hazard assumption*. The ratio of the hazard function of subject i with a set of covariates V_i and a subject j with a set of covariates V_j is equal to

$$\frac{h(t, V_i)}{h(t, V_j)} = \frac{h_0(t) \cdot exp(V_i' \beta)}{h_0(t) \cdot exp(V_j' \beta)} = exp([V_i - V_j]\beta) = const. \tag{2.4}$$

The ratio does not depend on t, i. e. the effect of the covariates on the hazard rate is proportional over the time. The ratio is also called *hazard ratio.*
Given that $h_0(t)$ is not specified, it is not possible to conduct a usual likelihood estimation for β. A so-called *partial likelihood* is employed, where $h_0(t)$ is considered as a nuisance parameter. The root of the derivative of the partial log-likelihood

$$\sum_{j=1}^{k} \left(V_j - \frac{\sum_{i \in R(\pi_j)} V_i \cdot exp(\beta' V_i)}{\sum_{i \in R(\pi_j)} exp(\beta' V_i)} \right) = 0, \tag{2.5}$$

provides the parameter estimation $\hat{\beta}$ (more details in Section 2.1.3).
Often in medical surveys covariates of interest cannot be recorded exactly. Surrogates have to be used instead. For example, the true blood pressure is difficult to investigate, usually an average over frequent measurements is taken. By using the surrogates W_i instead of V_i the naive estimator $\hat{\beta}_{naive}$ of β is obtained by maximizing the likelihood function. Substituting V_i by W_i in the score function of the Cox model result in

$$\sum_{j=1}^{k} \left(W_j - \frac{\sum_{i \in R(\pi_j)} W_i \cdot exp(\beta' W_i)}{\sum_{i \in R(\pi_j)} exp(\beta' W_i)} \right) = 0. \tag{2.6}$$

The root of this so-called *naive score function* is the naive parameter estimator $\hat{\beta}_{naive}$. Ignoring the measurement error, $\hat{\beta}_{naive}$ is biased and does not converge to the true parameter, as has been illustrated by Prentice (1982). Statistical models, including covariates which are subject to measurement errors, also termed as *error-in-variables models*. For this model class several methods exit to adjust for error in variables (see Section 2.1.2). A detailed description is given in Carroll et al. (2006).

2.1.1. Measurement Error Model

The adjustment for measurement errors is only possible due to the knowledge of the relationship between the true variables V_i and the surrogates W_i. An *additive measurement error model* (Carroll et al., 2006, Chapter 1) is assumed:

$$W_i = V_i + U_i, \tag{2.7}$$

with $\mathbb{E}(U_i) = 0$, $i = 1, \ldots, n$, and U_i is independent of V_i, T_i, δ_i, and of U_l $(l \neq i)$, i. e. the errors U_i are taken to be independent among each other. Furthermore, the distribution of the measurement error U_i has to be known. An identically distribution is not necessary, so that heteroscedastic or homoscedastic errors can be considered. The *moment generating function* of the specified distribution has to exist for all a and is defined by

$$M_{U_i}(a) = \mathbb{E}(exp(a' U_i)), \tag{2.8}$$

which is twice differentiable with respect to a. $M_{U_i}(\cdot)$ is required for the correction of the likelihood and accordingly of the score function (see Section 2.1.3). The outcome variable T_i and δ_i are assumed to be error free. The covariates are independent of time and the variables measured with error are taken to be continuous.

2.1.2. Adjustment for Measurement Error in Cox Model

Several authors have dealt with the question how to adjust for measurement error in the Cox model (e. g. Prentice, 1982; Wang et al., 1997; Buzas, 1998; Hu et al., 1998; Kong et al., 1998; Nakamura, 1992; Augustin, 2004), an overview is given in Augustin and Schwarz (2002). Moreover, Huang and Wang (2000); Hu and Lin (2002); Gorfine et al. (2004); Liu et al. (2004); Li and Ryan (2006); Yi and Lawless (2007); Zucker and Spiegelman (2008) and Wen (2009) have concentrated on this subject.
For the inference on the regression coefficients β, there are two kinds of measurement error models, the *functional* and the *structural approach* (Carroll et al., 2006, Chapter 2). The main difference between these two approaches is the assumption about the distribution of true variables V_i. In the structural method the distribution of V_i is taken to follow a known class of distributions. The functional approach gets along without any parametric assumption on the distribution of V_i.

Structural Approach

The two basic approaches are *regression calibration* and *integrating the likelihood*, where the measurement error is assumed to be homoscedastic. In the following only the regression calibration is outlined, which is used as a comparison to the new implemented approach in the simulation study (see Chapter 4). The basis of the regression calibration is to replace the true but unobservable variables V_i by the regression of V_i on W_i. For inference on β via the likelihood the conditional expectation $\mathbb{E}(V_i|W_i)$ is used instead of V_i. Assuming that V_i is i.i.d. normally distributed with unknown mean μ_V and non-singular covariance matrix Σ_V. Then $W_i \sim N(\mu_V; \Sigma_V + \Sigma_U)$ and $V_i|W_i \sim N(\bar{\mu_i}, \bar{\Sigma})$. Σ_U is an estimator of the measurement error covariance matrix, usually derived from validation data or repeated measurements. In addition, a low failure rate is assumed, also named as *rare disease assumption*. The conditional expectation $\mathbb{E}(V_i|W_i)$ is given by

$$\bar{\mu_i} = \mu_V + \Sigma_V \cdot (\Sigma_V + \Sigma_U)^{-1} \cdot (W_i - \mu_V) \tag{2.9}$$

and $\bar{\Sigma} = \Sigma_V - \Sigma_V \cdot (\Sigma_V + \Sigma_U)^{-1}$ (Augustin and Schwarz, 2002). The estimated conditional expectation is computed as follows

$$\hat{\bar{\mu}}_i = \hat{\mu}_W + (\hat{\Sigma}_W - \hat{\Sigma}_U) \cdot (\hat{\Sigma}_W)^{-1} \cdot (W_i - \hat{\mu}_W), \tag{2.10}$$

given that under the additive measurement error model $\mu_W = \mu_V + \mu_U$ and $\mu_U = 0$ as well as $\Sigma_W = \Sigma_V + \Sigma_U \Leftrightarrow \Sigma_V = \Sigma_W - \Sigma_U$.
Replacing V_i by the conditional expectation $\bar{\mu}_i$ in Equation (2.5) leads to

$$\sum_{j=1}^{k} \left(\bar{\mu}_i - \frac{\sum_{i \in R(\pi_j)} \bar{\mu}_i \cdot exp(\beta' \bar{\mu}_i)}{\sum_{i \in R(\pi_j)} exp(\beta' \bar{\mu}_i)} \right) = 0. \tag{2.11}$$

It can be shown that the corrected parameter estimation $\hat{\beta}_{corr}$ is equal to $\Sigma_V^{-1} \cdot (\Sigma_V + \Sigma_U) \cdot \hat{\beta}_{naive}$. This estimator corresponds to the method in the linear model by multiplying $\hat{\beta}$ with the inverse reliability ratio (see Carroll et al., 2006, Chapter 3). The adjusted estimates are not necessarily consistent, but the bias reduction is substantial (Augustin and Schwarz, 2002). In addition, the approach is easy to implement. Wang et al. (1997) derive this approach for the Cox model in a functional way, where the estimation of the distribution of $V_i|W_i$ is carried out via validation data.

Functional Approach

An established procedure is suggested by Nakamura (1992), which is based on the methodology of the *corrected score function* (see Nakamura, 1990; Carroll et al., 2006, Chapter 7). A corrected score function is proposed, which leads to approximately unbiased estimates. For the correction no distribution assumptions have to be made about the covariates V_i.
In general, a corrected score function to estimate the parameter vector ϑ is one whose conditional expectation with respect to the measurement error distribution coincides with the usual score function based on the unknown true variables V_i (Nakamura, 1990)(see Equation (2.12) below). Given is the sample $\tilde{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_n)$ with covariates $V = (V_1, \dots, V_n)$ and surrogates $W = (W_1, \dots, W_n)$. $s^V(\tilde{Y}, V, \vartheta)$ denotes the *ideal score function*, since V is unobservable. Replacing V by W yields to the *naive score function* $s^V(\tilde{Y}, W, \vartheta)$. In general, $\mathbb{E}_{\vartheta_0}(s^V(\tilde{Y}, W, \vartheta_0))$, where ϑ_0 is the true parameter value of ϑ, is not equal to zero and therefore the naive score function leads to a biased estimator of ϑ. The main idea is to find a corrected score

function $s^W(\tilde{Y}, W, \vartheta)$ and the corrected log-likelihood $l^W(\tilde{Y}, W, \vartheta)$, respectively, such that

$$\mathbb{E}_\vartheta(s^W(\tilde{Y}, W, \vartheta | V, \tilde{Y})) = s^V(\tilde{Y}, V, \vartheta) \tag{2.12}$$

$$\mathbb{E}_\vartheta(l^W(\tilde{Y}, W, \vartheta | V, \tilde{Y})) = l^V(\tilde{Y}, V, \vartheta) \tag{2.13}$$

$$\text{with } s^W(\tilde{Y}, W, \vartheta) := \frac{\partial}{\partial\vartheta} l^W(\tilde{Y}, W, \vartheta) \;\; \forall \vartheta. \tag{2.14}$$

In addition, $\mathbb{E}_{\vartheta_0}(s^V(\tilde{Y}, V, \vartheta_0)) = 0$ holds, thus by the law of iterated expectation

$$\mathbb{E}_{\vartheta_0}(s^W(\tilde{Y}, W, \vartheta_0)) = \mathbb{E}_{\vartheta_0}(\mathbb{E}_{\vartheta_0}(s^W(\tilde{Y}, W, \vartheta_0 | V, \tilde{Y}))) = \mathbb{E}_{\vartheta_0}(s^V(\tilde{Y}, V, \vartheta_0)) = 0. \tag{2.15}$$

The estimator $\hat{\vartheta}$ obtained by solving $s^W(\tilde{Y}, W, \hat{\vartheta}) = 0$, such that $-\partial s^W(\tilde{Y}, W, \hat{\vartheta})/\partial\vartheta$ is positive definite, is called the corrected estimator for ϑ. Given ϑ is of finite dimension, under certain regularity conditions $\hat{\vartheta}$ is a consistent and asymptotically normal estimator of ϑ (Nakamura, 1990; Carroll et al., 2006, Appendix A.6). M-estimation techniques for the calculation of standard errors and for inference on $\hat{\vartheta}$ are used (Carroll et al., 2006, Chapter 7).

Stefanski (1989) shows that the concept of the corrected score function cannot be applied to the usual likelihood inference for the Cox model. The denominator of the partial estimation function in Equation (2.5) shows a (complex) singularity, thus an <u>exact</u> correction of the partial likelihood does not exist (Nakamura, 1992; Augustin and Schwarz, 2002). Therefore, Nakamura (1992) suggests an adjustment via a first and second order Taylor approximation under the assumption of homoscedastic normal measurement error. Kong and Gu (1999) show that the estimator suggested by Nakamura (1992) is consistent and has an asymptotic normal distribution. Kong et al. (1998) derive the corresponding corrected estimator of the cumulative baseline hazard rate. An extension to homoscedastic non-normal error is captured in Kong and Gu (1999).

Augustin (2004) shows that the concept of the first order estimator by Nakamura (1992) and the corrected baseline hazard rate estimator by Kong et al. (1998) can be applied to the *Breslow likelihood* (Breslow, 1972, 1974). An exact corrected likelihood approach is derived, which allows to estimate a corrected version of the regression coefficients β and the cumulative baseline hazard under heteroscedastic and non-normal measurement error. The suggested concept can also be applied to a form of regression models, where the hazard rate is parameterised, e. g. the Weibull model. The focus in the master's thesis is on the approach suggested by

Augustin (2004) and is described in the next section in detail. In addition, the corrected log-likelihood is extended for more than one error-prone covariate and for error-free covariables.

2.1.3. Corrected Log-Likelihood and Score Function in Cox and Weibull Model

Let $Y_i = min(T_i, C_i)$, $i = 1, \ldots, n$, be the censored failure times and δ_i be the indicator function for censoring. Let $\pi_1 < \pi_2 < \ldots < \pi_k$ be the different ordered failure times and be $d_j := |D_j(\pi_j)|$, $j = 1, \ldots, k$, the number of all subjects, which failed at π_j. If d_j is greater than 1, so-called *ties* exist in the data set. This means that more than one event occur at the same time. In such cases, the partial likelihood has to be corrected. Different methods exist to adjust for ties, here the correction by Peto (1972) and Breslow (1974) is considered, see Equation (2.18). $R(\pi_j)$ denotes the *risk set*, which includes all subjects, which are alive immediately before $\pi_j, j = 0, 1, \ldots, k, \pi_0 := 0$. A set of covariates is considered, which consists of covariates measured with and without error. Let $A_i = \begin{pmatrix} W_i \\ F_i \end{pmatrix}$ be the observed covariates vector, which consists of W_i subject to measurement error and underlie an additive error model from Equation (2.7) $W_i = V_i + U_i$ with $U_i \perp \{V_i, T_i, \delta_i\}$ and $U_i \perp U_l$ $(i \neq l)$, as well as F_i, which are free of measurement error. Let $X_i = \begin{pmatrix} V_i \\ F_i \end{pmatrix}$, which includes the true but unobservable covariables V_i. In matrix notation the design matrix is given by

$$X = \underbrace{\begin{pmatrix} 1 & V_{11} & \cdots & V_{1p} & F_{1p+1} & \cdots & F_{1m} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & V_{i1} & \cdots & V_{ip} & F_{ip+1} & \cdots & F_{im} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & V_{n1} & \cdots & V_{np} & F_{np+1} & \cdots & F_{nm} \end{pmatrix}}_{(n \times m+1)} \text{ and } X_i = \underbrace{\begin{pmatrix} 1 \\ V_{i1} \\ \vdots \\ V_{ip} \\ F_{ip+1} \\ \vdots \\ F_{im} \end{pmatrix}}_{(m+1 \times 1)}, \tag{2.16}$$

with m = number of covariates and n = number of observations.
The adjustment for measurement error via the moment generating function $M_{U_i}(\beta_X)$ is only required for the error-prone variables W_J, $J = 1, \ldots, p$. For the error-free

variables F_J, $J = p+1, \ldots, m$, the moment generating function is set to 1. β_X is a $(m+1)$-dimensional vector defined as follows

$$\beta_X = (\beta_0, \beta_{V_1}, \cdots, \beta_{V_p}, \beta_{F_{p+1}}, \cdots, \beta_{F_m})'. \tag{2.17}$$

In the case of the Cox model, X is defined without the column with 1s, X_i without the row with 1 and β_X without the intercept β_0, because the Cox model includes the intercept via the baseline hazard rate.

Cox Model

In Cox (1972, 1975) an estimation of the parameter vector β by maximizing the partial likelihood is suggested and is given by

$$L^X(\pi, X, \beta_X) = \prod_{j=1}^{k} \left(\frac{exp(\beta_X' X_j)}{\left[\sum_{i \in R(\pi_j)} exp(\beta_X' X_i)\right]^{d_j}} \right)^{\delta_j}, \tag{2.18}$$

where $\pi = (\pi_1, \ldots, \pi_k)$ and a correction for ties by Peto (1972) and Breslow (1974) is included. From the log-likelihood

$$l^X(\pi, X, \beta_X) = \sum_{j=1}^{k} \left[\sum_{i \in D(\pi_j)} \beta_X' X_i - d_j \cdot ln \left[\sum_{i \in R(\pi_j)} exp(\beta_X' X_i) \right] \right] \tag{2.19}$$

the estimation equation is derived as

$$s^X(\pi, X, \beta_X) = \sum_{j=1}^{k} \left[\sum_{i \in D(\pi_j)} X_i - d_j \cdot \frac{\sum_{i \in R(\pi_j)} X_i \cdot exp(\beta_X' X_i)}{\sum_{i \in R(\pi_j)} exp(\beta_X' X_i)} \right] = 0, \tag{2.20}$$

which is equal to Equation (2.5), besides the correction for ties. For inference on the baseline hazard rate $h_0(t)$ the *Breslow estimator* (Breslow, 1972, 1974) of the cumulative baseline hazard

$$\hat{H}_0^{Br}(t) = \sum_{j:\pi_j \leq t} \frac{d_j}{\sum_{i \in R(\pi_j)} exp(\hat{\beta}_X' X_i)} \tag{2.21}$$

with $H_0(t) = \int_0^t h_0(u)du$ is considered.

As already mentioned, the concept of corrected score function cannot be applied to the score function derived by the partial log-likelihood in Equation (2.20). Thus

Augustin (2004) used the *Breslow likelihood* for derivation of a corrected parameter estimator for β instead. The *Breslow likelihood* was developed by Breslow (1972, 1974) to justify the partial likelihood (2.18) and to obtain the baseline hazard estimator from Equation (2.21).

Assuming piecewise constant hazard rate $h_0(t) \equiv h_j > 0, \pi_{j-1} < t < \pi_j, j = 1, \dots, k$, the Breslow likelihood and log-likelihood is given, respectively, by

$$L^X_{Br}(\pi, X, \beta_X, h_j) = \prod_{i=1}^{n} \left[\left(h_0(t_i) exp(\beta'_X X_i) \right)^{\delta_i} exp\left(-exp(\beta'_X X_i) \cdot \int_0^{t_i} h_0(u)du \right) \right], \tag{2.22}$$

$$l^X_{Br}(\pi, X, \beta_X, h_j) = \sum_{j=1}^{k} \left[d_j ln(h_j) + \sum_{i \in D(\pi_j)} \beta'_X X_i - h_j(\pi_j - \pi_{j-1}) \sum_{i \in R(\pi_j)} exp(\beta'_X X_i) \right] \tag{2.23}$$

Corrected Log-Likelihood and Score Function of the Cox Model

Equation (2.23) does not possess any singularity, thus it is used for the derivation of the corrected score or log-likelihood function. In Augustin (2004) a corrected log-likelihood is considered, where in the Breslow log-likelihood X_i is replaced by A_i and the moment generating function is used for correction. In comparison to Augustin (2004) the corrected log-likelihood is extended for more than one error-prone variable. In addition, error-free covariables can be considered. Under the assumptions of the error model from Section 2.1.1 the corrected log-likelihood is given as follows

$$l^A_{corr}(\pi, A, \beta_X, h_j) = \sum_{j=1}^{k} \left[d_j ln(h_j) + \sum_{i \in D(\pi_j)} \beta'_X A_i - h_j(\pi_j - \pi_{j-1}) \sum_{i \in R(\pi_j)} \underbrace{\frac{exp(\beta'_X A_i)}{M_{U_i}(\beta_X)}}_{:=(*)} \right] \tag{2.24}$$

The correction for measurement error is only necessary for W_i. For the term (*) in Equation (2.24) this means in detail

$$
\begin{aligned}
(*) &= \frac{exp\left((\beta_{V_1}, \cdots, \beta_{V_p}, \beta_{F_{p+1}}, \cdots, \beta_{F_m}) \cdot A_i\right)}{M_{U_i}(\beta_X)} \\
&= \frac{exp\left(\beta_{V_1} W_{i1} + \ldots + \beta_{V_p} W_{ip} + \beta_{F_{p+1}} F_{ip+1} + \ldots + \beta_{F_m} F_{im}\right)}{M_{U_i}(\beta_X)} \\
&\overset{U_i \perp \{V_i, F_i\}}{=} \frac{exp(\beta_{V_1} W_{i1})}{M_{U_i}(\beta_{V_1})} \cdot \ldots \cdot \frac{exp(\beta_{F_m} F_{im})}{M_{U_i}(\beta_{F_m})} \\
&= exp(\beta_{V_1} W_{i1} - log[M_{U_i}(\beta_{V_1})]) \cdot \ldots \cdot exp(\beta_{F_m} F_{im} - log[M_{U_i}(\beta_{F_m})]) \\
&= exp(\beta_{V_1} W_{i1} - log[M_{U_i}(\beta_{V_1})] + \ldots + \beta_{F_m} F_{im} - log[\underbrace{M_{U_i}(\beta_{F_m})}_{=1}]) \\
&= exp(\beta_{V_1} W_{i1} - log[M_{U_i}(\beta_{V_1})] + \ldots + \beta_{V_p} W_{ip} - log[M_{U_i}(\beta_{V_p})] + \\
&\beta_{F_{p+1}} F_{ip+1} + \ldots + \beta_{F_m} F_{im}).
\end{aligned}
$$

The corresponding corrected score function $s^A_{corr_{\beta_X}}(\pi, A, \beta_X, h_j)$ is given by

$$
\sum_{j=1}^{k} \left[\sum_{i \in D(\pi_j)} A_i - \frac{d_j}{\sum_{i \in R(\pi_j)} exp(\beta'_X A_i)/M_{U_i}(\beta_X)} \cdot \sum_{i \in R(\pi_j)} K_i(\beta_X, A_i, M_{U_i}) \right] = 0, \tag{2.25}
$$

where $K_i(\beta_X, A_i, M_{U_i}) = exp(\beta'_X A_i)/M_{U_i}(\beta_X) \cdot \left(A_i - \frac{\partial}{\partial \beta_X} ln M_{U_i}(\beta_X)\right)$. The root of the corrected score function provides a corrected parameter estimation $\hat{\beta}^*_X$.
A corrected estimation of the cumulative baseline hazard $\hat{H}^*_0(t)$ is derived as

$$
\hat{H}^*_0(t) = \sum_{j: \pi_j \leq t} \frac{d_j}{\sum_{i \in R(\pi_j)} exp(\beta'_X A_i)/M_{U_i}(\hat{\beta}^*_X)}. \tag{2.26}
$$

The detailed calculations of $s^A_{corr_{\beta_X}}(\cdot)$ and $\hat{H}^*_0(t)$ are located in the Appendix B.2. It can be shown that the corrected log-likelihood satisfies condition (2.13) of being an unbiased estimation equation and accordingly the corrected score function fulfils the condition (2.12) with $\vartheta = (h_1, \ldots h_k, \beta_{V_1}, \cdots, \beta_{V_p}, \beta_{F_p+1}, \ldots, \beta_{F_m})$ and $\tilde{Y} = (min(T_i, C_i), \delta_i)$, $i = 1, \ldots, n$. The proof is given in the Appendix B.1.
In the case of normal homoscedastic errors $M_{U_i}(\beta) \equiv M_U(\beta)$ and is defined as

$$
M_U(\beta_X) = exp\left(\frac{1}{2} \beta'_X \Sigma_U \beta_X\right), \tag{2.27}
$$

with the m-dimensional vector $\beta_X = (\beta_{V_1}, \ldots, \beta_{V_p}, \beta_{F_{p+1}}, \ldots, \beta_{F_m})'$.

http://www.springer.com/978-3-658-08504-9

Model-Based Recursive Partitioning with Adjustment for Measurement Error
Applied to the Cox's Proportional Hazards and Weibull Model
Birke, H.
2015, XXIV, 240 p. 65 illus., Softcover
ISBN: 978-3-658-08504-9