



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Augustin:

An exact corrected log-likelihood function for Cox's
proportional hazards model under measurement error
and some extensions

Sonderforschungsbereich 386, Paper 277 (2002)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions

T. Augustin
University of Bielefeld

Abstract

This paper studies Cox's proportional hazards model under covariate measurement error. Nakamura's (1990) methodology of corrected log-likelihood will be applied to the so called Breslow likelihood, which is, in the absence of measurement error, equivalent to partial likelihood. For a general error model with possibly heteroscedastic and non-normal additive measurement error, corrected estimators of the regression parameter as well as of the baseline hazard rate are obtained. The estimators proposed by Nakamura (1992), Kong, Huang and Li (1998) and Kong and Gu (1999) are reestablished in the special cases considered there. This sheds new light on these estimators and justifies them as *exact* corrected score estimators. Finally, the method will be extended to some variants of the Cox model.

Key words: corrected log-likelihood, corrected score function, Cox

model, error-in-variables, measurement error, proportional hazards model.

1 Introduction and sketch of the main arguments

The most popular model for analyzing survival data is Cox's (1972) proportional hazard model. For every unit i , $i = 1, \dots, n$, the corresponding hazard rate $\lambda(t|X_i)$ is related to the vector X_i of covariates by

$$\lambda(t|X_i) = \lambda_0(t) \cdot \exp(X_i' \beta), \quad (1)$$

where throughout this paper X_i is assumed not to depend on time. The baseline hazard rate $\lambda_0(t)$ can be left completely unspecified making the model semiparametric and therefore quite flexible.

The common random censorship model is used: rather than always observing T_i , only the pair (Y_i, Δ_i) is available where $Y_i = \min(T_i, C_i)$ and Δ_i is the indicator function of $\{T_i \leq C_i\}$. The censoring variable C_i is stochastically independent of T_i and describes the maximal time span which unit i can be in the study.

Let $\tau_1 < \tau_2 < \dots < \tau_j < \dots < \tau_k$ be the distinct observed true failure times, ordered in increasing magnitude. Further let $\mathcal{D}(\tau_j)$, $j = 1, \dots, k$, with $|\mathcal{D}(\tau_j)| =: d_j$, be the set of all units failing at τ_j , and define the *risk set* $\mathcal{R}(\tau_j)$ to be the set of all units being alive immediately before τ_j , $j = 0, 1, \dots, k$, $\tau_0 := 0$.

Cox (1972) proposed to estimate the parameter vector β by maximizing the so called *partial likelihood*, which does not involve $\lambda_0(t)$. Following Peto's (1972) adjustment for ties, one has

$$\mathcal{PL} = \prod_{j=1}^k \frac{\exp(\sum_{i \in \mathcal{D}(\tau_j)} (\beta' X_i))}{[\sum_{i \in \mathcal{R}(\tau_j)} \exp(\beta' X_i)]^{d_j}}. \quad (2)$$

Then the *partial likelihood estimator* $\hat{\beta}_{\text{PL}}$ is obtained as the root of the deriva-

tive of

$$\ln(\mathcal{PL}) = \sum_{j=1}^k \left(\sum_{i \in \mathcal{D}(\tau_j)} \beta' X_i - d_j \ln \left(\sum_{i \in \mathcal{R}(\tau_j)} \exp(\beta' X_i) \right) \right) \quad (3)$$

with respect to β ; i.e., one has to solve

$$\sum_{j=1}^k \left(\sum_{i \in \mathcal{D}(\tau_j)} X_i - d_j \frac{\sum_{i \in \mathcal{R}(\tau_j)} X_i \cdot \exp(\hat{\beta}'_{PL} X_i)}{\sum_{i \in \mathcal{R}(\tau_j)} \exp(\hat{\beta}'_{PL} X_i)} \right) = 0. \quad (4)$$

For inference on the baseline hazard rate customarily the *Breslow estimator* (Breslow, 1972, 1974)

$$\hat{\Lambda}_0^{Br}(t) = \sum_{j: \tau_j \leq t} \frac{d_j}{\sum_{i \in \mathcal{R}(\tau_j)} \exp(\hat{\beta}'_{PL} \cdot X_i)} \quad (5)$$

of the cumulative baseline hazard rate $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ is used. From this an estimator for $\lambda_0(t)$ can be derived by appropriate smoothing (cf. Ramlau-Hansen, 1983). Asymptotic properties of the estimator are established in Andersen and Gill (1982) by embedding the problem into the theory of counting processes.

In many applications, the accurate values of the covariates X_i are unascertainable — surrogates W_i have to be used instead, for instance, imprecise measurements or improper operationalizations of X_i . Neglecting the difference between X_i and W_i , and just plugging in W_i instead of X_i , may lead to a severe bias in the analysis. Measurement error modeling, also called error-in-variables modeling, is serious about this fact and develops procedures to adjust for measurement error (see Carroll, Ruppert and Stefanski, 1995, and Cheng and Van Ness, 1999, for comprehensive monographs on this topic).

In the last years also a vivid discussion (e.g., Buzas, 1998, Hu, Tsiatis and Davidian, 1998, Kong, 1999, Huang and Wang, 2000, and the references below) has taken place how to adjust for measurement error in the Cox model (see also Augustin and Schwarz, 2002, for a survey and for a comparison of the basic methods).

Special attention has been paid to the work by Nakamura (1992), who had, in a previous paper (Nakamura, 1990), developed a powerful methodology for functional measurement error correction by the notion of corrected score functions and corrected log-likelihood (cf., also, Stefanski, 1989, Buonaccorsi, 1996). From Stefanski (1989) a necessary condition for the existence of a such a corrected function can be concluded: the underlying log-likelihood and score function have to be entire functions in the complex plane. This condition, however, is violated in the Cox model: the log-partial likelihood (3) as well as the estimating equation (4) derived from it possess singularities in the complex numbers. For this reason, Nakamura (1992) proposes approximate corrections based on the first and second order Taylor expansion of the left hand side in (4), which are entire functions. Using heuristic arguments, it is convincing that this procedure reduces the bias compared to naive estimation, and Nakamura also reported some simulations supporting this conjecture. Surprisingly, much stronger asymptotic properties hold: the estimator based on the first order correction is consistent and asymptotically normal, as was shown by Kong and Gu (1999). Additionally, Kong, Huang and Li (1998) managed to derive a corresponding estimator of the cumulative baseline hazard rate. Both result are extended in Kong and Gu (1999, Section 5) to non-normal homoscedastic measurement error.

The non-existence of an exact correction of partial likelihood estimation has generally been understood to imply that Nakamura's methodology is not applicable to the Cox model. This conclusion, however, is premature. As this paper shows, Nakamura's concept can be applied in a straightforward way to the so called Breslow likelihood, which is, in the absence of measurement error, equivalent to partial likelihood and provides the forgotten basis of the Breslow estimator of the cumulative baseline hazard rate. An exact corrected likelihood is obtained that allows for estimation of the regression coefficients β as well as of the cumulative baseline hazard rate under a general error model, which, for instance, contains (mixtures of) normals. In particular, a corrected cumulative baseline hazard rate estimator under heteroscedastic normal measurement error is derived.

Nakamura's first order corrected estimator and the baseline hazard rate estimator proposed in Kong et al. (1998) and Kong and Gu (1999) are reestablished in the cases considered there — a fact which gives new and strong support to these estimators. It is shown that Nakamura's first order estima-

tor can be understood as an *exact* corrected estimator, and so, for instance, its consistency becomes much less surprising. Moreover, the point of view taken here very much justifies the baseline hazard rate estimator introduced in Kong et al. (1998), which was gained there by genuine intuition. That estimator is not one estimator among possibly many others, it is *the* consequent generalization of the Breslow estimator arising from the methodology of corrected log-likelihood.

Additionally, the method proves successful in adjusting for measurement error in some related models, where the hazard rate is represented by a certain smooth function (cf., e.g., the polynomial approaches of Taulbee, 1979, Ciampi and Etezadi-Amoli, 1985, the spline approach of Whittemore and Keller, 1986, and the econometric models of Flinn and Heckman, 1982, and of Gritz, 1993).

This paper is organized as follows: Section 2 briefly describes the error model and recalls the basic ideas of the concept of corrected log-likelihood. Section 3 derives the exact corrected log-likelihood, which in Section 4 is applied to some special cases considered in the literature. In Section 5 the corrected log-likelihood for some variants of the Cox model is constructed.

2 Measurement error and the concept of corrected log-likelihood

2.1 Modeling the measurement error

Evidently, measurement error corrected inference is only possible under knowledge (or assumptions) on the relationship between the true, but unobservable variables X_i and the surrogates W_i .

Throughout this paper an additive error model is considered, i.e., for $i = 1, \dots, n$,

$$W_i = X_i + U_i \tag{6}$$

where U_i , with $\mathbb{E}(U_i) = 0$ and finite second moments, is independent of X_i ,

T_l and Δ_l , $l = 1, \dots, n$, and of U_l , $l \neq i$.¹ Additionally, the measurement errors U_i are taken to be independent among each other. The distribution of U_i must be known (or must have been estimated consistently, for instance from validation data). Here a mild additional assumption has to be made: the distributions are such that the moment generating functions $M_{U_i}(a) = \mathbb{E}(\exp(a'U_i))$ exist and are two times differentiable with respect to a , at least in a sufficiently large domain to make the expressions below well-defined for 'realistic' ranges of β and X_i .

This model covers the standard case of normal measurement error, but additionally offers a substantially broader scope. For instance, it contains all error distributions with bounded support as well as the very flexible class of finite mixtures of normals (see, for instance, Everitt & Hand, 1981, p. 28f.), who give an impression of the huge variety of different shapes which can be produced by the mixture of two normals).

Notice further that the variables U_i were not required to be identically distributed. Hence the method developed can also deal with heteroscedastic measurement error, which is often suggested by subject matter considerations (cf., for instance, Willett, 1998, pp. 33-48, for nutritional epidemiology).

2.2 Corrected log-likelihood

Consider the estimation of a parameter ϑ based on a sample $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \dots, \tilde{Y}_n)$ of possibly vector-valued observations in a regression model with covariates $(X_1, \dots, X_n) =: \mathbf{X}$ and surrogates $(W_1, \dots, W_n) =: \mathbf{W}$. Denote by $l^X(\tilde{\mathbf{Y}}, \mathbf{X}, \vartheta)$ and by $s^X(\tilde{\mathbf{Y}}, \mathbf{X}, \vartheta)$ the log-likelihood and the score function of ϑ , conditional on X_1, \dots, X_n , after having observed $\tilde{Y}_1, \dots, \tilde{Y}_n$. Since X_1, \dots, X_n are unobservable, $s^X(\tilde{\mathbf{Y}}, \mathbf{X}, \vartheta)$ will be called the *ideal score function* below.

Let $l^X(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta)$ and $s^X(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta)$ be the corresponding naive log-likelihood and the naive score function which arise from $l^X(\tilde{\mathbf{Y}}, \mathbf{X}, \vartheta)$ and $s^X(\tilde{\mathbf{Y}}, \mathbf{X}, \vartheta)$ by replacing \mathbf{X} with the matrix of surrogates \mathbf{W} . Then, $\mathbb{E}(s^X(\tilde{\mathbf{Y}}, \mathbf{X}, \vartheta)) =$

¹Note that the formulation is chosen such that it is not necessary to distinguish in notation between correctly measured and error-prone components of the covariate vector. If the j -th component $X_i[j]$ is measured without error then $U_i[j] \equiv 0$.

0 at the true parameter value ϑ_0 , but, in general, the absolute value of $\mathbb{E}(s^X(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta_0))$ is bounded away from zero, even for sample size going to infinity. As a consequence, the root of the naive score function $s^X(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta)$ is an inconsistent estimator of ϑ .

Measurement error correction removes this deficiency by constructing unbiased estimating functions in the observable variables $\tilde{\mathbf{Y}}$ and \mathbf{W} , i.e. functions $s^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta)$ such that $\mathbb{E}(s^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta_0)) = 0$. Then, under mild regularity conditions, standard asymptotic arguments (cf., for instance, Carroll, Rupert and Stefanski, 1995, Appendix A, Fahrmeir and Tutz, 2001, Appendix A2) can be applied to show that the corresponding root $\hat{\vartheta}$ is a \sqrt{n} -consistent and asymptotically normal estimator of ϑ .

Nakamura (1990) and Stefanski (1989) (cf. also Buonaccorsi, 1996) developed independently a general method to arrive at unbiased estimating functions. One of its main advantages is that it is a so called functional approach (cf., e.g., Carroll et al., 1995, Chapter 1), i.e., no assumptions concerning the distribution of the vector X_i of true covariates have to be made. The X_i 's may be non-random or random, but even if they are random their distribution does not play any role in the correction procedure.

The essential idea is to use the ideal score function

$$s^X(\tilde{\mathbf{Y}}, \mathbf{X}, \vartheta)$$

as a building block and to look for a function

$$s^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta)$$

with the property that

$$\mathbb{E}(s^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta) | \mathbf{X}, \tilde{\mathbf{Y}}) = s^X(\tilde{\mathbf{Y}}, \mathbf{X}, \vartheta). \quad (7)$$

Every such function is called a *corrected score function*.

Indeed, by construction and by the law of iterated expectation,

$$\mathbb{E}(s^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta_0)) = \mathbb{E}(\mathbb{E}(s^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta_0) | \mathbf{X}, \tilde{\mathbf{Y}})) = \mathbb{E}(s^X(\tilde{\mathbf{Y}}, \mathbf{X}, \vartheta_0)) = 0.$$

Often it is advantageous to get there by a circuitous route, using the log-likelihood and looking for a function $l^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta)$ with

$$\mathbb{E}(l^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta) | \mathbf{X}, \tilde{\mathbf{Y}}) = l^X(\tilde{\mathbf{Y}}, \mathbf{X}, \vartheta). \quad (8)$$

Then, again under appropriate regularity conditions,

$$s^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta) := \frac{\partial}{\partial \vartheta} l^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta) \quad (9)$$

is a corrected score function, and so $l^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta)$ is called a *corrected log-likelihood*.

More generally, the methodology applies in the same way, if $s^X(\tilde{\mathbf{Y}}, \mathbf{X}, \vartheta)$ is an arbitrary unbiased estimation function instead of the likelihood score function.

Within this framework also estimators of the asymptotic variance of the corrected estimator $\hat{\vartheta}$ are available from the standard theory of estimating functions (cf., e.g., the summaries in Serfling (1980, Chapter 7), Carroll, Ruppert and Stefanski, 1995, Appendix A).

3 The exact corrected log-likelihood for the Cox model — general form

In this section the corrected log-likelihood in its general form will be derived. To prepare for this, some early work on the Cox model has to be recalled. Breslow's (1972, 1974) likelihood construction, which was originally developed as a justification of (4) and to obtain a baseline hazard rate estimator, will prove to be of great importance in the context of measurement error modeling. It will allow a straightforward and powerful application of the methodology of corrected log-likelihood.

Breslow shifts, for every $j = 1, \dots, k$, all censoring times occurring in the interval $[\tau_{j-1}, \tau_j)$ to τ_{j-1} and assumes piecewise constant hazard,

$$\lambda_0(t) \equiv \lambda_j > 0, \quad \tau_{j-1} < t \leq \tau_j, \quad j = 1, \dots, k. \quad (10)$$

Then, based on the observed values t_1, \dots, t_n and $\delta_1, \dots, \delta_n$ of Y_1, \dots, Y_n and of $\Delta_1, \dots, \Delta_n$, the likelihood

$$\mathcal{L} = \prod_{i=1}^n \left((\lambda_0(t_i) \exp(\beta' X_i))^{\delta_i} \exp \left(- \exp(\beta' X_i) \cdot \int_0^{t_i} \lambda_0(u) du \right) \right) \quad (11)$$

arising under independent censoring from Cox's model (1) yields the log-likelihood

$$\ln \mathcal{L}^{Br} = \sum_{j=1}^k \left(d_j \ln \lambda_j + \sum_{i \in \mathcal{D}(\tau_j)} \beta' X_i - \lambda_j (\tau_j - \tau_{j-1}) \sum_{i \in \mathcal{R}(\tau_j)} \exp(\beta' X_i) \right). \quad (12)$$

Differentiating with respect to β and to $\lambda_1, \dots, \lambda_k$ leads indeed just to the partial likelihood equation (4) and to the Breslow estimator (5).

Notice that, in contrast to (3) and to (4), the log-likelihood from (12) does not contain any singularities ruling out the existence of the corrected log-likelihood or the corrected score function. Indeed, the methodology described in Section 2.2 can directly be applied with $\vartheta = (\beta_1, \dots, \beta_p, \lambda_1, \dots, \lambda_k)$ and $\tilde{Y}_i = (Y_i, \Delta_i)$, $i = 1, \dots, n$. This yields well-justified corrected estimators for the regression parameters as well as for the baseline hazard rate under heteroscedastic, possibly non-normal measurement error.

Theorem 3.1 *Under the error model from Section 2.1,*

$$l^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta) = \sum_{j=1}^k \left(d_j \ln \lambda_j + \sum_{i \in \mathcal{D}(\tau_j)} \beta' W_i - \lambda_j (\tau_j - \tau_{j-1}) \sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' W_i)}{M_{U_i}(\beta)} \right) \quad (13)$$

is a corrected log-likelihood for Breslow's log-likelihood (12).

The corresponding corrected estimators of the regression parameters β and the cumulative baseline hazard rate $\Lambda_0(t)$ are $\hat{\beta}^$ solving*

$$\sum_{j=1}^k \left(\sum_{i \in \mathcal{D}(\tau_j)} W_i - \frac{d_j}{\sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\hat{\beta}^* W_i)}{M_{U_i}(\hat{\beta}^*)}} \cdot \sum_{i \in \mathcal{R}(\tau_j)} K_i(\hat{\beta}^*; W_i, M_{U_i}) \right) = 0 \quad (14)$$

where

$$K_i(\beta; W_i, M_{U_i}) = \frac{\exp(\beta' W_i)}{M_{U_i}(\beta)} \cdot \left(W_i - \frac{\partial \ln M_{U_i}(\beta)}{\partial \beta} \right), \quad (15)$$

and $\hat{\Lambda}_0^(t)$ with*

$$\hat{\Lambda}_0^*(t) = \sum_{j: \tau_j \leq t} \frac{d_j}{\sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\hat{\beta}^* W_i)}{M_{U_i}(\hat{\beta}^*)}}. \quad (16)$$

Proof: For the first part one has to confirm (8), the second one then follows from (9). The details of the calculation are given in the Appendix. The essential point is that, by (6) and by the independence assumptions formulated in Section 2.1,

$$\mathbb{E} \left(\exp(\beta' W_i) | \mathbf{X}, \tilde{\mathbf{Y}} \right) = \exp(\beta' X_i) \cdot M_{U_i}(\beta)$$

and

$$\mathbb{E} \left(\beta' W_i | \mathbf{X}, \tilde{\mathbf{Y}} \right) = \beta' X_i.$$

4 Some applications of Theorem 3.1

In this section Theorem 3.1 will be applied to some special cases. This will distinguish several estimators which have previously been proposed in the literature as exact corrected log-likelihood estimators. Additionally, a solution for the practically important extension from homoscedastic to heteroscedastic normal measurement error will be developed.

4.1 Justification of previously proposed estimators

First the case of homoscedastic measurement error is considered, where the error variables U_i are identically distributed with $M_{U_i}(\beta) \equiv M_U(\beta)$. Then Theorem 3.1 justifies the proposal of Kong and Gu (1999, p. 964f.).

Corollary 4.1 (Homoscedastic measurement error) *In the case of homoscedastic measurement error, the estimators $\hat{\beta}^*$ and $\hat{\Lambda}_0^*(t)$ are obtained as the root of*

$$\sum_{j=1}^k \left(\sum_{i \in D(\tau_j)} W_i - d_j \left(\frac{\sum_{i \in \mathcal{R}(\tau_j)} W_i \exp(\beta' W_i)}{\sum_{i \in \mathcal{R}(\tau_j)} \exp(\beta' W_i)} - \frac{\frac{\delta M_U(\beta)}{\delta \beta}}{M_U(\beta)} \right) \right) = 0 \quad (17)$$

and as

$$\hat{\Lambda}_0^*(t) = \sum_{j: \tau_j \leq t} \frac{d_j \cdot M_U(\beta)}{\sum_{i \in \mathcal{R}(\tau_j)} \exp(\hat{\beta}^{*'} W_i)}, \quad (18)$$

respectively.

In the literature homoscedastic normal measurement error is customarily assumed, where $U_i \sim N(0, \Sigma_U)$, $i = 1, \dots, n$, $M_U(\beta) = \exp\left(\frac{1}{2}\beta' \Sigma_U \beta\right)$, and $\frac{\delta}{\delta \beta} \ln M_U(\beta) = \Sigma_U \cdot \beta$. Then one obtains

Corollary 4.2 (Homoscedastic normal measurement error) *If additionally all U_i 's are normally distributed with covariance Σ_U , then the estimator for β is the root $\hat{\beta}^*$ of*

$$\sum_{j=1}^k \left(\sum_{i \in \mathcal{D}(\tau_j)} W_i - d_j \left(\frac{\sum_{i \in \mathcal{R}(\tau_j)} W_i \cdot \exp(\beta' W_i)}{\sum_{i \in \mathcal{R}(\tau_j)} \exp(\beta' W_i)} - \Sigma_U \beta \right) \right) = 0. \quad (19)$$

The corresponding cumulative baseline hazard rate estimator is

$$\hat{\Lambda}_0^*(t) = \exp\left(\frac{1}{2} \hat{\beta}^{*'} \Sigma_U \hat{\beta}^*\right) \cdot \sum_{j: \tau_j \leq t} \frac{d_j}{\sum_{i \in \mathcal{R}(\tau_j)} \exp(\hat{\beta}^{*'} W_i)}. \quad (20)$$

Corollary 4.2 is puzzling in this context. The estimating equation for β resulting from Theorem 3.1 coincides with Nakamura's (1992) proposal: what was derived as a crude first order approximation now turns out to be an *exact* corrected score function. This helps to understand the appealing mathematical properties like consistency and asymptotic normality as well as the astonishing power in a comparative simulation study (Augustin and Schwarz, 2002, Section 5).

The second lesson from Corollary 4.2 addresses baseline hazard rate estimation. It distinguishes the estimator (20), which was introduced by Kong et al. (1998) and by Kong and Gu (1999), as *the* consequent generalization of the Breslow estimator to the situation under homoscedastic normal measurement error.

4.2 Heteroscedastic normal measurement error

Keeping the assumption of normality, but returning to the practically very important situation of heteroscedastic measurement error, i.e. considering the case $U_i \sim \mathcal{N}(0, \Sigma_{U_i})$, a straightforward extension of (19) would be to use the estimating equation

$$\sum_{j=1}^k \left(\sum_{i \in \mathcal{D}(\tau_j)} (W_i + \Sigma_{U_i} \beta) - d_j \frac{\sum_{i \in \mathcal{R}(\tau_j)} W_i \cdot \exp(\beta' W_i)}{\sum_{i \in \mathcal{R}(\tau_j)} \exp(\beta' W_i)} \right) = 0 .$$

However, the methodology of corrected log-likelihood leads to another proposal. Applying Theorem 3.1 yields the estimating equation

$$\begin{aligned} \sum_{j=1}^k \left(\sum_{i \in \mathcal{D}(\tau_j)} W_i - \frac{d_j}{\sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' W_i)}{\exp(0.5 \beta' \Sigma_{U_i} \beta)}} \right. \\ \left. \cdot \sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' W_i)}{\exp(0.5 \beta' \Sigma_{U_i} \beta)} \cdot (W_i - \Sigma_{U_i} \beta) \right) = 0 \end{aligned} \quad (21)$$

for the regression parameter β .

Additionally, Theorem 3.1 tells how to extend (20) appropriately to obtain a corrected estimator $\hat{\Lambda}_0^*(t)$ of the cumulative baseline hazard rate. The corresponding generalization of the Breslow estimator is

$$\hat{\Lambda}_0^*(t) = \sum_{j: \tau_j \leq t} \frac{d_j}{\sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' W_i)}{\exp(0.5 \beta' \Sigma_{U_i} \beta)}} .$$

5 Measurement error correction in some variants of the Cox model

The likelihood based approach taken in this paper also proves successful in dealing with some variants and generalizations of the Cox model. Straightforward is the extension to the case considered in Kalbfleisch and Prentice

(1973) (cf. also Kay, 1977, p. 229), where the intervals of constant hazard are given exogenously in advance.

Several authors have studied variants of the Cox model where in (1) the baseline hazard rate is modeled as

$$\lambda_0(t) = g(\eta, t), \quad (22)$$

with a smooth, known function $g(\cdot)$ and a parameter vector η of finite dimension. Specific examples motivated by biometrical applications are the polynomial approaches of Taulbee (1979) and of Ciampi and Etezadi-Amoli (1985), as well as the spline approach of Whittemore and Keller (1986), each developed to obtain a flexible but smooth estimator of the baseline hazard rate. Representations in the spirit of (22) are also of great interest in econometrics, where they are used to model and test specific dynamic effects in labor force data (cf., e.g., Flinn and Heckman, 1982, Gritz, 1993).

Defining $G(\eta, t) = \int_0^t g(\eta, u) du$, and assuming again time-independent covariates, the likelihood (11) of η and β becomes

$$\mathcal{L} = \prod_{i=1}^n \left(g(\eta, t_i) \cdot \exp(\beta' X_i) \right)^{\delta_i} \cdot \exp \left(- \exp(\beta' X_i) \cdot G(\eta, t_i) \right) \quad (23)$$

leading to the log-likelihood

$$\ln \mathcal{L} = \sum_{i=1}^n \delta_i \cdot \left(\ln(g(\eta, t_i)) + \beta' X_i \right) - \exp(\beta' X_i) \cdot G(\eta, t_i). \quad (24)$$

This likelihood is such that the techniques used in the proof of Theorem 3.1 can be applied in a very similar way. This allows powerful measurement error corrections in all the variants characterized and exemplified above.

Prop. 5.1 *Under the error model from Section 2.1,*

$$\sum_{i=1}^n \delta_i \cdot \left(\ln(g(\eta, t_i)) + \beta' W_i \right) - \frac{\exp(\beta' W_i)}{M_{U_i}(\beta)} \cdot G(\eta, t_i) \quad (25)$$

is a corrected log-likelihood for the log-likelihood (24).

Finally, the parameterization

$$g(\eta, t) = \eta \cdot t^{\eta-1}$$

with real-valued, nonnegative η deserves some attention. In this case the Cox model (1) specializes to the Weibull model with dependence parameter η and covariate effects $\eta^{-1} \cdot \beta$. Then Proposition 5.1 extends the results of Gimenez, Bolfarine and Colosimo (1999) on Nakamura's correction method in the Weibull model under homoscedastic normal measurement error to the general error model considered here.

6 Concluding remarks

This paper applies the method of corrected log-likelihood to Breslow's likelihood for the Cox model, yielding measurement error corrected estimators under a general error model containing heteroscedastic and non-normal measurement error. In the special cases which have been considered in the literature, the estimators proposed there are reestablished; they obtain a neat justification as exact corrected score estimators.

The correction procedure belongs to the class of functional methods which manage to adjust for measurement error without making any distributional assumptions on the true, but unobservable vector of covariates. This may be of great importance, for instance, in biometrical application where exposure variables often are distributed in a way which is difficult to be handled by standard distributions.

The main result was given in terms of the moment generating function of the error distribution. On the one hand, this general form allows for a very flexible modeling of the error, if knowledge on its distribution is present (for instance in the case of validation data). On the other hand, if such knowledge is lacking, Theorem 3.1 provides a rich framework to perform a comprehensive sensitivity analysis.

Acknowledgements I am grateful to Helmut Küchenhoff and Hans Schneeweiß for their very valuable comments.

References

- ANDERSEN, P.K. and GILL, R.D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-1120.
- AUGUSTIN, T. and SCHWARZ, R. (2002). Cox's proportional hazards model under covariate measurement error — A review and comparison of methods. In: S. Van Huffel and P. Lemmerling (eds.). *Total least squares and errors-in-variables modeling: analysis, algorithms and applications*. Kluwer, Dordrecht, 175-184.
- BUONACCORSI, J.P. (1996). A modified estimating equation approach to correcting for measurement error in regression. *Biometrika* **83**, 433-440.
- BRESLOW, N.E. (1972). Contribution to the discussion of Cox (1972). *J. Roy. Stat. Soc. Ser. B Stat. Methodol.* **34**, 216-217.
- BRESLOW, N.E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89-99.
- BUZAS, J.S. (1998). Unbiased scores in proportional hazards regression with covariate measurement error. *J. Statist. Plann. Inference* **67**, 247-257
- CARROLL, R.J. and RUPPERT, D. and STEFANSKI, L.A. (1995). *Measurement error in nonlinear models*. Chapman and Hall, London.
- CIAMPI, A. and ETEZADI-AMOLI, J. (1985). A general model for testing the proportional hazards and the accelerated failure time hypothesis in the analysis of censored survival data with covariates. *Comm. Statist. Theory Methods* **14**, 651-667.
- CHENG, C.-L. and VAN NESS, J.W. (1999). *Statistical regression with measurement error*. Arnold, London.
- COX, D.R. (1972). Regression models and life tables (with discussion). *J. Roy. Stat. Soc. Ser. B Stat. Methodol.* **34**, 187-220.
- EVERITT, B.S. and HAND, D.J. (1981). *Finite mixture distributions*. Chapman and Hall, London.

- FAHRMEIR, L. and TUTZ, G. (2001). *Multivariate statistical models based on generalized linear models*. Springer, New York, second edition.
- FLINN, C.J. and HECKMAN, J.J. (1982). Models for the analysis of labor force dynamics. *Advances in Econometrics* **1**, 35-95.
- GIMENEZ, P. and BOLFARINE, H., and COLOSIMO, E.A. (1999). Estimation in Weibull regression model with measurement error. *Comm. Statist. Theory Methods* **28**, 495-510.
- GRITZ, M. (1993); The impact of training on the frequency and duration of employment. *J. Econometrics* **57**, 21-51.
- HU, P. and TSIATIS, A. and DAVIDIAN, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* **54**, 1407-1419.
- HUANG, Y. and WANG, C.Y. (2000). Cox regression with accurate covariates unascertainable: A nonparametric-correction approach, *J. Amer. Statist. Assoc.* **45**, 1209-1219
- KALBFLEISCH, J.D. and PRENTICE, R.L. (1973). Marginal likelihood based on Cox's regression and life model. *Biometrika* **60**, 267-278.
- KAY, R. (1977). Proportional hazard regression models and the analysis of censored survival data. *J. Roy. Statist. Soc. Ser. C* **26**, 227-237.
- KONG, F.H. (1999). Adjusting regression attenuation in the Cox proportional hazards model, *J. Statist. Plann. Inference* **79**, 31-44.
- KONG, F.H. and GU, M. (1999). Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statistica Sinica* **9**, 953-969.
- KONG, F.H. and HUANG, W. and LI, X. (1998). Estimating survival curves under proportional hazards model with covariate measurement errors. *Scand. J. Statist.* **25**, 573-587.
- NAKAMURA, T. (1990). Corrected score functions for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika* **77**, 127-137.

- NAKAMURA, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics* **48**, 829–838.
- PETO, R. (1972). Contribution to the discussion of Cox (1972). *J. Roy. Statist. Soc. Ser. B Stat. Methodol.* **34**, 205–207.
- RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 804–813.
- SERFLING, R. J. (1980). *Approximation theorems of mathematical statistics*. Wiley, New York.
- STEFANSKI, L. A. (1989). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Comm. Statist. Theory Methods* **18**, 4335–4358.
- TAULBEE, J. D. (1979). A general model for the hazard rate with covariables. *Biometrics* **35**, 439–450.
- WHITTEMOORE, A. S. and KELLER, J. B. (1986). Survival estimation using splines. *Biometrics* **42**, 495–506.
- WILLETT, W. (1998). *Nutritional epidemiology*. Oxford University Press. Oxford, second edition.

Thomas Augustin
Institute for Population Research and Social Policy Research
University of Bielefeld
Box 100131
D – 33501 Bielefeld
Germany
thomas@stat.uni-muenchen.de

Appendix: Detailed Proof of Theorem 3.1

Firstly, (8) has to be confirmed.

$$\begin{aligned}
& \mathbb{E} \left(l^w(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta | \mathbf{X}, \tilde{\mathbf{Y}}) \right) \\
&= \mathbb{E} \left(\sum_{j=1}^k \left(\left(d_j \ln \lambda_j + \sum_{i \in \mathcal{D}(\tau_j)} \beta' W_i \right. \right. \right. \\
&\quad \left. \left. \left. - \lambda_j(\tau_j - \tau_{j-1}) \sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' W_i)}{M_{U_i}(\beta)} \right) \middle| \mathbf{X}, \tilde{\mathbf{Y}} \right) \right) . \\
&= \mathbb{E} \left(\sum_{j=1}^k \left(d_j \ln \lambda_j + \sum_{i \in \mathcal{D}(\tau_j)} \beta' (X_i + U_i) \right. \right. \\
&\quad \left. \left. - \lambda_j(\tau_j - \tau_{j-1}) \cdot \sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' (X_i + U_i))}{M_{U_i}(\beta)} \right) \middle| \mathbf{X}, \tilde{\mathbf{Y}} \right) \\
&= \sum_{j=1}^k \left(d_j \ln \lambda_j + \sum_{i \in \mathcal{D}(\tau_j)} \beta' X_i + \sum_{i \in \mathcal{D}(\tau_j)} \mathbb{E}(\beta' U_i) | \mathbf{X}, \tilde{\mathbf{Y}} \right. \\
&\quad \left. - \lambda_j(\tau_j - \tau_{j-1}) \cdot \sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' X_i)}{M_{U_i}(\beta)} \mathbb{E}(\exp(\beta' U_i)) | \mathbf{X}, \tilde{\mathbf{Y}} \right) .
\end{aligned}$$

Because U_i is independent of X_i , T_l and Δ_l , $l = 1, \dots, n$,

$$\mathbb{E} \left(\exp(\beta' U_i) | \mathbf{X}, \tilde{\mathbf{Y}} \right) = \mathbb{E}(\exp(\beta' U_i)) = M_{U_i}(\beta)$$

and

$$\mathbb{E} \left(\beta' U_i | \mathbf{X}, \tilde{\mathbf{Y}} \right) = \mathbb{E}(\beta' U_i) = 0 .$$

Therefore

$$\begin{aligned} \mathbb{E} \left(l^w(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta | \mathbf{X}, \tilde{\mathbf{Y}}) \right) &= \sum_{j=1}^k \left(d_j \ln \lambda_j + \sum_{i \in \mathcal{D}(\tau_j)} \beta' X_i \right. \\ &\quad \left. - \lambda_j (\tau_j - \tau_{j-1}) \cdot \sum_{i \in \mathcal{R}(\tau_j)} \exp(\beta' X_i) \right) \\ &= \ln \mathcal{L}^{Br}, \end{aligned}$$

and (8) is shown.

The second part applies the fact that, according to (9) corrected score functions can be derived via differentiation with respect to the parameters. Starting with $\lambda_1, \dots, \lambda_k$, one has

$$\frac{\partial l^w(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta)}{\partial \lambda_j} = \left(\frac{d_j}{\lambda_j} - (\tau_j - \tau_{j-1}) \cdot \sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' W_i)}{M_{U_i}(\beta)} \right), \quad j = 1, \dots, k,$$

and the corresponding root is

$$\hat{\lambda}_j = \frac{d_j}{(\tau_j - \tau_{j-1}) \cdot \sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' W_i)}{M_{U_i}(\beta)}}. \quad (26)$$

Then, in dependence on β ,

$$\hat{\Lambda}_0(t) = \int_0^t \sum_{j=1}^k \hat{\lambda}_j \cdot \mathbf{1}_{\{\tau_{j-1} < u \leq \tau_j\}} du = \sum_{j: \tau_j \leq t} \frac{d_j}{\sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' W_i)}{M_{U_i}(\beta)}}. \quad (27)$$

For the regression parameter one obtains

$$\begin{aligned} \frac{\partial l^W(\tilde{\mathbf{Y}}, \mathbf{W}, \vartheta)}{\partial \beta} &= \left(0 + \sum_{i \in \mathcal{D}(\tau_j)} W_i - \lambda_j (\tau_j - \tau_{j-1}) \right) \\ &\quad \cdot \sum_{i \in \mathcal{R}(\tau_j)} \frac{W_i \cdot \exp(\beta' W_i) \cdot M_{U_i}(\beta) - \frac{\delta}{\delta \beta} M_{U_i}(\beta) \exp(\beta' W_i)}{(M_{U_i}(\beta))^2}, \end{aligned}$$

which yields, together with (26), the estimating function

$$\sum_{j=1}^k \left(\sum_{i \in \mathcal{D}(\tau_j)} W_i - \frac{d_j}{\sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' W_i)}{M_{U_i}(\beta)}} \sum_{i \in \mathcal{R}(\tau_j)} \exp(\beta' W_i) \left(\frac{W_i}{M_{U_i}(\beta)} - \frac{\partial \ln M_{U_i}(\beta)}{\partial \beta} \right) \right).$$