Augustin, Schwarz:

Cox's Proportional Hazards Model under Covariate Measurement Error - A Review and Comparison of Methods

Sonderforschungsbereich 386, Paper 260 (2001)

COX'S PROPORTIONAL HAZARDS MODEL UNDER COVARIATE MEASUREMENT ERROR

A Review and Comparison of Methods

Thomas Augustin and Regina Schwarz
Department of Statistics, University of Munich
D 80799 Munich, Germany.
thomas@stat.uni-muenchen.de

Abstract This contribution studies the Cox model under covariate measurement error. Methods proposed in the literature to adjust for measurement error are reviewed. The basic structural and functional approaches are discussed in some detail, important modifications and further developments are briefly sketched. Then the basic methods are compared in a simulation study.

Keywords: Measurement error, error-in-variables, Cox model, survival analysis

1. Introduction

Probably the most common statistical model in biometrics is Cox's proportional hazards model for analyzing survival data. Searching in the Medline database yields about 8000 entries for the term 'proportional hazards' during the last ten years. In most biometric studies, measurement error is an important issue. Often variables of interest can not be measured without substantial error, often they are even not available in principle (like the average protein intake over the last five years). Surrogates have to be used instead. The naive estimate, which just plugs in the surrogate instead of the true covariate, may be expected to be severely biased. Therefore, several methods have been developed to remove the bias by taking the measurement error appropriately into account. This paper reviews them and compares the basic approaches underlying them. For this purpose, we proceed as follows: Section 2 collects some facts on Cox's proportional hazards model, Section 3 states precisely the basic form of the error models underlying this study. The literature on covariate measurement error in the Cox model is surveyed in Section 4, where special attention is paid to those methods which have laid the foundations for further developments. In Section 5 several correction methods are compared by a simulation study.

2. Cox's proportional hazards model

According to Cox [9], for every unit $i = 1, \dots, n$, the corresponding hazard rate $\lambda(t) = \lim_{\varepsilon \downarrow 0} \varepsilon^{-1} \cdot P(\{t \le T_i \le t + \varepsilon\} \,|\, \{T_i \ge t\})$ of the failure time T_i is related to the p-dimensional vector X_i of covariates by

$$\lambda(t|X_i) = \lambda_0(t) \cdot \exp(\beta^T X_i)\,. \tag{1}$$

The hazards of the units are proportional to each other, because the baseline hazard rate $\lambda_0(t)$ is assumed to be the same for all i. It can be left completely unspecified, making the model a semiparametric and therefore quite flexible tool.

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 $\{T_i \le C_i\}$. 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. Assume that no ties occur, and order the observed true failure times in increasing magnitude: $\tau_1 < \tau_2 < \ldots < \tau_j < \ldots < \tau_k$, $\tau_0 := 0$. Define for $j = 0, 1, \dots, k$ the *risk set* $\mathcal{R}(\tau_j)$ to be the set of all units being alive immediately before τ_j.

Estimation of the parameter vector β is based on the so called *partial likelihood*, which does not involve $\lambda_0(t)$. The *partial likelihood estimate* $\hat\beta_{\mathrm{PL}}$ is then obtained as the root of

$$\sum_{j=1}^{k} \left(X_j - \frac{\sum_{i \in \mathcal{R}(\tau_j)} X_i \cdot \exp\left(\beta^T X_i\right)}{\sum_{i \in \mathcal{R}(\tau_j)} \exp\left(\beta^T X_i\right)} \right) = 0\,. \tag{2}$$

For inference on the baseline hazard rate customarily the *Breslow estimate* (cf. [4, 5]) of the cumulative baseline hazard rate $\Lambda_0(t) = \int_0^t \lambda(u)du$ is used.

3. The basic error model

Unless it is explicitly mentioned the classical, homoscedastic error model in its basic form is considered throughout the paper: Take all true covariates X_i to be continuous and assume that the surrogates W_i are related to X_i by $W_i = X_i + U_i$ .The measurement error U_i is required to be independent of T_j, X_j, $j = 1, \dots, n$, as well as of $U_j, j \ne i$. Furthermore, the variables U_i are assumed to be i.i.d. normally distributed with mean zero and known (or consistently estimated) covariance matrix Σ_U.[1] The dependent variables Y_i and Δ_i are taken to be error free.

Altogether this leads to *nondifferential measurement error*: given X_i, the variables T_i and W_i are conditionally independent. This means that knowledge of X_i would make observation of W_i superfluous; W_i really is only a surrogate providing no information which is not contained in X_i.

4. Correction methods for the Cox model

Apart from some comments in Section 6 on baseline hazard rate estimation, we concentrate on measurement error corrected inference on the regression coefficients β. To structure this part of the presentation, we order the different methods according to the assumptions they require for the distribution of the true, unknown covariate X_i. We begin with the *structural approaches*, where the distribution of X_i is assumed to belong to a known class of parametric distributions. Then we turn to the *functional methods* which manage to do without any parametric assumption on the distribution law of the X_is.

4.1 Structural approaches

Generally there are two basic structural approaches: regression calibration, which will be discussed at the end of this subsection, and 'integrating the likelihood.'

The latter one uses the conditional distribution of X_i given W_i and the assumption of non-differentiality to integrate out the influence of the measurement error. Prentice [20] made it clear that applying this idea to the Cox model leads to unexpected difficulties: Under non-differential measurement error, which can be shown to be equivalent to $\lambda(t|X_i, W_i) = \lambda(t|X_i)$, one obtains in general

$$\begin{aligned}\lambda(t|W_i) &= \lim_{\varepsilon\downarrow 0} \varepsilon^{-1} \cdot P(\{T_i \le t+\varepsilon\}|\{T_i \ge t\}, W_i)\\ &= \lim_{\varepsilon\downarrow 0} \varepsilon^{-1} \cdot \mathbf{E}\left(P(\{T_i \le t+\varepsilon\}|X_i, \{T_i \ge t\}, W_i)\big|\{T_i \ge t\}, W_i\right)\\ &= \lim_{\varepsilon\downarrow 0} \varepsilon^{-1} \cdot \mathbf{E}\left(P(\{T_i \le t+\varepsilon\}|\{T_i \ge t\}, X_i)\big|\{T_i \ge t\}, W_i\right)\\ &= \mathbf{E}(\lambda(t|X_i)|\{T_i \ge t\}, W_i)\,,\end{aligned}$$

and for the Cox model

$$\lambda(t|W_i) = \lambda_0(t) \cdot \mathbf{E}(\exp(\beta^T X_i)|W_i, \{T_i \ge t\})\,. \tag{3}$$

Note that, via the event $\{T_i \ge t\}$ appearing in the condition, the second factor, the so called induced relative risk, depends on the previous history of the process. Because of this complex dependence on the unknown baseline hazard rate, the characteristic form of (1) is lost, and partial likelihood estimation can not be directly applied any more.[2]

However, as Prentice also argued, the effect of this time dependence can be expected to be small if the failure intensity is very low, because then the condition $\{T_i \geq t\}$ is almost always satisfied. Under this so called *rare disease assumption*,

$$\begin{aligned} \lambda(t|W_i) = \mathbf{E}\left(\lambda(t|X_i)|\{T_i \geq t\}, W_i\right) &\approx \mathbf{E}\left(\lambda(t|X_i)|W_i\right) \\ &= \lambda_0(t) \cdot \mathbf{E}\left(\exp(\beta^T X_i)|W_i\right) , \end{aligned} \tag{4}$$

and the induced relative risk can be explicitly calculated in some cases approximately. (See [19, p. 1170] for a brief discussion of the exactness of this approximation.)

If $X_i|W_i$ is normal with mean $\bar{\mu}_i$ and common covariance $\bar{\Sigma}$, then

$$\lambda(t|W_i) = \lambda_0(t) \cdot \exp(\beta^T \bar{\mu}_i + 0.5\beta^T \bar{\Sigma}\beta) =: \lambda_0^*(t) \cdot \exp(\beta^T \bar{\mu}_i) \,. \tag{5}$$

One important example is the situation where X_i itself is i.i.d. normally distributed, with unknown mean μ_X and non-singular covariance matrix Σ_X. Then $W_i \sim \mathcal{N}\left(\mu_X; \Sigma_X + \Sigma_U\right)$, and indeed $X_i|W_i \sim \mathcal{N}\left(\bar{\mu_i}, \bar{\Sigma}\right)$, with

$$\bar{\mu_i} = \mu_X + \Sigma_X \cdot \left(\Sigma_X + \Sigma_U\right)^{-1} \cdot \left(W_i - \mu_X\right) \tag{6}$$

and $\bar{\Sigma} = \Sigma_X - \Sigma_X \cdot \left(\Sigma_X + \Sigma_U\right)^{-1} \Sigma_X$. Now (5) reads as

$$\begin{aligned} \lambda(t|W_i) &= \lambda_0^*(t) \cdot \exp(\beta^T \mu_X + \beta^T \Sigma_X \cdot \left(\Sigma_X + \Sigma_U\right)^{-1} \cdot \left(W_i - \mu_X\right) \\ &=: \lambda_0^{**}(t) \cdot \exp(\beta^T \cdot \Sigma_X \cdot \left(\Sigma_X + \Sigma_U\right)^{-1} \cdot W_i). \end{aligned}$$

This shows that, under the assumptions stated above, the simple attenuation factor known from linear regression also applies for the Cox model; the corrected estimate $\hat{\beta}_{corr}$ is

$$\hat{\beta}_{corr} = \Sigma_X^{-1} \cdot \left(\Sigma_X + \Sigma_U\right) \cdot \hat{\beta}_{naive} \ . \tag{7}$$

Notice further that, given the measurement error covariance Σ_U, the nuisance parameters μ_X and Σ_X can be efficiently estimated from the observations $W_1, \ldots, W_n$.

[13] arrived at (7) as an ad-hoc proposal, motivated by the attenuation known from linear regression. In a series of simulations he observes a strong dependence of the bias on the true β and on the amount of censoring. Both issues are in accordance with the deviation given above: the lower, ceteris paribus, the true β and the higher the proportion of censored observation are, the better the rare disease assumption (4) is satisfied.

Probably the most universal tool for measurement error correction is *regression calibration* (c.f., e.g., [7, Chapter 3]). One uses the knowledge of W_i to predict X_i and replaces X_i by its expectation given W_i. In general, the estimates derived are not necessarily consistent, but the bias is considerably reduced. The main advantage of the method is its easy implementation; simply by proceeding with $\mathbf{E}(X_i|W_i)$ instead of X_i in the calculation, the estimates are obtainable by standard software.

In the presence of validation data, regression calibration can be used in a functional way, because then $\mathbf{E}(X_i \mid W_i)$ can be estimated in a non-parametric manner. [21] elaborated this idea for the Cox model. If no validation data are available, structural modeling is necessary. In the simplest case also considered above, where X_i is i.i.d. normally distributed, regression calibration coincides with the method discussed above: Substituting in Equation (2) the variable X_i by its conditional expectation $\bar{\mu_i}$ from (6) yields the estimating equation

$$\sum_{j=1}^{k}\left(\bar{\mu_j} - \frac{\sum_{i\in\mathcal{R}(\tau_j)} \bar{\mu_i}\cdot\exp\left(\beta^T\bar{\mu_i}\right)}{\sum_{i\in\mathcal{R}(\tau_j)}\exp\left(\beta^T\bar{\mu_i}\right)}\right) = 0\,.$$

After simplification and multiplication with $\left(\Sigma_X\cdot(\Sigma_X+\Sigma_U)^{-1}\right)^{-1}$ one obtains

$$\sum_{j=1}^{k}\left(W_j - \frac{\sum_{i\in\mathcal{R}(\tau_j)} W_i\cdot\exp\left(\beta^T\Sigma_X\cdot(\Sigma_X+\Sigma_U)^{-1}W_i\right)}{\sum_{i\in\mathcal{R}(\tau_j)}\exp\left(\beta^T\Sigma_X\cdot(\Sigma_X+\Sigma_U)^{-1}W_i\right)}\right) = 0\,, \quad (8)$$

which indeed leads again to (7).

Since assumptions on the distribution of the latent variable may influence the behaviour of the estimates, more flexible models for the distribution of X_i, for instance mixtures of normals, may be very attractive. The main arguments given in [2] to develop measurement error corrected quasi-likelihood estimation carry over to the situation considered here. In this generalized setting, both structural approaches won't be equivalent any more.

Methods related to regression calibration are studied and developed further in [19, 8, 13]. As mentioned above, [21] integrate this approach into a model, where the conditional distribution needed to adjust for measurement error is estimated from validation data. An alternative way to incorporate validation data is discussed in [23].

Under the assumption of piecewise constant hazard rates and based on numerical integration, [11] developed three likelihood based methods, which differ with respect to the modelling of the covariate distribution (nonparametric, semiparametric and parametric.)

4.2 Functional approaches

There are also several strictly functional approaches. The most popular one is [18], where Nakamura adopts his general methodology of corrected score functions (cf. [17]) to the Cox model. Nakamura's basic idea is to construct unbiased estimating equations by looking for a function in the observable quantities $\vec{Y} := (Y_1, \dots, Y_n)$, $\vec{\Delta} := (\Delta_1, \dots, \Delta_n)$ and $\mathbf{W} := (W_1, \dots, W_n)$ such that the conditional expectation given X_i is equal to the original score function. Then, by the theorem of iterative expectation, the overall expectation of this so called *corrected score function* equals zero, from which, under mild regularity conditions, consistency and asymptotic normality of the resulting estimate can be derived. This general framework, however, can not be directly applied to Equation (2) for partial likelihood estimation in the Cox model: the fact that the denominator possesses a (complex) singularity makes the existence of a corrected score function impossible.

Nakamura [18] therefore proposes an approximate solution based on a first and second order Taylor approximation of the fraction in (2). Denoting the naive estimating function by $\Psi^{-}(\vec{Y}, \vec{\Delta}, \mathbf{W}; \beta)$ and defining

$$K_j(\mathbf{W}) = 1 - \left(\sum\nolimits_{i \in \mathcal{R}(\tau_j)} \exp\left(2W_i{}^T\beta\right)\right) \cdot \left(\sum\nolimits_{i \in \mathcal{R}(\tau_j)} \exp\left(W_i{}^T\beta\right)\right)^{-2},$$

one obtains the first order corrected estimating function $\Psi_1(\vec{Y}, \vec{\Delta}, \mathbf{W}; \beta)$ and the second order correction $\Psi_2(\vec{Y}, \vec{\Delta}, \mathbf{W}; \beta)$ as

$$\begin{aligned} \Psi_1\left(\vec{Y}, \vec{\Delta}; \mathbf{W}; \beta\right) &= \Psi^{-}\left(\vec{Y}, \vec{\Delta}; \mathbf{W}; \beta\right) + \Sigma_U \cdot \beta \qquad (9) \\ \Psi_2\left(\vec{Y}, \vec{\Delta}; \mathbf{W}; \beta\right) &= \Psi^{-}\left(\vec{Y}, \vec{\Delta}; \mathbf{W}; \beta\right) + \sum\nolimits_{j=1}^{k} \left(K_j(\mathbf{W})\right) \cdot \Sigma_U \cdot \beta \end{aligned}$$

[15] study asymptotic properties of the resulting first order estimate and suggest an extension to non-normal measurement error. [3] gives a justification of (9) as an exact corrected likelihood estimate for Breslow's (cf. [4, 5]) likelihood approach to the Cox model. Based on replication data, [12] develop a nonparametric correction method which manages to do without parametric assumptions on the measurement error.

Another functional approach is studied by [6], who obtains a different unbiased score equation. [14] derives an expression for the asymptotic bias of the naive partial likelihood estimate, which can also be used for bias reduction.

5. Simulation study

By simulation we examined the behaviour of the basic estimates in the situation of one normally distributed covariate under normally distributed, homoscedastic measurement error. We compared the naive estimate, the first and second order Nakamura estimates and the two elementary structural methods, which coincide here. We varied the sample size, the distribution of the true covariate, the amount of censorship, the measurement error variance Σ_U and the underlying distribution of the survival time. Here is a brief summary of the results:

We generated Weibull distributed survival times. Varying the shape parameter had some effect on the magnitude of the bias, but did not change the phenomenological picture depicted below. As may be broadly expected, for very small measurement error the naive estimate still is superior, while for large measurement error the bias is intolerable: for $\Sigma_U = \Sigma_X$ the observed attenuation is about one half.

5.1 Structural correction

In the designs we studied, the amount of censoring did not play a very important role: this suggests that the rare disease assumption (4) underlying (7) may be interpreted liberally to some degree (c.f., however, [13], and Section 4.1.) The effect of the distribution of the true covariate X_i was surprisingly small. Misspecification (by a symmetric mixture of normals and a uniform distribution) did not substantially worsen the behaviour of the structural methods. This insensitivity, on the other hand, is also responsible for the fact that the structural methods were not able to beat Nakumara's first and second order estimates even in situations where all the assumptions on which the structural methods are based were fairly met.

5.2 The Nakamura estimates

The estimates obtained from Nakamura's correction method showed some quite remarkable features: the often observed excellent behaviour was contrasted by sometimes completely wrong results and many numerical difficulties. In a, by far not negligible, number of situations we were confronted with the problems of non-convergence or of wrong convergence, which also had been reported by some other authors. As described in [15], the estimate may not always exist because the derivative of the corrected score functions is not always negative in the neighborhood of the true β. This effect regularly happens when the measurement error gets large. Even when the measurement error variance was half of the covariates variance, Nakamura's estimates often failed to converge. We additionally want to stress that a lot of care is needed with respect to the numerical calculation of the root of the corrected estimating equations. Experimenting with different root finders we got quite often completely different estimates. This is in particular urgent for the second order estimate and if the slope of the hazard rate differs considerably from zero. The need to use numerically expensive procedures makes calculation of the estimate rather slow.

When we restricted ourselves to those situations where the solutions were apparently reliable, the bias was very small indeed. Then Nakamura's methodology proved to be very powerful for correcting the measurement error. After having removed the 'outlying values', the functional methods performed almost always better than the structural methods. As already said above, this was even true in that situation which was used to derive und justify the structural approaches.

Comparing both Nakamura estimates with each other, the second order estimate was in most situations slightly superior to the first order estimate, as long as no numerical problems appeared. Whether this gain in bias reduction is large enough to compensate the additional numerical difficulties and the higher danger to produce artefacts, has to be decided on a case-by-case basis.

6. Concluding Remarks

In this paper we reviewed methods to estimate regression parameters in the Cox model under homoscedastic measurement error. In some more detail, we discussed Nakamura's method as well as the application of the two main structural methods, which were additionally shown to be equivalent in a special case. Then the basic estimates were compared in a simulation study. The overall conclusion resulting from it is somewhat ambiguous, because the Nakamura estimates showed very extreme behaviour. On the one hand, they can lead to a lot of numerical difficulties and may produce artificial results, on the other hand, in those constellations where they behave not irregularly, they are very powerful.

We did not discuss the estimation of the baseline hazard rate under measurement error. Results on this issue can be found in [21, 16, 15, 12, 3]. A comparison of the different methods is still lacking.

Another topic of further research, also quite important for practical application, is the extension of the methods to heteroscedastic measurement error. For instance, this is of particular interest in nutritional studies, where subject matter considerations suggest comparatively high heteroscedastic measurement errors (see, e.g., [22, page 33-48].)

In the last years parametric survival models have attracted much attention, but up to now not much is known how to correct for measurement error in this context. Some results are directly available from [17] and are extended by [10]. A structural approach to measurement error correction in parametric survival models is proposed in [1].

Notes

1. We allow for the border case that (some of) the rows of Σ_U are zero, so that we do not need 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 $U_i[j] \equiv 0$.

2. Cf., however, [19, p. 1169], who characterize a family of distributions where the adapted partial likelihood can be dealt with in a closed form.

Acknowledgments

This work was partially supported by the Deutsche Forschungsgemeinschaft (DFG) within the frame of the Sonderforschungsbereich SFB386. We thank Angela Döring, Helmut Küchenhoff and Hans Schneeweiß for valuable comments and discussions.

References

[1] T. Augustin. Correcting for measurement error in parametric duration models by quasi likelihood. *Biometrical Journal,* accepted for publication, 2001.

[2] T. Augustin. Some basic results on the extension of quasi-likelihood based measurement error correction to multivariate and flexible structural models. In W. Gaul and G. Ritter, editors, *Classification, Automation, and New Media.* Springer, 2001. To appear.

[3] T. Augustin. *An exact corrected log-likelihood for Cox's proportional hazards model*, 2001. Manuscript, University of Munich.

[4] N.E. Breslow. Contribution to Cox (1972). *Journal of the Royal Statistical Society, Series B*, 34:216–217, 1972.

[5] N.E. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30:89–99, 1974.

[6] J. Buzas. Unbiased scores in proportional hazards regression with covariate measurement error. *Journal of Statistical Planning and Inference*, 67:247–257, 1998.

[7] R.J. Caroll, D. Ruppert, and L.A. Stefanski. *Measurement Error in Nonlinear Models.* Chapman and Hall, 1995.

[8] D.G. Clayton. Models for the analysis of cohort and case-control studies with inaccurately measured exposures. In J.H. Dwyer, M. Feinleib, and Lipsert et al., P., editors, *Statistical Models for Longitudinal Studies of Health*, pages 301–331. Oxford University Press, 1991.

[9] D.R. Cox. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34:187–220, 1972.

[10] P. Gimenez, H. Bolfarine, and E.A. Colosimo. Estimation in Weibull regression model with measurement error. *Communications Statistics. Theory and Methods*, 28:495–510, 1999.

[11] P. Hu, A. Tsiatis, and M. Davidian. Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics*, 54:1407–1419, 1998.

[12] Y. Huang and C.Y. Wang. Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. *Journal of the American Statistical Association*, 45:1209–1219, 2000.

[13] M.D. Hughes. Regression dilution in the proportional hazards model. *Biometrics*, 49:1056–1066, 1993.

[14] F.H. Kong. Adjusting regression attenuation in the Cox proportional hazards model. *Journal of Statistical Planning and Inference*, 79:31–44, 1999.

[15] F.H. Kong and M. Gu. Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statistica Sinica*, 9:953–969, 1999.

[16] F.H. Kong, W. Huang, and X. Li. Estimating survival curves under proportional hazards model with covariate measurement errors. *Scandinavian Journal of Statistics*, 25:573–587, 1998.

[17] T. Nakamura. Corrected score functions for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77:127–137, 1990.

[18] T. Nakamura. Proportional hazards model with covariates subject to measurement error. *Biometrics*, 48:829–838, 1992.

[19] M.S. Pepe, S.G. Self, and R. L. Prentice. Further results in covariate measurement errors in cohort studies with time to response data. *Statistics in Medicine*, 8:1167–1178, 1989.

[20] R.L. Prentice. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69:331–342, 1982.

[21] C.Y. Wang, L. Hsu, Z.D. Feng, and R.L. Prentice. Regression calibration in failure time regression. *Biometrics*, 53:131–145, 1997.

[22] W. Willett. *Nutritional Epidemiology*. Oxford Univ. Press, 1998^2.

[23] H. Zhou and C.Y. Wang. Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society, Series B*, 62:657–665, 2000.