Bachelor Thesis

Corrected Score Functions under Additive Berkson Error

Daniel Schlichting

Supervisor: Prof. Dr. Thomas Augustin

Department of Statistics, LMU Munich

7.6.2022

Abstract

Measurement error affects statistical inference. On the other hand, there are methods to correct for measurement error. One of these methods are corrected score functions, that were developed to correct for classical additive measurement error, that itself is just one of many statistical error types. The aim of this thesis is to try to apply the concept of corrected score functions to additive Berkson error in the linear, Poisson and Cox's proportional hazards model with a response and an erroneous covariate variable.

It is possible to derive theoretical corrected score functions. To evaluate the corrected estimators, the newly derived functions are implemented in `R` and a simulation study is performed. For different combinations of sample size and error variance, 500 data sets each are simulated and the ideal maximum-likelihood, naive and corrected estimators are estimated. These are then compared in boxplots to evaluate their bias and variance.

It turns out, that the corrected estimator tends towards a steady state near the true parameter value with increasing sample size. In the same turn, its variance decreases. Increasing error variance made for more variance and in some scenarios with very high error variance the corrected estimator gets biased. Compared to the naive estimator, the corrected estimator is most of the times more accurate at cost of a higher variance.

The simulation study is though very limited, since the parameter values are fixed in advance and only one covariate variable is included. Furthermore, a sensitivity analysis reveals that a misspecification of the error variance leads to remarkable changes regarding variance and bias of the corrected estimator.

All in all, corrected score functions are applicable to additive Berkson error and are worth further investigation.

Contents

1 Introduction **3**

2 Methodology of Measurement Error and Corrected Score Functions **4**
2.1 Measurement Error 4
2.1.1 Classical Additive Measurement Error 4
2.1.2 Additive Berkson Measurement Error 5
2.1.3 Comparison Classical and Berkson Error 5
2.2 Corrected Score Functions 5
2.2.1 Maximum-Likelihood- and M-Estimation 5
2.2.2 Idea of Corrected Score Functions 6
2.2.3 Example: Corrected Score Function Under Classical Additive Measurement Error . 7

3 Derivation of Corrected Score Functions under Berkson Error **8**
3.1 General Procedure to Construct Corrected Score Functions 8
3.2 Assumptions and Prerequisites 8
3.2.1 Stochastic Assumptions 9
3.2.2 Consequences and Important Prerequisites 9
3.3 Linear Model 10
3.4 Poisson Model 13
3.5 Cox's Proportional Hazards Model 14
3.6 Remark on the Derived Corrected Score Functions 17

4 Simulation Study **18**
4.1 ADEMP 18
4.1.1 The ADEMP Scheme 18
4.1.2 Application of ADEMP 18
4.2 Technical Details 19
4.2.1 Generation of Random Numbers 19
4.2.2 Determining the Corrected Estimator 19
4.2.3 Technical Difficulties 19

5 Results of the Simulation Study **21**
5.1 Linear Model 21
5.2 Poisson Model 23
5.3 Cox's Proportional Hazards Model 26
5.4 Poisson Model under False Assumptions 29

6 Concluding Remarks **31**

A Further Graphics **32**

List of Figures **40**

Bibliography **41**

Notation

$Y = (Y_1, Y_2, \dots, Y_n)^T$	vector of response variable, $Y_i \in \mathbb{R}, Y \in \mathbb{R}^n$
$X = (X_1, X_2, \dots, X_n)^T$	vector of true unobserved covariate, $X_i \in \mathbb{R}^p, X \in \mathbb{R}^{n\times p}$
$U = (U_1, U_2, \dots, U_n)^T$	vector of additive measurement error, $U_i \in \mathbb{R}^p, U \in \mathbb{R}^{p\times n}$
$W = (W_1, W_2, \dots, W_n)^T$	vector of erroneous observed covariate, $W_i \in \mathbb{R}^p, W \in \mathbb{R}^{p\times n}$
Δ_i, δ_i	censoring indicator (1 = not censored), δ_i realisation of Δ_i
T_i	event time of observation i
C_i	censoring time of observation i
$\tau_j, j = 1, \dots, k$	observed event times
$\ell_{X_i}(\beta, Y_i, X_i), \ell_X(\beta, Y, X)$	true log-likelihood without erroneous variables, $\ell_{X_i}(\beta, Y_i, X_i) = \log(f_i(Y_i \mid X_i, \beta)), \ell_X(\beta, Y, X) = \sum_{i=1}^n \ell_{X_i}(\beta, Y_i, X_i)$
$S_{X_i}(\beta, Y_i, X_i), S_X(\beta, Y, X)$	true score function without erroneous variables, $S_{X_i}(\beta, Y_i, X_i) = \frac{\partial}{\partial\beta}\ell_{X_i}(\beta, Y_i, X_i), S_X(\beta, Y, X) = \sum_{i=1}^n S_{X_i}(\beta, Y_i, X_i)$
$\ell_{X_i}(\beta, Y_i, W_i), \ell_X(\beta, Y, W)$	naive log-likelihood, use of W instead of X in the true log-likelihood
$S_{X_i}(\beta, Y_i, W_i), S_X(\beta, Y, W)$	naive score function, use of W instead of X in the true score function
$\ell_{W_i}(\beta, Y_i, W_i), \ell_W(\beta, Y, W)$	corrected log-likelihood
$S_{W_i}(\beta, Y_i, W_i), S_W(\beta, Y, W)$	corrected score function
$\lambda(t \mid \cdot), \Lambda(t)$	hazard rate and cumulative hazard rate
$\lambda_0(t), \Lambda_0(t)$	baseline hazard rate and cumulative baseline hazard rate
β_{true}	true parameter, $\beta_{true} \in \mathbb{R}^p$
$\hat{\beta}_{ideal}$	ideal Maximum-Likelihood-estimator, solution to $S_X(\beta, Y, X) = 0$
$\hat{\beta}_{naive}$	naive estimator, solution to $S_X(\beta, Y, W) = 0$
$\hat{\beta}_{cor}$	corrected estimator, solution to $S_W(\beta, Y, W) = 0$
Σ_{U_i}	covariance matrix of $U_i, \Sigma_{U_i} \in \mathbb{R}^{p\times p}$

1 Introduction

Imagine a statistical model, e.g. Poisson regression. If a covariate suffers some sort of measurement error the inference may be affected. *Measurement error in clinical research, yes it matters* is the title of a recent publication by Groenwold and Dekkers [2020] and it is only one of many scientific articles that deals with measurement error. However, measurement error is no the end of the world. If the properties of the measurement error are known, one can correct it and (hopefully) improve the results. Corrected score functions are such a measurement error correction method developed by Nakamura [1990] for the classical additive error that itself is just one of a few different error types. This raises the question if corrected score functions can also be applied to other sorts of measurement error.
This thesis focuses on the question if corrected score functions under additive Berkson error for three selected models, linear and Poisson regression and Cox's proportional hazards model (hereafter Cox's PH model), can be derived, and if so, how the corrected estimators behave in different scenarios.
Berkson and classical error are recapitulated in chapter 2.1 together with a general repetition of measurement error. After that, the eponymous corrected score functions are presented. The whole chapter 2 benefitted from Carroll et al. [2006], a book that is worth reading. With this theoretical basis section 3 guides through the derivation process for the corrected score functions, first in general and then for all three models. The behavior of the corrected estimators is evaluated in a simulation study following the ADEMP scheme by Morris et al. [2019], that is shortly recalled in section 4.1. The remaining part of chapter 4 covers the application of ADEMP to this simulation study and technical details in section 4.2. The open-source software `R` [R Core Team, 2020] is used to implement the corrected score functions and perform the simulation study. Chapter 5 presents the results of the simulation study for the three models and a sensitivity analysis what happens if the corrected score algorithm receives false assumptions in section 5.4. This work concludes with some remarks and limitations.

2 Methodology of Measurement Error and Corrected Score Functions

This chapter focuses on the methodological foundations of measurement errors in statistics and the idea and properties of corrected score functions. The first subsection 2.1 gives a brief overview over measurement error and shows differences between the classical and Berkson error. Subsection 2.2 recalls the idea and theory of corrected score functions, a measurement error correction for the classical additive error developed by Nakamura [1990].

2.1 Measurement Error

Statistical models are based on measured data. The higher the data quality the more likely are more accurate estimation results. However, not all variables can be measured directly and instead only an erroneous substitute is measurable. In this case the variable is subject to measurement error. In general a measurement error is defined as the difference of the true value X and the available value W. Independent of the cause or origin of the error, in all cases the inference is affected. The coefficients may for example be biased or the variance may be underestimated. [Carroll et al., 2006, page 2] shows an example for a measurement error that completely masks the true underlying data structure.
The simple expression *measurement error* does not even describe the complexity of the topic. First of all measurment errors exist for all types of attributes, whether discrete or continuous. In the first case one also speaks of misclassification. An example would be a diagnosis as healthy though the person is in fact ill, and for the continuous case the measurement of protein intake in a nutrition study. The feature of interest may be subject to additive $(X + U)$ or multiplicative $(X * U)$ error with U a random error. In multiplicative error case higher absolute values of the covariate lead to higher or lower deviation of the true values. In some settings the error may even contain information of another or the response variable, which is called differential error. There are lots more error types (and corresponding correction methods) that are all not relevant for this study, for more information see Carroll et al. [2006]. The differences between Berkson and classical measurement error will be explained in detail in the subsequent sections.
If the properties of the error are known, one can do some kind of reverse calculations and, with some uncertainty, try to deduce the true from the false values. The crux is the knowledge about the distribution and properties of the measurement error. In practice this knowledge is most often not available or even worse one does not know if the variables are subject to measurement error. Doing replication studies or other methods for figuring out the type and properties of the underlying measurement error are very expensive and time consuming. Because most correction method are based on some assumptions about the error, it is hard to realize measurement error correction in real world applications.
This work has a focus on theory that will afterwards be evaluated in a simulation study. In consequence all parameters and coefficients are always known, in contrast to real world applications. The expression *measurement error* refers to additive non-differential error for continuous covariates in regression models, if not labelled separately. The true target values are always available. The two subsequent subsections cover the classical and the Berkson error, two types of measurement error. More detailed information and mathematical properties can be found in section 3.

2.1.1 Classical Additive Measurement Error

The classical measurement error is given if the measurement W is the sum of the true value X and the error U. X and U are assumed stochastically independent (see [Carroll et al., 2006, chapter 2]).

$$W = X + U, \qquad X \perp U \tag{1}$$

If one assumes the errors U to be normally distributed with mean 0 and some variance σ_u^2, the expectation of the measured values given the true values is

$$\mathbb{E}(W|X) = X \tag{2}$$

So if one knows the true variable the measurement does not contain any additional information and this also applies to the error.

Classical	Berkson Error
$W = X + U$	$X = W + U$
$W \perp U$	$X \perp U$
$\mathbb{E}(W\|X) = X$	$\mathbb{E}(X\|W) = W$
$Var(W) = Var(X) + Var(U)$	$Var(X) = Var(W) + Var(U)$
$Var(W) > Var(X)$	$Var(X) > Var(W)$

Table 1: Comparison of Classical and Berkson measurement error; on top stochastic assumptions and below implications if U and X or W respectively are normally distributed

An example are questionaires to personal nutrition. The answers from the participants in the questionaire (W) are related to their true nutrition (X) and their subjective perception (U).

2.1.2 Additive Berkson Measurement Error

On the other side, the additive Berkson error assumes stochastic independence of U and W instead of X. Thus X and U are now stochastic dependent which leads to different conditional distributions and expectations (see [Carroll et al., 2006, chapter 2]).

$$X = W + U, \qquad W \perp U \tag{3}$$

In the same scenario as above with normally distributed error U the conditional expectation of the measurement given the true values is in general not X (for details see section 3):

$$\mathbb{E}(W|X) \neq X \tag{4}$$

The dust exposure of a group of miners is an example of Berkson error. The true individual dust pollution (X) is given as the dust pollution of the whole group (W) and some individual component (U) such as smoking behaviour etc. As in this example the Berkson error can be seen more as some sort of deviation than error.

2.1.3 Comparison Classical and Berkson Error

Both types of error affect inference in a negative way. This can be e.g. coefficient bias in regression models, loss of power or underestimation of variance. Though of course there are differences. In the linear model the classical error introduces coefficient bias while the Berkson error does not bias the coefficients (see [Carroll et al., 2006, chapter 3]). Table 1 compares the classical and Berkson measurement error. The upper part shows the general definition and the lower part some more properties if the errors and true values or measurements respectively are normally distributed.

2.2 Corrected Score Functions

This section first gives a very brief repetition of M- and ML-estimation and second describes the idea of corrected score functions. An example of a corrected score function in the linear model closes this section.

2.2.1 Maximum-Likelihood- and M-Estimation

M-estimation is a general approach for parameter estimation. The M-estimator $\hat{\theta}$ for the true parameter θ is defined as the solution of an estimating equation of the form

$$\frac{1}{n}\sum_{i=1}^{n} \psi_i(Y_i, \hat{\theta}) \overset{!}{=} 0 \tag{5}$$

and ψ_i an estimating function that maps to $\mathbb{R}^p, p := dim(\theta)$ (see for example [Carroll et al., 2006, appendix A.6]).
If the estimating function is unbiased, that means $\mathbb{E}(\psi_i(Y_i, \theta)) = 0$ holds, and under certain regularity conditions $\hat{\theta}$ is a consistent estimator for θ (see [Carroll et al., 2006, appendix A.6] or Nakamura [1990]). With the law of large numbers the asymptotic distribution of $\hat{\theta}$ can be derived

$$\begin{aligned} \hat{\theta} &\overset{a}{\sim} N(\theta, \frac{1}{n} A_n^{-1}(\theta) B_n(\theta) A_n^{-1}(\theta)^T) \text{ with} \\ A_n(\theta) &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left\{ \frac{\partial}{\partial \theta^T} \psi_i(Y_i, \theta) \right\} \text{ and} \\ B_n(\theta) &= \frac{1}{n} \sum_{i=1}^{n} Cov(\psi_i(Y_i, \theta)). \end{aligned} \tag{6}$$

The variance of the estimator can be estimated either by evaluating $A_n(\hat{\theta})$ and $B_n(\hat{\theta})$ in a model or without distributional assumptions with the sandwich estimator formula.

$$\begin{aligned} \hat{A}_n &= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \hat{\theta}^T} \psi_i(Y_i, \hat{\theta}) \text{ and} \\ \hat{B}_n &= \frac{1}{n} \sum_{i=1}^{n} \psi_i(Y_i, \hat{\theta}) \psi_i(Y_i, \hat{\theta})^T \end{aligned} \tag{7}$$

2.2.2 Idea of Corrected Score Functions

The idea of corrected score functions and of Nakamura [1990] is, not to correct the estimator but the estimating function, namely the score function, whose root is the estimator. You construct a new estimating function that is centered around the ideal ML-estimator $\hat{\beta}_{ideal}$ and therefore around the true parameter β_{true}.
The basic and ideal scenario is when X is given. Then the true score function $S_{X_i}(\beta, Y_i, X_i)$ yields the *ideal* estimator $\hat{\beta}_{ideal}$. Now if the true values are unknown, one may simply replace X with the erroneous surrogates W which leads to the *naive* score function $S_{X_i}(\beta, Y_i, X_i)$. The root of $\sum_{i=1}^{n} S_{X_i}(\beta, Y_i, W_i)$ is the *naive* estimator $\hat{\beta}_{naive}$. It is not necessarily consistent since $\mathbb{E}(S_{X_i}(\beta_{true}, Y_i, W_i)) = 0$ does not always hold, to say the expectation of the naive score function at the true parameter is not always equal to zero.
Let F be an open convex subset of a parameter space that includes β_{true}. A *corrected* log-likelihood $\ell_{W_i}(\beta, Y_i, W_i)$ is a function for that for all $\beta \in F$ holds that the expectation of $\ell_{W_i}(\beta, Y_i, W_i)$ given the true response and covariate is the true log-likelihood.

$$\mathbb{E}(\ell_{W_i}(\beta, Y_i, W_i)|Y_i, X_i) = \ell_{X_i}(\beta, Y_i, X_i) \quad \text{for } i = 1, \ldots, n \tag{8}$$

If $\ell_{W_i}(\beta, Y_i, W_i)$ is differentiable in F, then the derivative of a corrected log-likelihood $S_{W_i}(\beta, Y_i, W_i) = \frac{\partial}{\partial \beta} \ell_{W_i}(\beta, Y_i, W_i)$ is a corrected score function. If differential and integral are interchangeable, so also under regularity, the conditional expectation of the corrected score function given Y_i and X_i is the true score function.

$$\begin{aligned} \mathbb{E}(S_{W_i}(\beta, Y_i, W_i)|Y_i, X_i) &= \mathbb{E}(\frac{\partial}{\partial \beta} \ell_{W_i}(\beta, Y_i, W_i)|Y_i, X_i) \\ &= \frac{\partial}{\partial \beta} \mathbb{E}(\ell_{W_i}(\beta, Y_i, W_i)|Y_i, X_i) \\ &= \frac{\partial}{\partial \beta} \ell_{X_i}(\beta, Y_i, X_i) \\ &= S_{X_i}(\beta, Y_i, X_i) \end{aligned} \tag{9}$$

The corrected estimator $\hat{\beta}_{cor}$ is the root of $\sum_{i=1}^{n} S_{W_i}(\beta, Y_i, W_i)$, if unique. It is consistent under certain regularity conditions and as M-estimator asymptotically normal distributed with mean $\hat{\beta}_{ideal}$ ([Nakamura, 1990]).

Variance estimation can be done using the normality of the estimator, if the conditions are fulfilled, or by the methods of M-estimation described in the end of the previous section, for details see 2.2.1 and [Nakamura, 1990, proposition 2].
Independent of this idea Stefanski [1989] also developed unbiased score functions. Note that every corrected score function is also an unbiased one but the inversion is not always true.

2.2.3 Example: Corrected Score Function Under Classical Additive Measurement Error

This example is taken from [Nakamura, 1990, chapter 4.2] and discusses the method of corrected score functions in the linear model under classical additive measurement error ($W_i = X_i + U_i$).
The Y_i are assumed to be independent and identically normal distributed with mean $\beta^T X_i$ and variance σ^2. Then if the true values X_i are given the log-likelihood is

$$\ell_{X_i}(\beta, Y_i, X_i) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y_i - \beta^T X_i)^2 \tag{10}$$

and the score function

$$S_{X_i}(\beta, Y_i, X_i) = \frac{\partial}{\partial\beta}\ell_{X_i}(\beta, Y_i, X_i) = -\frac{1}{2\sigma^2}(-2Y_i X_i + 2X_i X_i^T \beta). \tag{11}$$

Therefore the ideal estimator is the solution to $\sum_{i=1}^n (Y_i X_i - X_i X_i^T \beta) = 0$ and identical with the Ordinary Least-Squares estimator $\hat{\beta}_{ideal} = (X^T X)^{-1} X^T Y$. To reduce the coefficient bias and receive a consistent estimator when only W_i instead of X_i is available, one can use the corrected score function. The procedure to derive corrected score functions is described in 3.1. It is

$$S_{W_i}(\beta, Y_i, W_i) = -\frac{1}{2\sigma^2}(-2Y_i W_i + 2W_i W_i^T \beta - 2\Sigma_{U_i}\beta) \tag{12}$$

with Σ_{U_i} the covariance matrix of U_i's. In this special case, one can derive a concrete closed form of the corrected estimator $\hat{\beta}_{cor}$ (see [Gauß, 2020, chapter 2.2])

$$\hat{\beta}_{cor} = (W^T W - \sum_{i=1}^n \Sigma_{U_i})^{-1} W^T Y. \tag{13}$$

3 Derivation of Corrected Score Functions under Berkson Error

This chapter deals with the derivation of the corrected score functions. At first the procedure to construct the corrected score functions will be explained. Second, the assumptions and basic stochastic concepts are recalled and based on this the corrected score functions of the linear, the Poisson and Cox's PH model with one response and one Berkson-erroneous covariate are derived in sections 3.3, 3.4 and 3.5, respectively.

3.1 General Procedure to Construct Corrected Score Functions

As described in [Carroll et al., 2006, chapter 7.4] one can find corrected score functions by being clever or numerical methods, like Monte Carlo, if the solution is too complex. The derivation in this work follows below scheme: At the beginning lie the stochastic assumptions for the errors U and the response variable given the true covariate vector $(Y|X)$. In this case of additive Berkson error, also a distribution for the errorneous surrogates W is assumed. The conditional distribution for $Y|X$ defines the model, e.g. for a normal distribution one is in the special case of the linear model with the ordinary least squares estimator (if the true covariate is available). Hence follows the true log-likelihood $\ell_X(\beta, Y, X)$. If you calculate the conditional expectation of the naive log-likelihood $\ell_X(\beta, Y, W)$ given Y and X the difference to the true log-likelihood becomes visible. By adding additive and multiplicative components this bias can be compensated. The new corrected term is the corrected log-likelihood $\ell_W(\beta, Y, W)$ and the conditional expectation of this expression is just the true log-likelihood, which can be shown by simply calculating it. As already described in chapter 2.2.2, the corrected score function is the derivative of the corrected log-likelihood with respect to the parameter of interest. Its root, if unique, is the corrected estimator $\hat{\beta}_{cor}$.

3.2 Assumptions and Prerequisites

The following rules will be used for the derivation of the corrected score functions in the next paragraphs. (16) is a conditional version of the variance and f is a fixed and not random function.

$$\mathbb{E}(f(Z)|Z) = f(Z) \tag{14}$$
$$\mathbb{E}(f(Z)Y|Z) = f(Z)\mathbb{E}(Y|Z) \tag{15}$$
$$Var(Y|Z) = \mathbb{E}(Y^2|Z) - (\mathbb{E}(Y|Z))^2 \tag{16}$$

For univariate normally distributed random variables with $a, b \in \mathbb{R}$ holds

$$Z \sim N(\mu_z, \sigma_z^2) \Rightarrow aZ + b \sim N(a\mu_z + b, a^2\sigma_z^2). \tag{17}$$

For the Poisson model the expectation of $\exp(W_i|X_i)$ will also be needed. In general it applies that if Z is normally distributed with mean $\mu_z \in \mathbb{R}$ and variance $\sigma_z^2 > 0$, then $\exp(Z)$ is log-normally dsitributed with parameters μ_w and σ_z^2 (Carroll et al. [2006], Appendix A.2).

$$Z \sim N(\mu_z, \sigma_z^2) \Rightarrow \exp(Z) \sim LN(\mu_z, \sigma_z^2) \tag{18}$$

Mean and variance of $\exp(Z)$ are given by

$$\begin{aligned} \mathbb{E}(\exp(Z)) &= \exp\left(\mu_z + \frac{1}{2}\sigma_z^2\right) \\ Var(\exp(Z)) &= \exp(2\mu_z)\left(\exp(2\sigma_z^2) - \exp(\sigma_z^2)\right). \end{aligned} \tag{19}$$

Especially, this means that the expectation of $\exp(Z)$ depends on the mean and variance of the normally distributed Z.

3.2.1 Stochastic Assumptions

For U and W below assumptions are made for all three models.

$$
\begin{aligned}
U_i &\sim N(0, \sigma_u^2), && (20)\\
W_i &\sim N(\mu_w, \sigma_w^2) && (21)\\
W_i &\perp U_i, && (22)\\
i &= 1, \dots, n && (23)
\end{aligned}
$$

By definition of the Berkson error it follows that X is also normally distributed.

$$X_i := W_i + U_i \Rightarrow X_i \sim N(\mu_w, \sigma_w^2 + \sigma_u^2) \tag{24}$$

The two-dimensional random variable $\binom{W_i}{X_i}$ follows a two-dimensional normal distribution.

$$\begin{pmatrix} W_i \\ X_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_w \\ \mu_w \end{pmatrix}, \begin{pmatrix} \sigma_w^2 & \sigma_w^2 \\ \sigma_w^2 & \sigma_w^2 + \sigma_u^2 \end{pmatrix}\right) \tag{25}$$

Based on above property one can derive the conditional distribution of $W_i|X_i$.

$$
\begin{aligned}
\mathbb{E}(W_i|X_i) &= \mu_w + \frac{\sigma_w^2}{\sigma_w^2 + \sigma_u^2}(X_i - \mu_w) = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_u^2} X_i + \frac{\sigma_u^2}{\sigma_w^2 + \sigma_u^2}\mu_w && (26)\\
Var(W_i|X_i) &= \sigma_w^2 - \frac{\sigma_w^4}{\sigma_w^2 + \sigma_u^2} = \frac{\sigma_w^2(\sigma_w^2 + \sigma_u^2) - \sigma_w^4}{\sigma_w^2 + \sigma_u^2} = \frac{\sigma_w^2 \sigma_u^2}{\sigma_w^2 + \sigma_u^2} && (27)\\
W_i|X_i &\sim N\left(\frac{\sigma_w^2}{\sigma_w^2 + \sigma_u^2} X_i + \frac{\sigma_u^2}{\sigma_w^2 + \sigma_u^2}\mu_w, \frac{\sigma_w^2 \sigma_u^2}{\sigma_w^2 + \sigma_u^2}\right) && (28)
\end{aligned}
$$

Both (25) and (28) are analogous to the scenario with classical additive measurement error and normal distribution, compare [Kauermann et al., 2021, chapter 11.4].

3.2.2 Consequences and Important Prerequisites

Now that we know the conditional expectation of $W_i|X_i$, we can construct an expression of the form $aW_i + b$ with $a, b \in \mathbb{R}$ that has the conditional expectation $\mathbb{E}(aW_i + b|X_i) = X_i$:

$$\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w \tag{29}$$

Comparing (28) and (29) yields that W_i is just multiplied with the inverse of the prefactor of X_i and after multiplying the second term with the same prefactor, it is just subtracted. It is advised to remember (29) since in every linear and also most non-linear parts of the three log-likelihoods containing X_i this formula serves as correction.
Proof:

$$
\begin{aligned}
&\mathbb{E}\left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w \middle| X_i\right)\\
&= \frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2}\mathbb{E}(W_i|X_i) - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\\
&\overset{(28)}{=} \frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2}\left(\frac{\sigma_w^2}{\sigma_w^2 + \sigma_u^2} X_i + \frac{\sigma_u^2}{\sigma_w^2 + \sigma_u^2}\mu_w\right) - \frac{\sigma_u^2}{\sigma_w^2}\mu_w && (30)\\
&= X_i + \frac{\sigma_u^2}{\sigma_w^2}\mu_w - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\\
&= X_i
\end{aligned}
$$

When calculating the conditional expectation of the naive log-likelihood later, we will also condition on Y_i. The response variable is known but does not contain any information about the surrogates W_i. So the expectation of W_i given X_i and Y_i is the same as the expectation of W_i given X_i.

$$\mathbb{E}(W_i|Y_i, X_i) = \mathbb{E}(W_i|X_i) \tag{31}$$

The following two expectations are relevant for the calculations in the next sections. First, by transposing (16) the expectation of $W_i^2|X_i$ can be computed.

$$\begin{aligned}
\mathbb{E}(W_i^2|X_i) &= (\mathbb{E}(W_i|X_i))^2 + Var(W_i|X_i) \\
&= \left(\frac{\sigma_w^2}{\sigma_w^2+\sigma_u^2}X_i + \frac{\sigma_u^2}{\sigma_w^2+\sigma_u^2}\mu_w\right)^2 + \frac{\sigma_w^2\sigma_u^2}{\sigma_w^2+\sigma_u^2} \\
&= \left(\frac{\sigma_w^2}{\sigma_w^2+\sigma_u^2}\right)^2 X_i^2 + 2\frac{\sigma_w^2\sigma_u^2}{(\sigma_w^2+\sigma_u^2)^2}X_i\mu_w + \left(\frac{\sigma_u^2}{\sigma_w^2+\sigma_u^2}\right)^2\mu_w^2 + \frac{\sigma_w^2\sigma_u^2}{\sigma_w^2+\sigma_u^2}
\end{aligned} \tag{32}$$

Second, with (17) and (18), one is able to derive the expected value of a log-normally distributed random variable. $\mu_{w|x}$ and $\sigma_{w|x}^2$ are substitutions for the mean and variance of (28). Let

$$a := \beta\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}, \quad \beta \in \mathbb{R}. \tag{33}$$

Then $aW_i|X_i$ and $\exp(aW_i|X_i)$ are normally and log-normally distributed respectively.

$$\begin{aligned}
aW_i|X_i &\sim N(a\mu_{w|x},\, a^2\sigma_{w|x}^2) \\
\exp(aW_i|X_i) &\sim LN(a\mu_{w|x},\, a^2\sigma_{w|x}^2)
\end{aligned} \tag{34}$$

Putting all pieces together the expectation of $\exp(aW_i|X_i)$ is given by

$$\begin{aligned}
\mathbb{E}(aW_i|X_i) &= \beta\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\left(\frac{\sigma_w^2}{\sigma_w^2+\sigma_u^2}X_i + \frac{\sigma_u^2}{\sigma_w^2+\sigma_u^2}\mu_w\right) = \beta\left(X_i + \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right) \\
Var(aW_i|X_i) &= \beta^2\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2\frac{\sigma_w^2\sigma_u^2}{\sigma_w^2+\sigma_u^2} = \beta^2\frac{(\sigma_w^2+\sigma_u^2)\sigma_u^2}{\sigma_w^2} \\
\Rightarrow \mathbb{E}(\exp(aW_i|X_i)) &= \exp\left\{\beta\left(X_i + \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right) + \frac{1}{2}\beta^2\frac{(\sigma_w^2+\sigma_u^2)\sigma_u^2}{\sigma_w^2}\right\}
\end{aligned} \tag{35}$$

Note that this is not the naive conditional expectation since we multiplied W_i with a prefactor a. The naive expectation is derived the same way with $a = 1$.

3.3 Linear Model

The underlying distribution in the linear model is a normal distribution for $Y_i|X_i$ where the mean is a linear combination of X_i and some variance:

$$Y_i|X_i \sim N(\beta_0 + \beta_1 X_i, \sigma_y^2) \tag{36}$$

(37), (39), (40) und (41) show the true density (or likelihood) (Kauermann et al. [2021], chapter 2), log-likelihood and score functions for β_0 and β_1, respectively.

$$f_{Y_i|X_i}(Y_i|X_i,\beta_0,\beta_1) = \frac{1}{\sqrt{2\pi\sigma_y^2}}\exp\left(-\frac{(Y_i-(\beta_0+\beta_1X_i))^2}{2\sigma_y^2}\right) \tag{37}$$

$$\ell_{X_i}(\beta_0,\beta_1,Y_i,X_i) = -\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2}(Y_i^2-2Y_i(\beta_0+\beta_1X_i)+(\beta_0+\beta_1X_i)^2) \tag{38}$$

$$= -\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2}(Y_i^2-2Y_i\beta_0-2Y_i\beta_1X_i+\beta_0^2+2\beta_0\beta_1X_i+\beta_1^2X_i^2) \tag{39}$$

$$S_{X_i}(\beta_0,Y_i,X_i) = -\frac{1}{2\sigma_y^2}(-2Y_i+2\beta_0+2\beta_1X_i) = \frac{1}{\sigma_y^2}(Y_i-\beta_0-\beta_1X_i) \tag{40}$$

$$S_{X_i}(\beta_1,Y_i,X_i) = -\frac{1}{2\sigma_y^2}(-2Y_iX_i+2\beta_0X_i+2\beta_1X_i^2) = \frac{1}{\sigma_y^2}(Y_iX_i-\beta_0X_i-\beta_1X_i^2) \tag{41}$$

A short remark on the linear model under additive Berkson error: If in this special model the covariate is subject to Berkson error the coefficients β_0 and β_1 are still unbiased (see for example Carroll et al. [2006] chapter 3). If no bias is present to be corrected one could argue that measurement error correction is redundant. Nevertheless this model will be used as a benchmark model to see how the corrected estimator behaves and if numerical problems occur.
The conditional expectation of the naive log-likelihood is

$$\begin{aligned}
&\mathbb{E}(\ell_{X_i}(\beta_0,\beta_1,Y_i,W_i)|Y_i,X_i) = \\
&= \mathbb{E}\left(-\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2}(Y_i^2-2Y_i\beta_0-2Y_i\beta_1W_i+\beta_0^2+2\beta_0\beta_1W_i+\beta_1^2W_i^2)\Big|Y_i,X_i\right) \\
&= -\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2}(Y_i^2-2Y_i\beta_0-2Y_i\beta_1\mathbb{E}(W_i|X_i)+\beta_0^2+2\beta_0\beta_1\mathbb{E}(W_i|X_i)+\beta_1^2\mathbb{E}(W_i^2|X_i)) \\
&\overset{(28)\&(32)}{=} -\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2}\left\{Y_i^2-2Y_i\beta_0-2Y_i\beta_1\left(\frac{\sigma_w^2}{\sigma_w^2+\sigma_u^2}X_i+\frac{\sigma_u^2}{\sigma_w^2+\sigma_u^2}\mu_w\right)\right. \\
&\quad +\beta_0^2+2\beta_0\beta_1\left(\frac{\sigma_w^2}{\sigma_w^2+\sigma_u^2}X_i+\frac{\sigma_u^2}{\sigma_w^2+\sigma_u^2}\mu_w\right) \\
&\quad \left.+\beta_1^2\left[\left(\frac{\sigma_w^2}{\sigma_w^2+\sigma_u^2}\right)^2X_i^2+2\frac{\sigma_w^2\sigma_u^2}{(\sigma_w^2+\sigma_u^2)^2}X_i\mu_w+\left(\frac{\sigma_u^2}{\sigma_w^2+\sigma_u^2}\right)^2\mu_w^2+\frac{\sigma_w^2\sigma_u^2}{\sigma_w^2+\sigma_u^2}\right]\right\}
\end{aligned} \tag{42}$$

Remember equation (31) that Y_i does not contain any information about the surrogates. By comparing the true log-likelihood (39) with the above expectation one can spot the differences and construct a corrected log-likelihood:

$$\begin{aligned}
\ell_{W_i}(\beta_0,\beta_1,Y_i,W_i) = &-\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2}\left\{Y_i^2-2Y_i\beta_0-2Y_i\beta_1\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i-\frac{\sigma_u^2}{\sigma_w^2}\mu_w\right)+\beta_0^2\right. \\
&+2\beta_0\beta_1\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i-\frac{\sigma_u^2}{\sigma_w^2}\mu_w\right)+\beta_1^2\left[\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2W_i^2\right. \\
&\left.\left.-2\frac{\sigma_u^2}{\sigma_w^2}\mu_w\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i-\frac{\sigma_u^2}{\sigma_w^2}\mu_w\right)-\left(\frac{\sigma_u^2}{\sigma_w^2}\right)^2\mu_w^2-\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2\frac{\sigma_w^2\sigma_u^2}{\sigma_w^2+\sigma_u^2}\right]\right\}
\end{aligned} \tag{43}$$

Like mentioned above, (29) serves as correction for linear combinations of X_i but furthermore appears in the non-linear correction of W_i^2. As it satisfies (8), (43) is a corrected log-likelihood.

Proof:

$$
\begin{aligned}
&\mathbb{E}(\ell_{W_i}(\beta_0,\beta_1,Y_i,W_i)|Y_i,X_i) = \\
&\quad = \mathbb{E}\left(-\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2}\left\{Y_i^2 - 2Y_i\beta_0 - 2Y_i\beta_1\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right) + \beta_0^2\right.\right.\\
&\quad + 2\beta_0\beta_1\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right) + \beta_1^2\left[\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2 W_i^2\right.\\
&\quad \left.\left.\left. -2\frac{\sigma_u^2}{\sigma_w^2}\mu_w\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right) - \left(\frac{\sigma_u^2}{\sigma_w^2}\right)^2\mu_w^2 - \left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2\frac{\sigma_w^2\sigma_u^2}{\sigma_w^2+\sigma_u^2}\right]\right\}\middle| Y_i, X_i\right)\\
&\quad = -\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2}\left\{Y_i^2 - 2Y_i\beta_0 - 2Y_i\beta_1\mathbb{E}\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\middle| X_i\right) + \beta_0^2\right.\\
&\quad + 2\beta_0\beta_1\mathbb{E}\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\middle| X_i\right) + \beta_1^2\left[\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2\mathbb{E}(W_i^2|X_i)\right.\\
&\quad \left.\left. -2\frac{\sigma_u^2}{\sigma_w^2}\mu_w\mathbb{E}\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\middle| X_i\right) - \left(\frac{\sigma_u^2}{\sigma_w^2}\right)^2\mu_w^2 - \left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2\frac{\sigma_w^2\sigma_u^2}{\sigma_w^2+\sigma_u^2}\right]\beta\right\}\\
&\quad \overset{(29)}{=} -\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2}\left\{Y_i^2 - 2Y_i\beta_0 - 2Y_i\beta_1X_i + \beta_0^2 + 2\beta_0\beta_1X_i\right.\\
&\quad \left. +\beta_1^2\left[\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2\mathbb{E}(W_i^2|X_i) - 2\frac{\sigma_u^2}{\sigma_w^2}\mu_wX_i - \left(\frac{\sigma_u^2}{\sigma_w^2}\right)^2\mu_w^2 - \left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2\frac{\sigma_w^2\sigma_u^2}{\sigma_w^2+\sigma_u^2}\right]\right\}\\
&\quad \overset{(32)}{=} -\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2}\left\{Y_i^2 - 2Y_i\beta_0 - 2Y_i\beta_1X_i + \beta_0^2 + 2\beta_0\beta_1X_i\right.\\
&\quad + \beta_1^2\left[\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2\left(\left(\frac{\sigma_w^2}{\sigma_w^2+\sigma_u^2}\right)^2X_i^2 + 2\frac{\sigma_w^2\sigma_u^2}{(\sigma_w^2+\sigma_u^2)^2}X_i\mu_w + \left(\frac{\sigma_u^2}{\sigma_w^2+\sigma_u^2}\right)^2\mu_w^2 + \frac{\sigma_w^2\sigma_u^2}{\sigma_w^2+\sigma_u^2}\right)\right.\\
&\quad \left.\left. -2\frac{\sigma_u^2}{\sigma_w^2}\mu_wX_i - \left(\frac{\sigma_u^2}{\sigma_w^2}\right)^2\mu_w^2 - \left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2\frac{\sigma_w^2\sigma_u^2}{\sigma_w^2+\sigma_u^2}\right]\right\}\\
&\quad = -\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2}\left\{Y_i^2 - 2Y_i\beta_0 - 2Y_i\beta_1X_i + \beta_0^2 + 2\beta_0\beta_1X_i + \beta_1^2\left[X_i^2 + 2\frac{\sigma_u^2}{\sigma_w^2}\mu_wX_i\right.\right.\\
&\quad \left.\left. + \left(\frac{\sigma_u^2}{\sigma_w^2}\right)^2\mu_w^2 + \left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2\frac{\sigma_w^2\sigma_u^2}{\sigma_w^2+\sigma_u^2} - 2\frac{\sigma_u^2}{\sigma_w^2}\mu_wX_i - \left(\frac{\sigma_u^2}{\sigma_w^2}\right)^2\mu_w^2 - \left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}\right)^2\frac{\sigma_w^2\sigma_u^2}{\sigma_w^2+\sigma_u^2}\right]\right\}\\
&\quad = -\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2}\left\{Y_i^2 - 2Y_i\beta_0 - 2Y_i\beta_1X_i + \beta_0^2 + 2\beta_0\beta_1X_i + \beta_1^2X_i^2\right\}\\
&\quad = \ell_{X_i}(\beta_0,\beta_1,Y_i,X_i)
\end{aligned} \tag{44}
$$

Differentiation of (43) with respect to β_0 and β_1 yields the corrected score functions.

$$
\begin{aligned}
S_{W_i}(\beta_0, Y_i, W_i) &= \frac{\partial}{\partial \beta_0} \ell_{W_i}(\beta_0, \beta_1, Y_i, W_i) \\
&= -\frac{1}{2\sigma_y^2} \left\{ -2Y_i + 2\beta_0 + 2\beta_1 \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2} \mu_w \right) \right\} \\
&= \frac{1}{\sigma_y^2} \left\{ Y_i - \beta_0 - \beta_1 \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2} \mu_w \right) \right\} \\
S_{W_i}(\beta_1, Y_i, W_i) &= \frac{\partial}{\partial \beta_1} \ell_{W_i}(\beta_0, \beta_1, Y_i, W_i) \\
&= -\frac{1}{2\sigma_y^2} \left\{ -2Y_i \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2} \mu_w \right) + 2\beta_0 \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2} \mu_w \right) \right. \\
&\quad + 2\beta_1 \left[\left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} \right)^2 W_i^2 - 2\frac{\sigma_u^2}{\sigma_w^2} \mu_w \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2} \mu_w \right) - \left(\frac{\sigma_u^2}{\sigma_w^2} \right)^2 \mu_w^2 \right. \\
&\quad \left. \left. - \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} \right)^2 \frac{\sigma_w^2 \sigma_u^2}{\sigma_w^2 + \sigma_u^2} \right] \right\} \\
&= \frac{1}{\sigma_y^2} \left\{ (Y_i - \beta_0) \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2} \mu_w \right) - \beta_1 \left[\left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} \right)^2 W_i^2 \right. \right. \\
&\quad \left. \left. -2\frac{\sigma_u^2(\sigma_w^2 + \sigma_u^2)}{\sigma_w^4} \mu_w W_i + \left(\frac{\sigma_u^2}{\sigma_w^2} \right)^2 \mu_w^2 - \frac{(\sigma_w^2 + \sigma_u^2)\sigma_u^2}{\sigma_w^2} \right] \right\}
\end{aligned}
\tag{45}
$$

3.4 Poisson Model

In contrast to the linear model, the Poisson model assumes a Poisson distribution for $Y_i|X_i$ with the only parameter a linear combintion of X_i and (in this work) the exponential function as link function:

$$
Y_i|X_i \sim Po(\exp(\beta_0 + \beta_1 X_i)) \tag{46}
$$

All other assumptions are and will be the same as in the linear and Cox's PH model. The following four equations show the true density (Kauermann et al. [2021], chapter 2), log-likelihood, and score functions, respectively.

$$
\begin{aligned}
f_{Y_i|X_i}(Y_i|X_i, \beta_0, \beta_1) &= \frac{\exp(\beta_0 + \beta_1 X_i)^{Y_i}}{Y_i!} \exp(-\exp(\beta_0 + \beta_1 X_i)) && (47) \\
\ell_{X_i}(\beta_0, \beta_1, Y_i, X_i) &= -\log(Y_i!) + Y_i(\beta_0 + \beta_1 X_i) - \exp(\beta_0 + \beta_1 X_i) && (48) \\
S_{X_i}(\beta_0, Y_i, X_i) &= Y_i - \exp(\beta_0 + \beta_1 X_i) && (49) \\
S_{X_i}(\beta_1, Y_i, X_i) &= Y_i X_i - \exp(\beta_0 + \beta_1 X_i) X_i && (50)
\end{aligned}
$$

After calculating the conditional expectation of the naive log-likelihood one can construct a corrected one.

$$
\begin{aligned}
\mathbb{E}(\ell_{X_i}(\beta_0, \beta_1, Y_i, W_i)|Y_i, X_i) &= \mathbb{E}\left(-\log(Y_i!) + Y_i\beta_0 + Y_i\beta_1 W_i - \exp(\beta_0 + \beta_1 W_i)|Y_i, X_i\right) \\
&= -\log(Y_i!) + Y_i\beta_0 + Y_i\beta_1 E(W_i|X_i) - \exp(\beta_0) E(\exp(\beta_1 W_i)|X_i) \\
&\overset{(28)\&(19)}{=} -\log(Y_i!) + Y_i\beta_0 + Y_i\beta_1 \left(\frac{\sigma_w^2}{\sigma_w^2 + \sigma_u^2} X_i + \frac{\sigma_u^2}{\sigma_w^2 + \sigma_u^2} \mu_w \right) \\
&\quad - \exp(\beta_0) \exp\left\{ \beta_1 \left(\frac{\sigma_w^2}{\sigma_w^2 + \sigma_u^2} X_i + \frac{\sigma_u^2}{\sigma_w^2 + \sigma_u^2} \mu_w \right) + \frac{1}{2}\beta_1^2 \frac{\sigma_w^2 \sigma_u^2}{\sigma_w^2 + \sigma_u^2} \right\}
\end{aligned}
\tag{51}
$$

The corrected log-likelihood is

$$\begin{aligned}\ell_{W_i}(\beta_0, \beta_1, Y_i, W_i) = & -\log(Y_i!) + Y_i\beta_0 + Y_i\beta_1 \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w \right) \\ & - \exp\left\{ \beta_0 + \beta_1 \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w \right) - \frac{1}{2}\beta_1^2 \frac{(\sigma_w^2 + \sigma_u^2)\sigma_u^2}{\sigma_w^2} \right\}.\end{aligned} \tag{52}$$

(29) can again be found in the linear term of X_i and even in the corrected exponential term. The principle is just the same: multiply by the inverse of some prefactor and afterwards subtract "inverse prefactor times the rest".
Proof for (8):

$$\begin{aligned}
&\mathbb{E}(\ell_{W_i}(\beta_0, \beta_1, Y_i, W_i)|Y_i, X_i) = \\
&\quad = \mathbb{E}\Bigg[-\log(Y_i!) + Y_i\beta_0 + Y_i\beta_1 \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w \right) \\
&\quad - \exp\left\{ \beta_0 + \beta_1 \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w \right) - \frac{1}{2}\beta_1^2 \frac{(\sigma_w^2 + \sigma_u^2)\sigma_u^2}{\sigma_w^2} \right\} \Bigg| Y_i, X_i \Bigg] \\
&\quad = -\log(Y_i!) + Y_i\beta_0 + Y_i\beta_1 \mathbb{E}\left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w \Big| X_i \right) \\
&\quad - \mathbb{E}\left(\exp\left\{ \beta_1 \frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i \right\} \Big| X_i \right) \exp\left\{ \beta_0 - \frac{\sigma_u^2}{\sigma_w^2}\beta_1\mu_w - \frac{1}{2}\beta_1^2 \frac{(\sigma_w^2 + \sigma_u^2)\sigma_u^2}{\sigma_w^2} \right\} \\
&\quad \overset{(28)\&(35)}{=} -\log(Y_i!) + Y_i\beta_0 + Y_i\beta_1 X_i - \exp\left\{ \beta_1 \left(X_i + \frac{\sigma_u^2}{\sigma_w^2}\mu_w \right) + \frac{1}{2}\beta_1^2 \frac{(\sigma_w^2 + \sigma_u^2)\sigma_u^2}{\sigma_w^2} \right\} \\
&\quad \exp\left\{ \beta_0 - \frac{\sigma_u^2}{\sigma_w^2}\beta_1\mu_w - \frac{1}{2}\beta_1^2 \frac{(\sigma_w^2 + \sigma_u^2)\sigma_u^2}{\sigma_w^2} \right\} \\
&\quad = -\log(Y_i!) + Y_i\beta_0 + Y_i\beta_1 X_i - \exp(\beta_0 + \beta_1 X_i) \\
&\quad = \ell_{X_i}(\beta_0, \beta_1, Y_i, X_i)
\end{aligned} \tag{53}$$

Differentiation again yields the corrected score functions.

$$\begin{aligned}
S_{W_i}(\beta_0, Y_i, W_i) &= \frac{\partial}{\partial\beta_0}\ell_{W_i}(\beta_0, \beta_1, Y_i, W_i) \\
&= Y_i - \exp\left\{ \beta_0 + \beta_1 \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w \right) - \frac{1}{2}\beta_1^2 \frac{(\sigma_w^2 + \sigma_u^2)\sigma_u^2}{\sigma_w^2} \right\} \\
S_{W_i}(\beta_1, Y_i, W_i) &= \frac{\partial}{\partial\beta_1}\ell_{W_i}(\beta_0, \beta_1, Y_i, W_i) \\
&= Y_i \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w \right) - \exp\left\{ \beta_0 + \beta_1 \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w \right) \right. \\
&\left. - \frac{1}{2}\beta_1^2 \frac{(\sigma_w^2 + \sigma_u^2)\sigma_u^2}{\sigma_w^2} \right\} \left[\left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w \right) - \beta_1 \frac{(\sigma_w^2 + \sigma_u^2)\sigma_u^2}{\sigma_w^2} \right]
\end{aligned} \tag{54}$$

3.5 Cox's Proportional Hazards Model

Cox's proportional hazards Modell [Cox, 1972] is a simple way to model survival data without distributional assumptions. The basic property is the assumption that every observation $i = 1, \ldots, n$ is connected to the hazard rate $\lambda(t|X_i)$ with covariate X_i by

$$\lambda(t|X_i) = \lambda_0(t) \exp(\beta_1 X_i) \tag{55}$$

The covariates are assumed independent of time t and the baseline hazard rate $\lambda_0(t)$ independent of the covariates and completely free in shape. The term proportional hazard refers to the difference between two observations X_1 and X_2, $\exp(\beta_1(X_1 - X_2))$, which is propotional as it does not depend on time. To

avoid identification problems, the model does not include an intercept β_0. Therefore, in the context of this work, the only parameter to be estimated is the coefficient of the covariate β_1.
Since measurement error correction with corrected score functions is impossible with the partial likelihood proposed by Cox (see Stefanski [1989]), instead the so called Breslow likelihood (see for example Augustin [2004]) is used. Let $\tau_j, j = 1, \ldots, k$, be the true observed event times, $D(\tau_j)$ the set of all observations with an event at τ_j and $R(\tau_j)$ the set of all observations just before τ_j that neither have had an event yet nor have been censored before τ_j. Censoring times occuring in the intervall $[\tau_{j-1}, \tau_j)$ are set to τ_{j-1} and $\tau_0 := 0$. The baseline hazard rate is assumed to be piecewise constant ($\lambda_0(t) \equiv \lambda_j > 0, \tau_{j-1} < t \leq \tau_j$). The response variable Y_i is the minimum of the event times T_i and the censoring times C_i and only works together with the censoring indicator δ_i ($\delta_i = 0$ means the observation is censored). In the absence of measurement errors the Breslow likelihood is identical to the partial likelihood and given by

$$L_X(\beta_1, \delta_i, Y_i, X_i) = \prod_{i=1}^{n} \left\{ (\lambda_0(t_i) \exp(\beta_1 X_i))^{\delta_i} \exp\left(-\exp(\beta_1 X_i) \int_0^{t_i} \lambda_0(u) du \right) \right\}. \tag{56}$$

In the below partial log-likelihood d_j is the cardinality of $D(\tau_j)$.

$$\ell_X(\beta_1, \delta_i, Y_i, X_i) = \sum_{j=1}^{k} \left\{ d_j \log \lambda_j + \sum_{i \in D(\tau_j)} \beta_1 X_i - \lambda_j(\tau_j - \tau_{j-1}) \sum_{i \in R(\tau_j)} \exp(\beta_1 X_i) \right\} \tag{57}$$

The partial likelihood estimator $\hat{\beta}_{PL}$ is the root of the score function, if unique.

$$S_X(\beta_1, \delta_i, Y_i, X_i) = \sum_{j=1}^{k} \left\{ \sum_{i \in D(\tau_j)} X_i - d_j \frac{\sum_{i \in R(\tau_j)} X_i \exp(\hat{\beta}_{PL} X_i)}{\sum_{i \in R(\tau_j)} \exp(\hat{\beta}_{PL} X_i)} \right\} \tag{58}$$

(57) also depends on λ_j which itself has to be estimated. (59) shows the Breslow estimator for the cumulative baseline hazard rate $\hat{\Lambda}_{0,X}(t)$

$$\hat{\Lambda}_{0,X}(t) = \sum_{j: \tau_j \leq t} \frac{d_j}{\sum_{i \in \mathbb{R}(\tau_j)} \exp(\beta_1 X_i)} \tag{59}$$

This cumulative estimator can be rebuilt to estimate the constant baseline hazard rate λ_j using X and β_1, given in (60). It is applied in the denominator of (58).

$$\hat{\lambda}_{j,X} = \frac{d_j}{\sum_{i \in \mathbb{R}(\tau_j)} \exp(\beta_1 X_i)} \tag{60}$$

A corrected log-likelihood is based on the same expectations and derivations like in linear and Poisson case and is shown below.

$$\begin{aligned} \ell_W(\beta_1, \delta_i, Y_i, W_i) = \sum_{j=1}^{k} \Bigg\{ & d_j \log \lambda_j + \sum_{i \in D(\tau_j)} \beta_1 \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2} \mu_w \right) - \lambda_j(\tau_j - \tau_{j-1}) \\ & \sum_{i \in R(\tau_j)} \exp\left[\beta_1 \left(\frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2} W_i - \frac{\sigma_u^2}{\sigma_w^2} \mu_w \right) - \frac{1}{2} \beta_1^2 \frac{(\sigma_w^2 + \sigma_u^2)\sigma_u^2}{\sigma_w^2} \right] \Bigg\} \end{aligned} \tag{61}$$

Proof for (8):

$$
\begin{aligned}
&\mathbb{E}(\ell_W(\beta_1,\delta_i, Y_i, W_i)|\delta_i, Y_i, X_i) = \\
&\quad \mathbb{E}\left(\sum_{j=1}^{k}\left\{d_j \log\lambda_j + \sum_{i\in D(\tau_j)}\beta_1\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right) - \lambda_j(\tau_j-\tau_{j-1})\right.\right.\\
&\quad\left.\left.\sum_{i\in R(\tau_j)}\exp\left[\beta_1\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right) - \frac{1}{2}\beta_1^2\frac{(\sigma_w^2+\sigma_u^2)\sigma_u^2}{\sigma_w^2}\right]\right\}\Bigg|\delta_i, Y_i, X_i\right)\\
&\quad = \sum_{j=1}^{k}\left\{d_j\log\lambda_j + \sum_{i\in D(\tau_j)}\beta_1\mathbb{E}\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\Big|X_i\right) - \lambda_j(\tau_j-\tau_{j-1})\right.\\
&\quad\left.\sum_{i\in R(\tau_j)}\mathbb{E}\left(\exp\left(\beta_1\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i\right)\Big|X_i\right)\exp\left[-\beta_1\frac{\sigma_u^2}{\sigma_w^2}\mu_w - \frac{1}{2}\beta_1^2\frac{(\sigma_w^2+\sigma_u^2)\sigma_u^2}{\sigma_w^2}\right]\right\} \qquad (62)\\
&\quad \overset{(30)\&(35)}{=} \sum_{j=1}^{k}\left\{d_j\log\lambda_j + \sum_{i\in D(\tau_j)}\beta_1 X_i - \lambda_j(\tau_j-\tau_{j-1})\right.\\
&\quad\quad \sum_{i\in R(\tau_j)}\exp\left[\beta_1 X_i + \beta_1\frac{\sigma_u^2}{\sigma_w^2}\mu_w + \frac{1}{2}\beta_1^2\frac{(\sigma_w^2+\sigma_u^2)\sigma_u^2}{\sigma_w^2}\right]\\
&\quad\quad\left.\times\exp\left[-\beta_1\frac{\sigma_u^2}{\sigma_w^2}\mu_w - \frac{1}{2}\beta_1^2\frac{(\sigma_w^2+\sigma_u^2)\sigma_u^2}{\sigma_w^2}\right]\right\}\\
&\quad = \sum_{j=1}^{k}\left\{d_j\log\lambda_j + \sum_{i\in D(\tau_j)}\beta_1 X_i - \lambda_j(\tau_j-\tau_{j-1})\sum_{i\in R(\tau_j)}\exp(\beta_1 X_i)\right\}\\
&\quad = \ell_X(\beta_1,\delta_i,Y_i,X_i)
\end{aligned}
$$

Because λ_j and X_i are unknown the baseline hazard rate has to be estimated using W_i. Therefore, the Breslow estimator for the cumulative baseline hazard rate $\hat{\Lambda}_{0,X}(t)$ (60) is corrected:

$$
\hat{\Lambda}_{0,W}(t) = \sum_{j:\tau_j\le t}\frac{d_j}{\sum_{i\in R(\tau_j)}\exp\left[\beta_1\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right) - \frac{1}{2}\beta_1^2\frac{(\sigma_w^2+\sigma_u^2)\sigma_u^2}{\sigma_w^2}\right]} \qquad (63)
$$

Just as before this version is rebuilt to estimate λ_j.

$$
\hat{\lambda}_{j,W} = \frac{d_j}{\sum_{i\in R(\tau_j)}\exp\left[\beta_1\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right) - \frac{1}{2}\beta_1^2\frac{(\sigma_w^2+\sigma_u^2)\sigma_u^2}{\sigma_w^2}\right]} \qquad (64)
$$

The correction of $\exp(\beta_1 W_i)$ is in fact the same as in the Poisson model, compare (52). First, differentiating with respect to (the only parameter of interest) β_1 and second, replacing λ_j with (64) yields the corrected score function.

$$
\begin{aligned}
S_W(\beta_1,\delta_i,Y_i,W_i) &= \frac{\partial}{\partial\beta_1}\ell_W(\beta_1,Y_i,W_i) = \sum_{j=1}^{k}\left\{\sum_{i\in D(\tau_j)}\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right)\right.\\
&- d_j\sum_{i\in R(\tau_j)}\exp\left[\beta_1\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right) - \frac{1}{2}\beta_1^2\frac{(\sigma_w^2+\sigma_u^2)\sigma_u^2}{\sigma_w^2}\right]\\
&\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w - \frac{(\sigma_w^2+\sigma_u^2)\sigma_u^2}{\sigma_w^2}\beta_1\right) \qquad (65)\\
&\left.\left(\sum_{i\in R(\tau_j)}\exp\left[\beta_1\left(\frac{\sigma_w^2+\sigma_u^2}{\sigma_w^2}W_i - \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right) - \frac{1}{2}\beta_1^2\frac{(\sigma_w^2+\sigma_u^2)\sigma_u^2}{\sigma_w^2}\right]\right)^{-1}\right\}
\end{aligned}
$$

3.6 Remark on the Derived Corrected Score Functions

All derived estimation functions are based on the normal distribution assumptions for surrogates W_i and errors U_i. Furthermore the correction requires knowledge of the true underlying variances. For the major part of the simulation study they are always available and accessible. Part 5.4 presents the results of a simulation with incorrect assumptions for σ_u^2, since in reality, like mentioned in chapter 2.1, it is pretty hard and expensive to acquire detailed information about the errors.
Compared to the corrected score functions in case of a classical additive measurement error (see Nakamura [1990], Gauß [2020] or Augustin [2004]), the correction terms seem more complex and of a different manner. While the moment generating function is very helpful for denoting corrected score functions for classical error, the correction of Berkson error involves additional additive and multiplicative elements. In general all derived corrections are of the form $aW_i + b$ with $a, b \in \mathbb{R}$. The simplest case is (29), the linear correction of W_i, with $a = \frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2}$ and $b = -\frac{\sigma_u^2}{\sigma_w^2}\mu_w$. The correction terms for W_i^2 and $\exp(\beta_1 W_i)$ can also be written in this form with non-random prefactor a and non-random additive constants b.

4 Simulation Study

This section covers the structure and technical implementation of the simulation study. At first, the ADEMP scheme will be repeated and the second part deals with technical details, like determining the corrected estimate.

4.1 ADEMP

4.1.1 The ADEMP Scheme

ADEMP is a blueprint for simulation studies proposed by Morris et al. [2019]. They analyzed simulation studies of various papers and showed ways for improvement, whether in the design, implementation or other aspects. This section will only describe the suggested scheme. More information can be found in the original paper [Morris et al., 2019].
The goal of ADEMP is to start a simulation study by thorough planning and design the study as a whole. The letters represent Aims, Data-generating mechanisms, Estimands and other targets, Methods and Performance Measures. The first step is defining the aims. Is the goal of the study to compare different methods in a special scenario or to evaluate one specific method in different scenarios? What properties should be checked, e.g. robustness or bias? Since the advantage of simulated data compared to real data is the knowledge of the true (usually hidden) parameters and properties, the way the artificial data are generated is crucial. The data-generating mechanism can be accessed on many levels, which requires rigorous thinking. For example should a parametric or non-parametric model be used, what parameter values and sample size are chosen? Estimands and other targets refers to the goal of the method of interest, e.g. whether it is for estimation, for prediction, ... Only relevant methods, which in best case are already implemented, should be analyzed. This also means to look for similar work with similar or even the same goal. The performance measures should give an answer to the questions formulated in the previous points. One should also keep in mind that these measures are estimates themselves and thus subject to errors. As always when working on any kind of project a well-thought structure and clear terminology enhance the outcome.

4.1.2 Application of ADEMP

The **Aims** of this thesis are to evaluate the corrected estimators in linear, Poisson and Cox's PH model under additive Berkson error in different settings of sample size n and error variance σ_u^2. Only parametric **Data-generating mechanisms** are used, that means all random values are drawn from parametric models: Errors U and surrogates W from a normal distribution yielding the true covariate X. The response variable Y in linear and Poisson case is drawn from a normal and Poisson distribution respectively, see 3.3 and 3.4. Realistic survival times are generated with the technique described in Bender et al. [2005]. Therefore uniformly distributed values are drawn and transformed to survival times following a Weibull based PH model. The censoring times are drawn from an exponential distribution independently from the event times. Y is constructed as the minimum of event and censoring time together with the censoring indicator δ_i, where $\delta_i = 0$ means censoring. The **Estimands** are the ideal, naive and corrected estimators. In this thesis, the **Method** of interest is the estimation procedure for $\hat{\beta}_{cor}$. Since the measurement error correction has been developed in this work and never been used before, a new implementation was needed and is described below in part 4.2. Bias and variability of the estimators are the **Performance measures** and are displayed in graphical form as boxplots.
So far the ADEMP setting. Sample size n and error variance σ_u^2 are modified while all other parameters are held constant. The number of observations is varied from 20, to 5000 and the variance from 0 to 100 percent of the surrogate variance σ_w^2. For a model, e.g. the linear model, a low and a high sample size are fixed. Then for both sample sizes the error variance is modified and a simulation is run for every parameter combination. Analogously n is varied for two fixed values of σ_u^2. This makes four simulation runs for a model. In each simulation run $N = 500$ datasets of sample size n and with error variance σ_u^2 are simulated and the three estimators are estimated. These are then compared with the true (and in a simulation study known) parameter value β_{true} in a boxplot.
The subsequent section covers the technical details and the specific parameter values are described in section 5 right prior to the results.

4.2 Technical Details

The technical details and implementation of the simulation are subject of this section. This includes the generation of random numbers, the optimization process and technical difficulties, like divergence of the estimation algorithm. The whole simulation was carried out in `R` [R Core Team, 2020] and the code can be found in the corresponding GitLab repository corr_scores_berkson_git.

4.2.1 Generation of Random Numbers

Random numbers were generated with R's built-in library `stats`. In detail, this includes the generation of random numbers from a uniform, Poisson, exponential and normal distribution. As described in the previous paragraph 4.1.2, the uniformly distributed values were transformed to survival times following a cox-weibull model. Reproducibility was assured by using `set.seed()`.

4.2.2 Determining the Corrected Estimator

In this work the corrected estimators for the linear and Poisson model will be derived by maximisation of the corrected log-likelihoods with R's function `optim` using method Nelder-Mead.
This simplex based algorithm developed by Nelder and Mead does not use derivatives but instead only function values. Imagine a one-dimensional function like $f(x) = -x^2$. You start with three points, random or chosen in advance, that are connected to a triangle. For maximisation in every iteration we evaluate another point and compare it with the edges of our triangle. If the function value is higher (in the sense of maximisation better) than some other edge, replace one of the previous edges according to a set of rules. Otherwise shrink the triangle towards the maximal edge. This way the triangle moves towards the maximum until some stopping criterion is reached. Triangle is an example, the algorithm is defined for arbitrarily large optimization problems. For more details see [Nocedal and Wright, 2006, chapter 9]. Although the algorithm is known to get stuck in local optima, it turned out the most robust method in the sense of convergence and numerical problems compared to the quasi-Newton method *BFGS* and conjugate gradients method *CG* of `optim`.
For the estimation in Cox's PH model a hand-written Newton-Raphson algorithm from the internet ([Swaminathan, 2021]) has been adjusted. Newton-Raphson is closely related to Fisher-Scoring and in this case the derivative of (65) with respect to β_1 was used as learning rate.
Both methods require starting points. As suggested by [Nakamura, 1990, chapter 6] the naive estimates for β_0 and β_1 serve as initial values. If the naive estimation terminated without yielding a valid estimate, the initial values were set to 0.
All three methods for calculating corrected estimators were verified the following way: First, the true log-likelihood was implemented and the ideal ML-estimators were derived. In the second step, these were compared to the results from R's functions `lm` and `glm` from package `stats`. For Cox's PH model the package `survival` with its function `coxph` was used as comparison. In all cases the difference between already implemented and the above described procedures was less than 10^{-3}.

4.2.3 Technical Difficulties

It is known that corrected score functions suffer from numerical problems meaning invalid estimators or divergence of optimization procedures (see for example [Kong and Gu, 1999] or the simulation study of Gauß [2020]). This was also the case in this work. Some brief explorative simulations indicated that the higher the term $\beta_0 + X_i\beta_1$, that also appears in the exponential function, the more likely divergence or invalid values are. One must say that also with lower values of the above term numerical problems occurred. Since, by technical reasons, optimization for β_0 and β_1 is performed simultaneously, the number of valid estimates for both coefficients is the same. The goal of this work is to evaluate and compare the corrected estimators, and not to analyse the numerical difficulties of corrected score functions. Therefore the parameter values were selected so that also in extreme cases (e.g. high variance and low sample size) at least a small number of estimates is valid.
An estimate is considered *valid* if the algorithm converges and yields a value that is reasonably close to the true parameter value (abolute relative bias less than 10, for relative bias see section 5). All other

estimates, that are either missing, infinite, mathematically undefined or very high absolute values, are classified as invalid and removed from the dataset.

Model	β_0	β_1	μ_w	σ_w^2	σ^2
Linear	-2	0.5	1	5	1
Poisson	-2	0.5	1	5	/
Cox PH	/	0.15	43	6	/

Table 2: True parameter values (except parameters of survival times). "/" indicates that the parameter does not appear in the model.

5 Results of the Simulation Study

This section presents the results of the simulation study. The presentation structure is the same for all three models: First, the true parameter values used for simulating the data are introduced and ongoing the results in graphical form are presented. Only selected results are shown, more graphics can be found in the appendix. The last part of this section is a simulation of the Poisson model under false assumptions for the corrected estimator.
For evaluating the estimators' accuracy and variance all valid estimates are visualized in a boxplot grouped by either sample size or error variance. The number of valid estimates is given above the corresponding boxplot in a label. For example $N = 500$ means that on all 500 simulated data sets the estimation procedure yielded a valid value whereas $N = 482$ means that 18 estimates are invalid for whatever reason. The x axis groups by estimation method, ideal, naive or corrected. The y axis shows the relative bias of an estimator $\hat{\beta}$ defined as $(\hat{\beta} - \beta_{true})/\beta_{true}$. So the relative bias 0 would be optimal as it means that $\hat{\beta}$ is β_{true}, indicated as a black line in the plots. A positive relative bias means overestimation of the parameter, and vice versa. For better visibility, outliers are winsorized. All values greater or less than 1 or -1 are set to 1 or -1 respectively.

5.1 Linear Model

The true parameter values and distributions of the linear model are

$$\begin{aligned}
W_i &\sim N(1, 5) \\
U_i &\sim N(0, \sigma_u^2) \\
X_i &\sim N(1, 5 + \sigma_u^2) \\
Y_i &\sim N(-2 + 0.5X_i, 1) \\
i &= 1, \ldots, n.
\end{aligned}$$

The intercept takes a negative value of -2 and the covariate coefficient the value 0.5. The surrogates W_i are normally distributed around mean 1 with variance 5 and the response variance σ_y^2 is 1. An overview over all true parameters for the three models is given in table 2.
Figure 1 shows estimates for β_0 with different sample sizes for fixed error variance $\sigma_u^2 = 0.7 \times \sigma_w^2 = 3.5$. Remember that there is no bias from additive Berkson error in the linear model and it thus just serves as some sort of benchmark. From top-left to bottom-right the sample size gets larger. One can see, that with higher n the variances of all estimators get smaller though the variance of $\hat{\beta}_{cor}$ is always bigger than that of $\hat{\beta}_{ideal}$. Almost all boxplots are centered around the relative bias zero. Note that with the lowest sample ($n = 20$) size a few scenarios did not yield a valid estimate ($N = 486$).
Sample size fixed at 500 figure 2 shows varying error variance. From top-left to bottom-right σ_u^2 (in the plot denoted as su) increases to 100 percent of σ_w^2. The top-left pane is a control scenario with $\sigma_u^2 = 0$. In this case covariates and surrogates are the same and the three boxplots are identical. The variance of the naive and the corrected estimator increase with higher error variance. $\hat{\beta}_{cor}$ has a higher deviation than $\hat{\beta}_{naive}$.
So far the results are as expected: Lower σ_u^2 and higher n decrease the variance of the naive and the corrected estimator and vice versa. Besides the fact that there is no bias in this model, neither of the estimators introduces a bias.

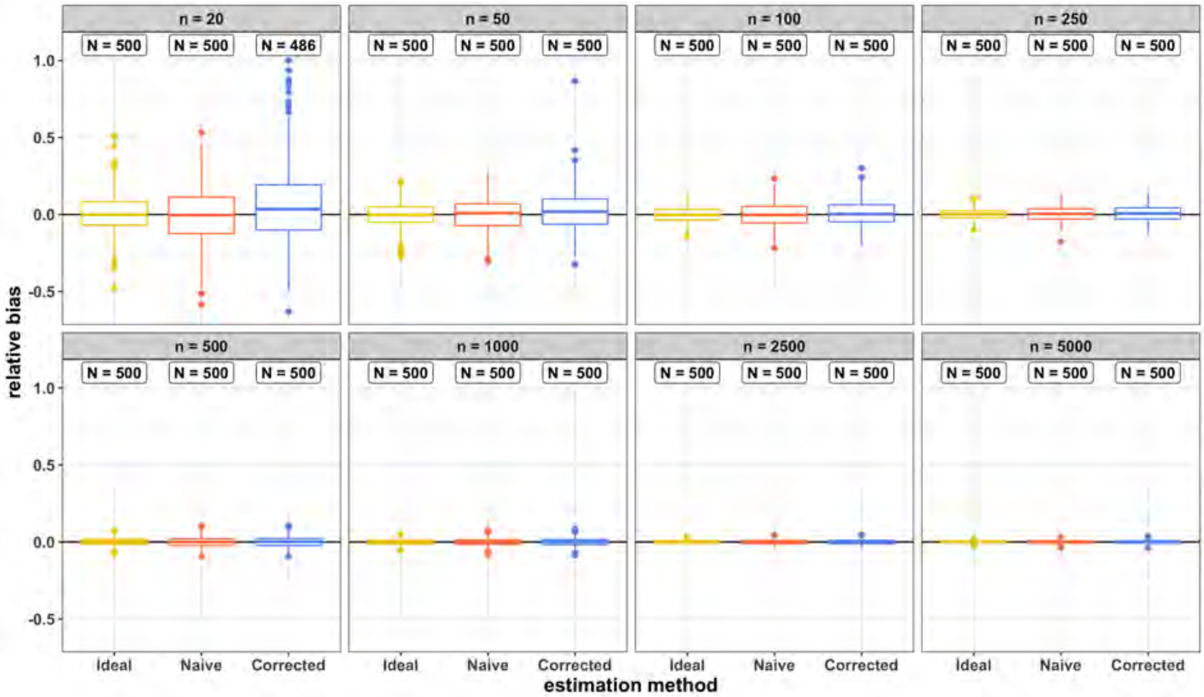

Figure 1: Estimates for β_0 with $\sigma_u^2 = 3.5$ and varying n, linear model

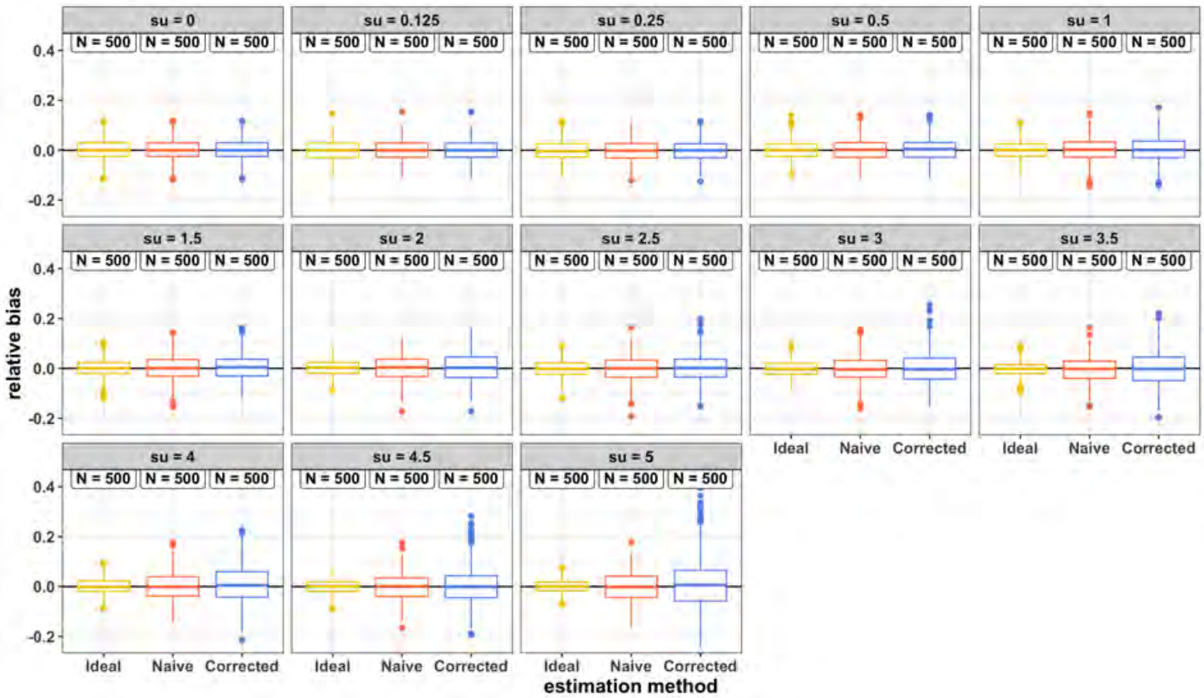

Figure 2: Estimates for β_1 with $n = 500$ and varying σ_u^2, linear model

5.2 Poisson Model

The true parameters in the Poisson model are the same as in the linear model, except that the Poisson distribution does not need a second parameter:

$$
\begin{aligned}
W_i &\sim N(1, 5) \\
U_i &\sim N(0, \sigma_u^2) \\
X_i &\sim N(1, 5 + \sigma_u^2) \\
Y_i &\sim P(-2 + 0.5X_i) \\
i &= 1, \ldots, n.
\end{aligned}
$$

In contrast to the linear model additive Berkson error introduces a coefficient bias in the Poisson model, as can be found in figure 3. The naive estimator for the intercept is clearly biased regardless of the sample sizes. The corrected estimator, on the other hand, tends towards about 0.05 and is thus clearly better. For all estimators, the variance decreases with increasing sample size. In the same manner the number of valid estimates increases with n, though at $n = 20$ less than a fifth of the estimation procedures yielded a valid estimate. One must admit that even the pre-defined and optimised routines from package *stats* did not succeed in all cases.

As seen, the Berkson error affects the intercept estimate. However, the naive estimator for the covariate coefficient is not biased, as figure 4 shows. σ_u^2 is again fixed at 3.5 and n is varied. The naive estimator is pretty accurate while the corrected estimator is biased for lower sample sizes and then tends towards 0.05. Just like the corrected intercept estimator the corrected estimator for β_1 has the highest variance and pretty much behaves the same in terms of the median.

Looking at the effects of higher error variance on the intercept estimator, shown in figure 5, one can see the bias of the naive estimator grow with σ_u^2. For both naive and corrected estimation the variance also increases with σ_u^2, though $\hat{\beta}_{cor}$ is a lot more uncertain. Until σ_u^2 is half as large as σ_w^2, the median tends towards some small positive value again. Interestingly, for σ_u^2 greater than $0.5\sigma_w^2$ it drifts away from zero in a negative direction and gets biased. In the scenarios depicted, $\hat{\beta}_{cor}$ is always better than $\hat{\beta}_{naive}$. This may not hold for more extreme scenarios where the error variance is greater than the surrogate variance. For the highest error variance still about a third or more of the estimates are valid.

Figure 6 shows the same settings for the covariate coefficient estimators. Again the naive estimation should be unbiased. However a small negative drift can be found for $\sigma_u^2 = 3$ upwards. The corrected estimator for β_1 behaves just like the one for β_0. It first tends to a small positive value and then, at the same point as the naive one, switches direction introducing a bias. Also the variance increases with the error variance and is most of the times the highest of the three estimators.

In the Poisson model one may conclude that in scenarios where σ_u^2 is (much) smaller than σ_w^2 and sample size is sufficiently large the corrected estimator is superior to the naive estimator. In rougher settings, where those two conditions are violated, the naive estimation is advised. This might be important in the context of diverging estimation algorithms, where the corrected estimation lacks stability, and especially if the intercept is not of interest or cannot be interpreted.

More simulations with lower sample size and error variance can be found in the appendix. Just as one might expect, a lower sample size introduces more uncertainty and a lower variance more certainty to the estimators.

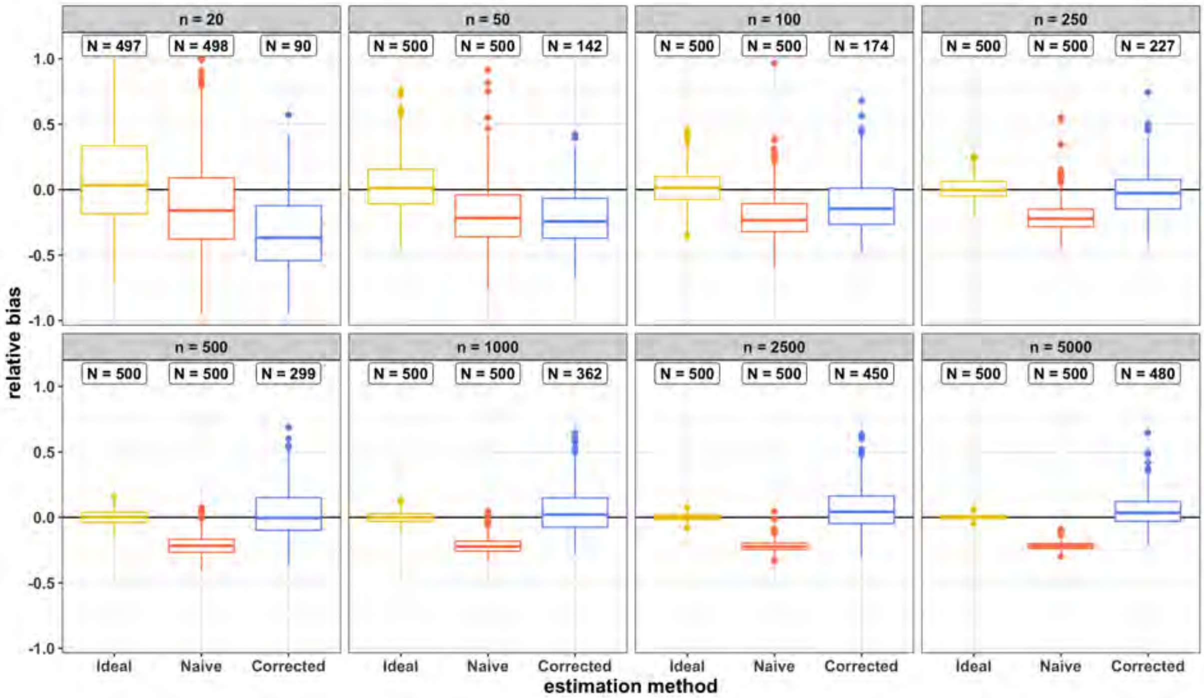

Figure 3: Estimates for β_0 with $\sigma_u^2 = 3.5$ and varying n, Poisson model

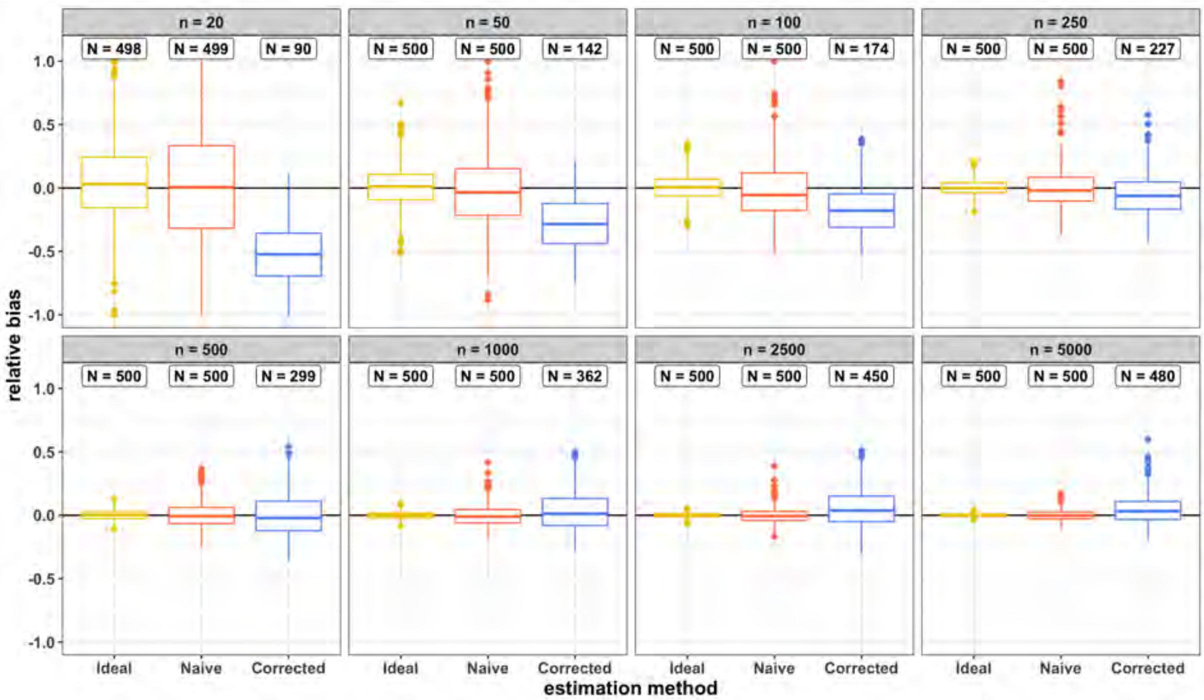

Figure 4: Estimates for β_1 with $\sigma_u^2 = 3.5$ and varying n, Poisson model

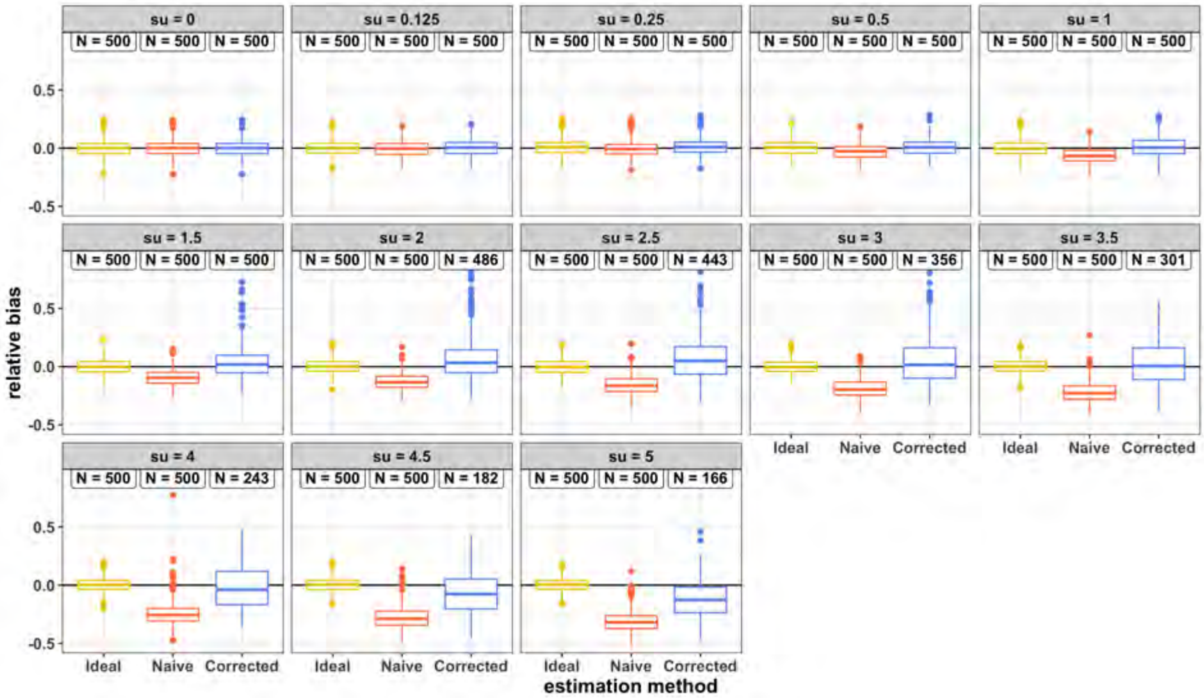

Figure 5: Estimates for β_0 with $n = 500$ and varying σ_u^2, Poisson model

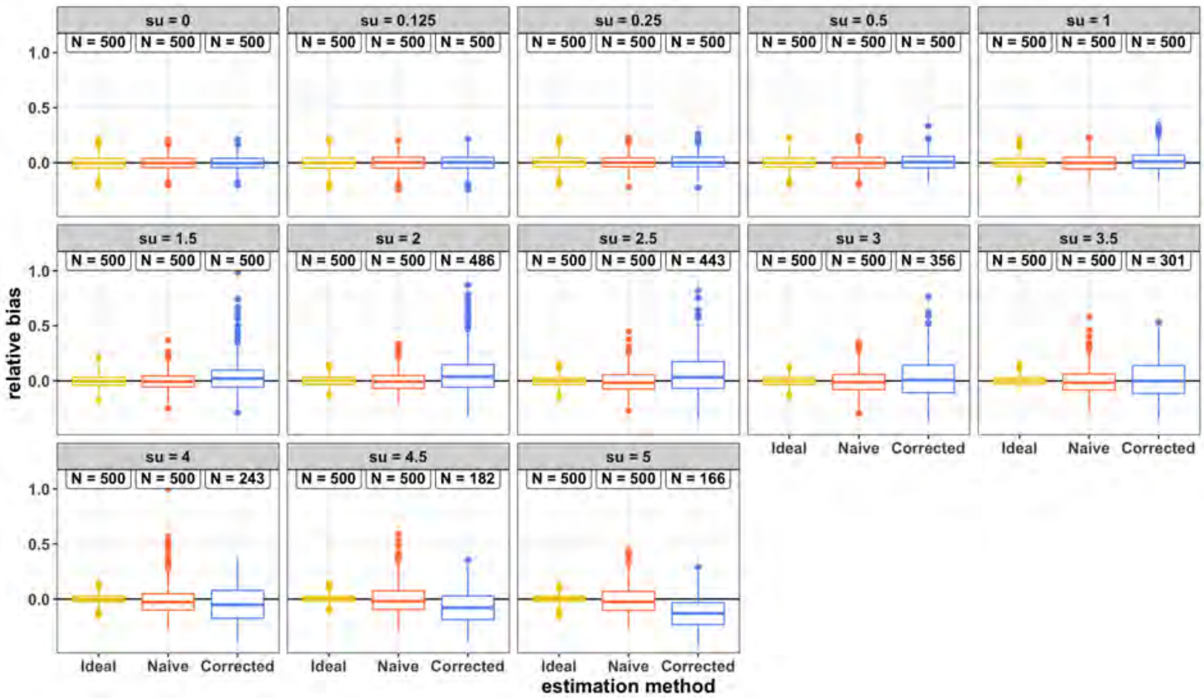

Figure 6: Estimates for β_1 with $n = 500$ and varying σ_u^2, Poisson model

5.3 Cox's Proportional Hazards Model

The survival times generated imitate the uranium miners data ([Kreuzer et al., 1999] or [Bender et al., 2005]). The basis are uniformly distributed random variables and a weibull distribution with shape 13.6141 and scale 69.4574 (in R's parameterization; with parameterization of [Bender et al., 2005] the scale is $69.4574^{-13.6141}$) so that the distribution's mean and variance are close to 66.86 and 6^2 respectively. Like in the original uranium miners study the true coefficient value is 0.15, and mean and variance of surrogates W are 43 and 6, respectively. In this work the surrogate and measurement error are normally distributed, so the uranium data cannot be copied perfectly. Nevertheless the simulated survival times are pretty realistic. As described in chapter 3.5, Cox's PH model has only one coefficient β_1 but also involves censoring. This censoring was simulated independently from an exponential distribution with parameter λ and the four simulation runs were carried out with two different values of λ: $1/66.86$ and $1/30$. First value corresponds to about 16 to 20 percent of censored observations and the latter to about 36 to 40 percent censoring, observed in some simulations.

The first point to mention when looking at figure 7, that shows estimates for β_1 with $\sigma_u^2 = 4.2$ and $\lambda = 1/66.86$ for varying n, is the higher uncertainty of all estimators compared to the other models. Therefore the bias of $\hat{\beta}_{naive}$ only becomes visible at really high sample sizes. $\hat{\beta}_{cor}$ on the other side tends towards zero, but has again the highest variance. If n increases, the variances of the estimators decrease. Regarding divergence of the optimization algorithm, this model is quite robust compared to the Poisson model as a sample size of $n = 50$ yielded slightly more than 70 percent valid estimates. Though it cannot be said if this is due to the different model or different implementation.

The absolute values of the levels of σ_u^2, when n is fixed, differ from those of the previous models due

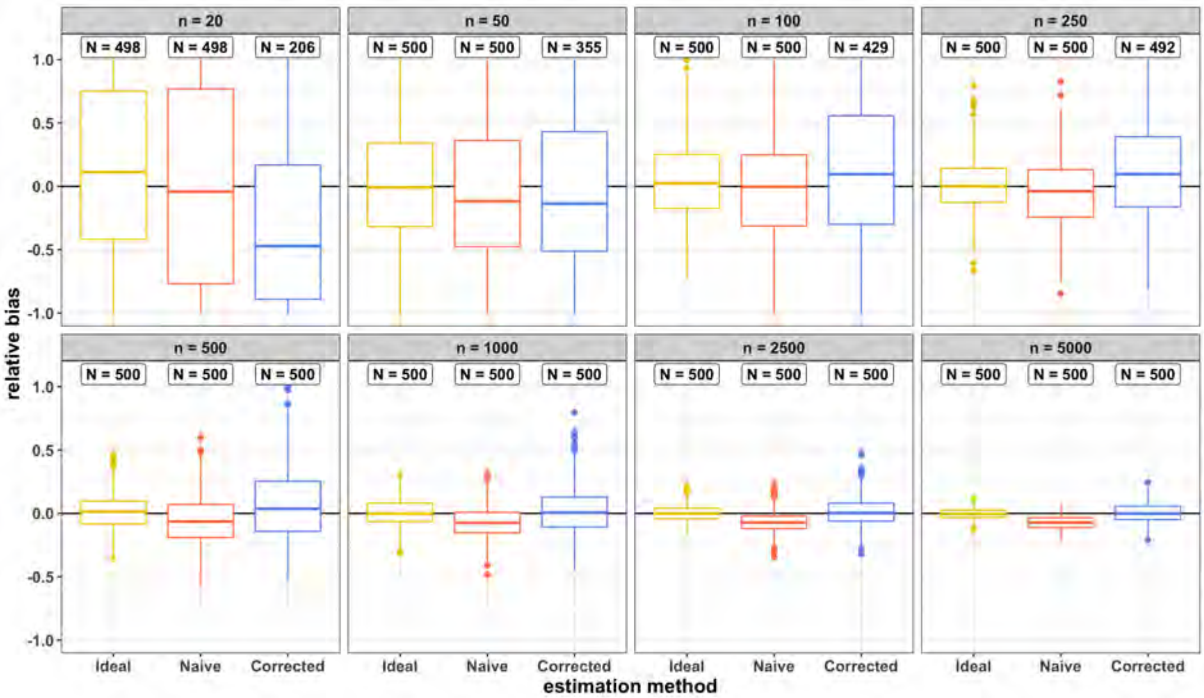

Figure 7: Estimates for β_1 with $\sigma_u^2 = 4.2$, $\lambda = 1/66.86$ and varying n, Cox's PH model

to the fact that σ_w^2 now takes the value 6 instead of 5. Actually, they are calculated in the same way as fractions of the surrogate variance. Figure 8 shows various error variances for $n = 500$ and lower censoring ($\lambda = 1/66.86$). From $\sigma_u^2 = 0.3\sigma_w^2 = 1.8$ upwards the naive estimator gets biased. In contrast to the Poisson model its variance does not seem to get much larger, whereas the uncertainty of the corrected estimator increases strongly. $\hat{\beta}_{cor}$ again tends to a small positive value and seems to have reached a steady state. In terms of numerical problems, the algorithm is robust as almost all estimates are valid in the roughest setting.

The last figure 9 depicts the same setting as before in figure 8 with more censoring ($\lambda = 1/30$). The figures look pretty similar and in fact the signs of the medians are always the same. The only difference is that more censoring makes for more uncertainty and thus the uncertainty of all estimators is larger. This makes sense as more censoring means less information. The number of valid estimates is also lower than in the scenarios with lower censoring.

In summary, it can be said that if there are many data, the corrected estimate does a good job and seems consistent. On the other hand, if the error variance is high compared to the surrogate variance, it is questionable if introducing more variance is worth compared to accepting a small bias. In fact, the variance of $\hat{\beta}_{cor}$ is always the largest. Numerically, the algorithm is satisfying as even in the toughest settings with $\sigma_u^2 = \sigma_w^2$, medium censoring and low sample size ($n = 50$) almost 40 percent of the estimates are valid (see appendix figure 24, bottom-right pane).

The appendix contains some more simulations with lower sample size or variance and medium censoring rate. The results are similar to those described in this section.

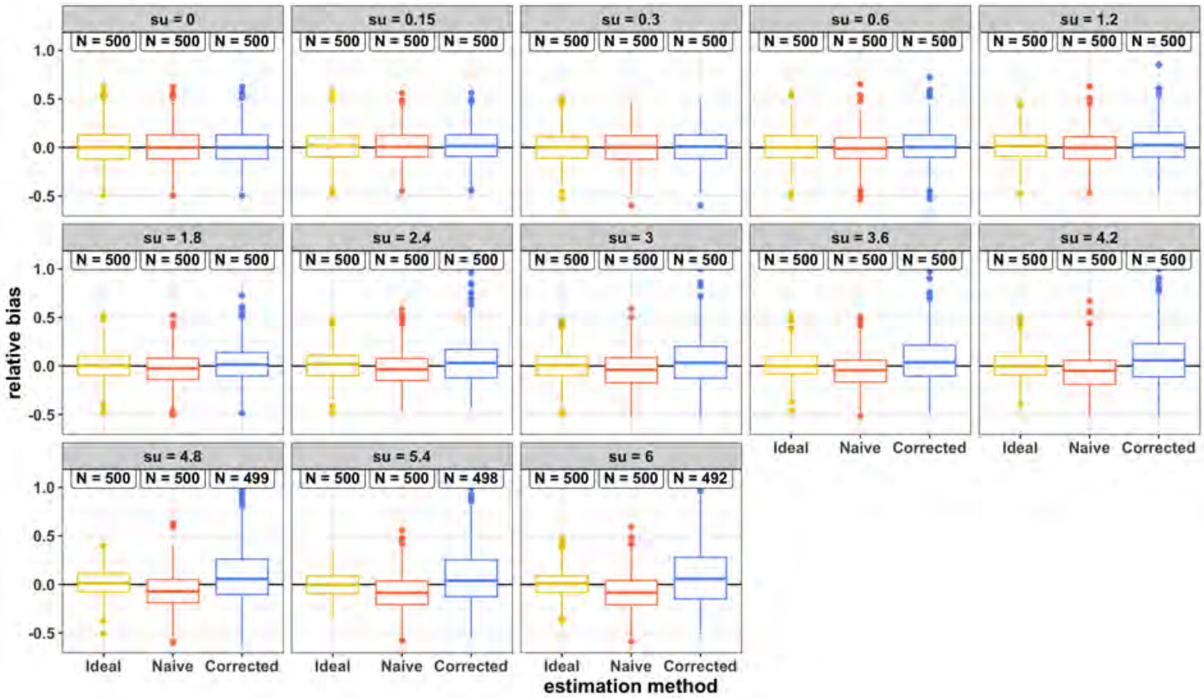

Figure 8: Estimates for β_1 with $n = 500$, $\lambda = 1/66.86$ and varying σ_u^2, Cox's PH model

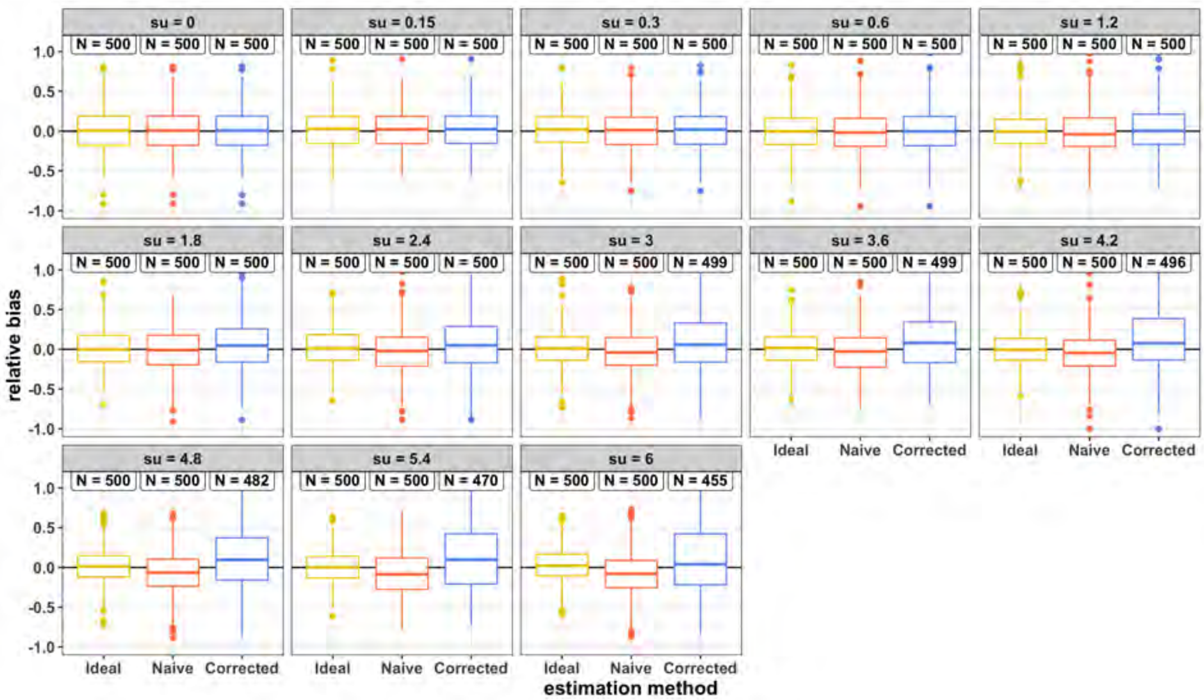

Figure 9: Estimates for β_1 with $n = 500$, $\lambda = 1/30$ and varying σ_u^2, Cox's PH model

estimation method	Corrected	Scenario 1	Scenario 2
true error variance	2.5	2.5	2.5
assumed error variance	2.5	1	4
relationship true to assumed	assumed = true	assumed < true	assumed > true

Table 3: True and assumed parameter values for simulation under false assumptions

5.4 Poisson Model under False Assumptions

The results presented so far were based on the knowledge of the measurement error. As mentioned in chapter 2.1, in practice it is very expensive and/or time consuming to obtain information about the measurement error. To address the question what happens if the true underlying error structure is misspecified, a small sensitivity analysis was performed. For a fixed true error variance of $\sigma_u^2 = 0.5\sigma_w^2 = 2.5$ with varying sample size in the Poisson model, two further estimators were calculated. In the first scenario (S1) a lower error variance of $0.2\sigma_w^2 = 1$ and in the second scenario (S2) a larger error variance of $0.8\sigma_w^2 = 4$ was assumed, which corresponds to an under- or overestimation of the true error variance respectively. Table 3 gives an overview over the true and assumed parameter values.

Figure 10 shows the ideal, naive and corrected estimators, just like before, and on the right side the estimators with assumed lower (S1) or higher (S2) error variance. The first three estimators behave like in figure 3, where σ_u^2 takes the value 3.5 instead of 2.5. Scenario 1 leads to a biased estimator that underestimates the true parameter value but is still better than the naive one. Scenario 2 also leads to a biased estimator. This time, however, the estimator overestimates the true parameter value. It seems that a higher assumed variance leads to an overestimation and vice versa. The higher the assumed error variance, the higher is also the variance of the corrected estimator, at least with larger sample sizes. Similarly, numerical problems increase and lead to fewer valid values.
Remember that only naive estimator for the intercept is biased in the Poisson model, but not the estimator for the covariate coefficient. Figure 11 shows the same as the previous figure for β_1. All versions of the corrected estimator tend towards a small positive value. This value is lower for lower assumed variances. Again, the higher the assumed error variance, the higher is the variance of the corresponding estimator. Interestingly, there seems to be a correlation of assumed error variance and drift behaviour. The estimator with assumed lower error variance (S1) reaches its stable state faster than the other two estimators. This holds for all combinations of the corrected estimators.

In summary, a wrongly assumed error variance has an impact on the corrected estimator. A higher assumed error variance leads to a higher variance of the estimator, a slower convergence to the steady state and fewer valid values as numerical problems are more likely to occur. In this simulation the estimator of scenario 2 turned out to be the worst. If σ_u^2 is assumed higher than it really is, the estimator gets biased and much more uncertain. Contrary to this, the estimator with an assumed lower error variance for β_0 is better and for β_1 as good as the naive estimator. On top of that, its variance is smaller than the corrected estimator with true assumed variance and not much larger than that of the naive one. An underestimation of the true error variance thus seems to have far less bad consequences than an overestimation.

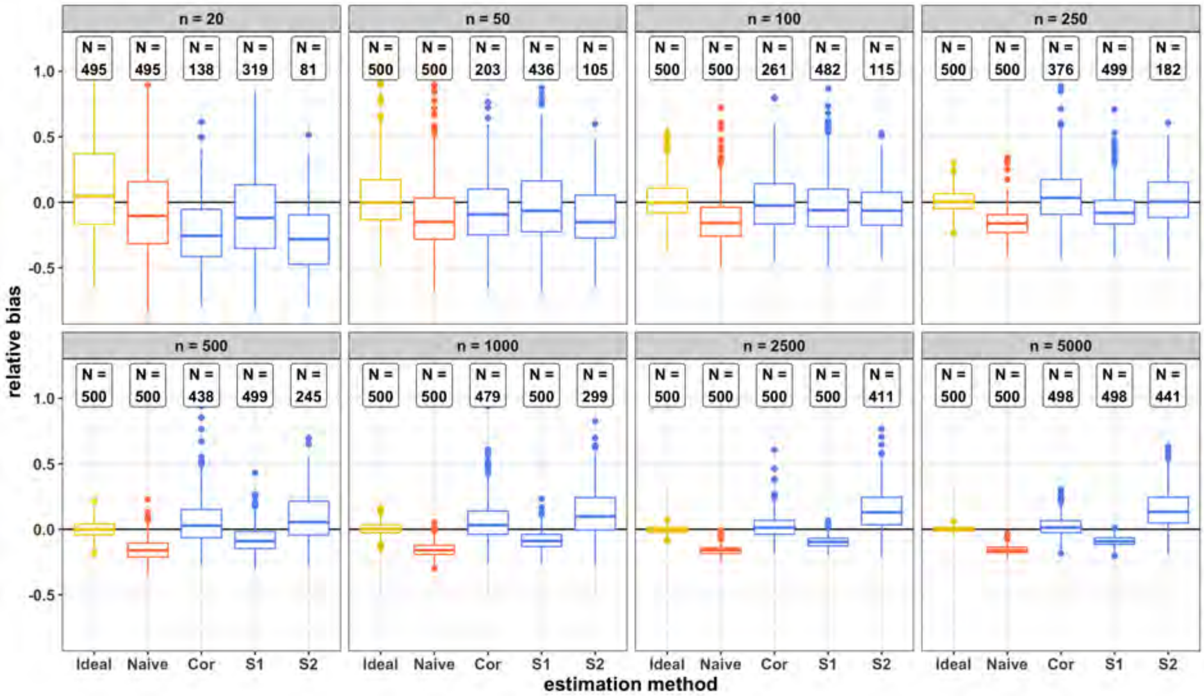

Figure 10: Estimates for β_0 under false assumptions and varying n, Poisson model

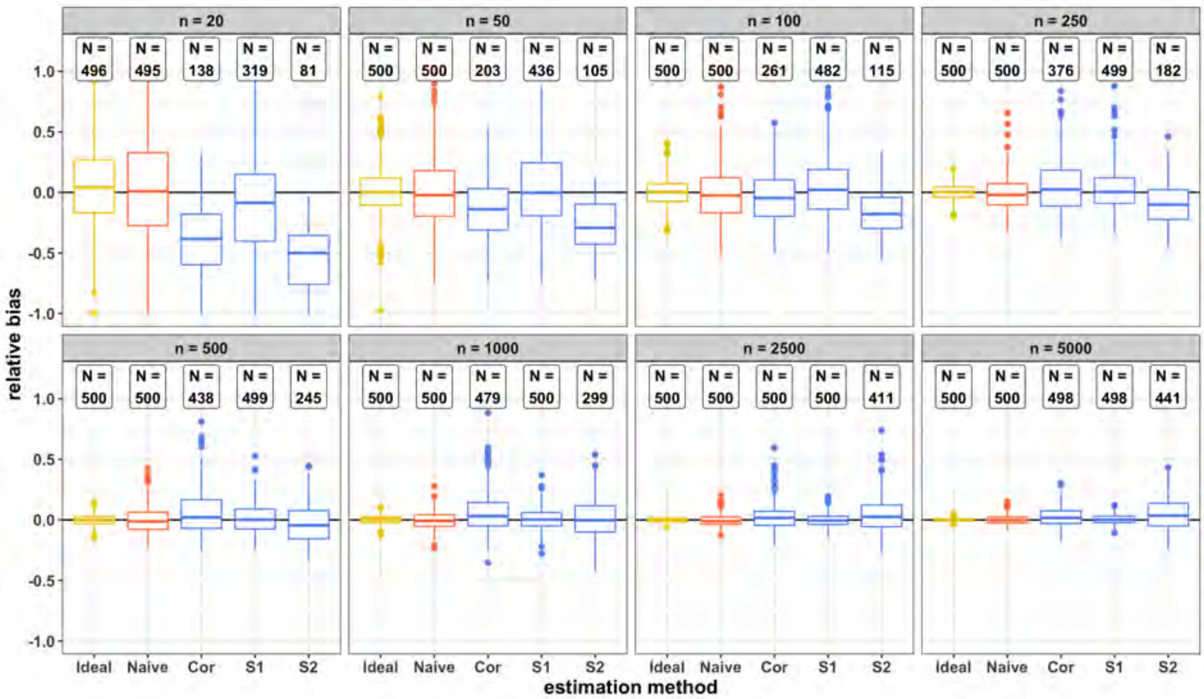

Figure 11: Estimates for β_1 under false assumptions and varying n, Poisson model

6 Concluding Remarks

The aim of this work was to investigate whether the method of corrected score functions can be applied to additive Berkson error. Therefore in the first step the corrected score functions for normal distributed error U and surrogate W were derived for linear and Poisson regression and Cox's PH model. In the next step, these functions were implemented in R in order to calculate corrected estimates. To evaluate the behavior of the corrected estimator, a simulation study was performed.
To answer the introductory question, corrected score functions are applicable to correct for additive Berkson error. Furthermore, the results of the simulation study do not object the theoretical consistency of the corrected estimator. More data or less error variance reduce the variance of the corrected estimator and help reaching the steady state faster. The other way around, less data or higher error variance increase the variance, cause more numerical problems and may even introduce a bias to the model.
As section 5.4 showed, the corrected estimator is sensitive towards misspecification. Hence, in practice a measurement error correction for additive Berkson error with corrected score functions should be based on thorough information about the error. It would be interesting to analyse the behavior of the corrected estimator under different misspecification settings. For example when a false error distribution is assumed or maybe even when the sample variance of the surrogates differs from the true surrogate variance. This is also true for other true parameter values, since the true parameters in this simulation study were determined, so that in all cases some estimates are valid.
Another point to discuss, is the censorship distribution in Cox's PH model. Since the exponential distribution is right-skewed and the event times imitate the age of adult miners in years, mainly small event times get censored. In the same context, the effects of a high censoring share of maybe 60 or 80 percent could be studied. When talking about the true parameter values in Cox's PH model, of course also the underlying model can be varied. For example a Cox-Gompertz model could be used, like in Bender et al. [2005].
The models in this thesis were limited to a response and a covariate variable, which may not be realistic in practice. The method of corrected score functions could be examined in models with more erroneous and/or correct covariates and different distribution of the response variable, e.g. logistic regression.
Speaking of more coefficients and larger models, the factors that cause the numerical problems are worth analysing. Subsection 5.4 proved that there are scenarios, where an estimate was valid and on the same dataset solely the increase of the assumed variance led to an invalid value. This can be seen by comparing the number of valid estimates for the three corrected estimators in a single pane of figures 10 or 11. As mentioned earlier in part 4.2.3, also the value of the term $\beta_0 + \beta_1 X_i$ seems to be correlated with diverging algorithm runs.
All in all, the method of corrected score functions is applicable to additive Berkson error and offers potential for further research.

A Further Graphics

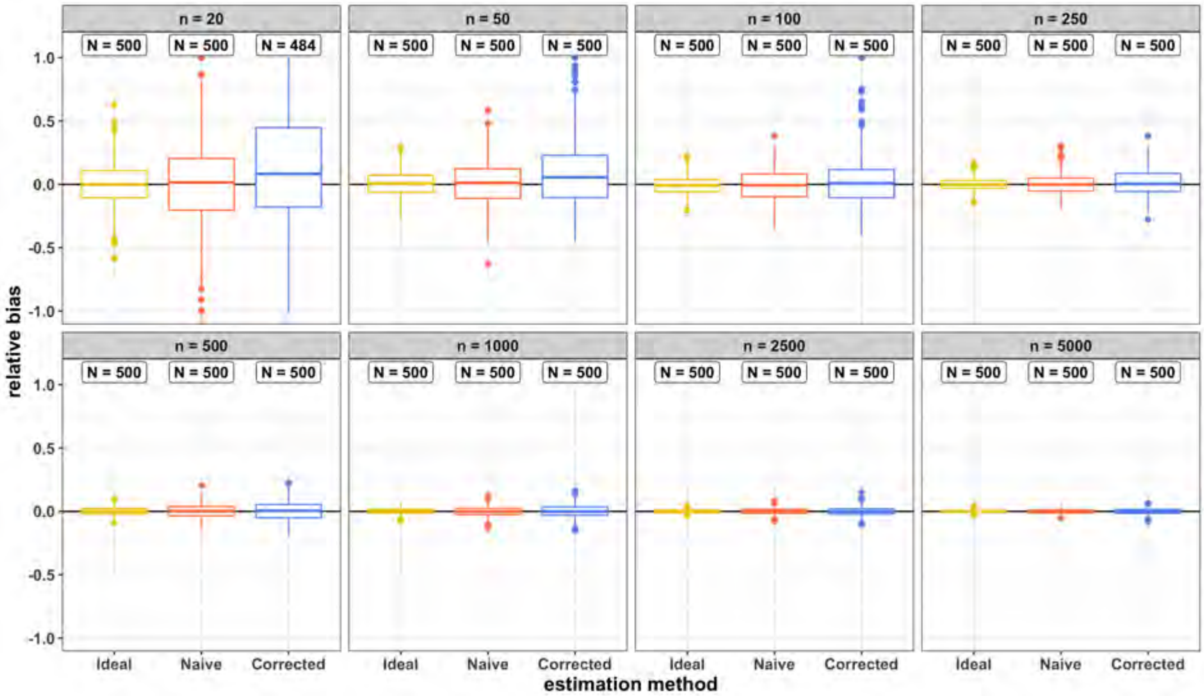

Figure 12: Estimates for β_1 with $\sigma_u^2 = 3.5$ and varying n, linear model

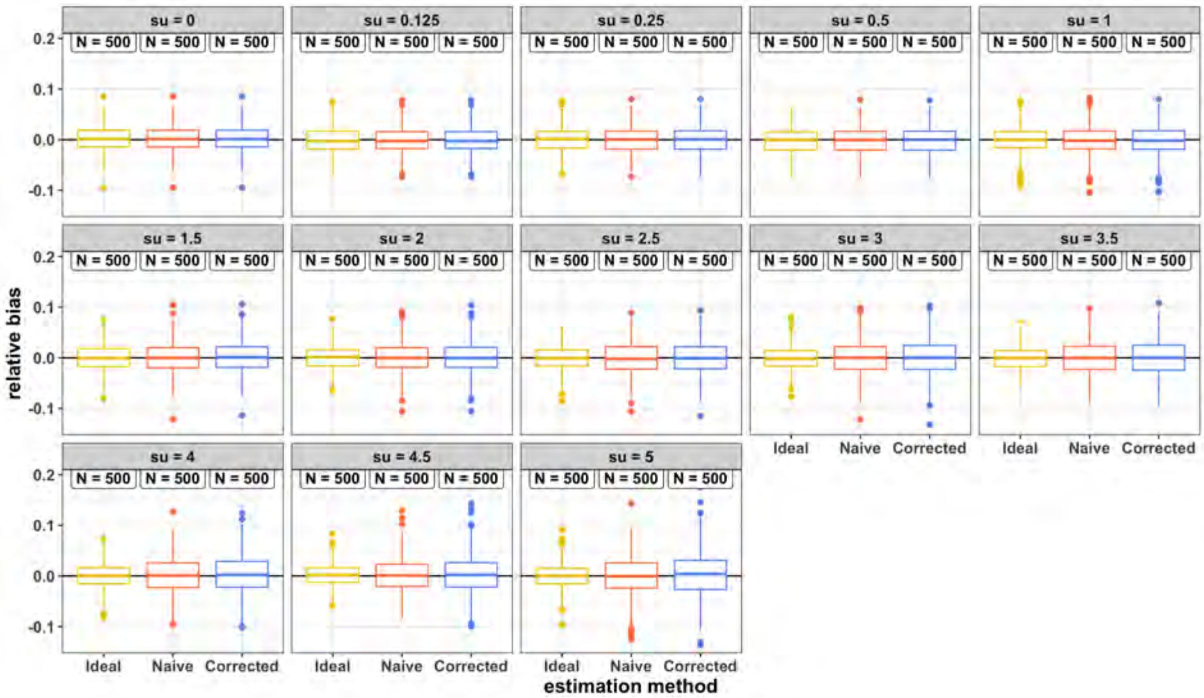

Figure 13: Estimates for β_0 with $n = 500$ and varying σ_u^2, linear model

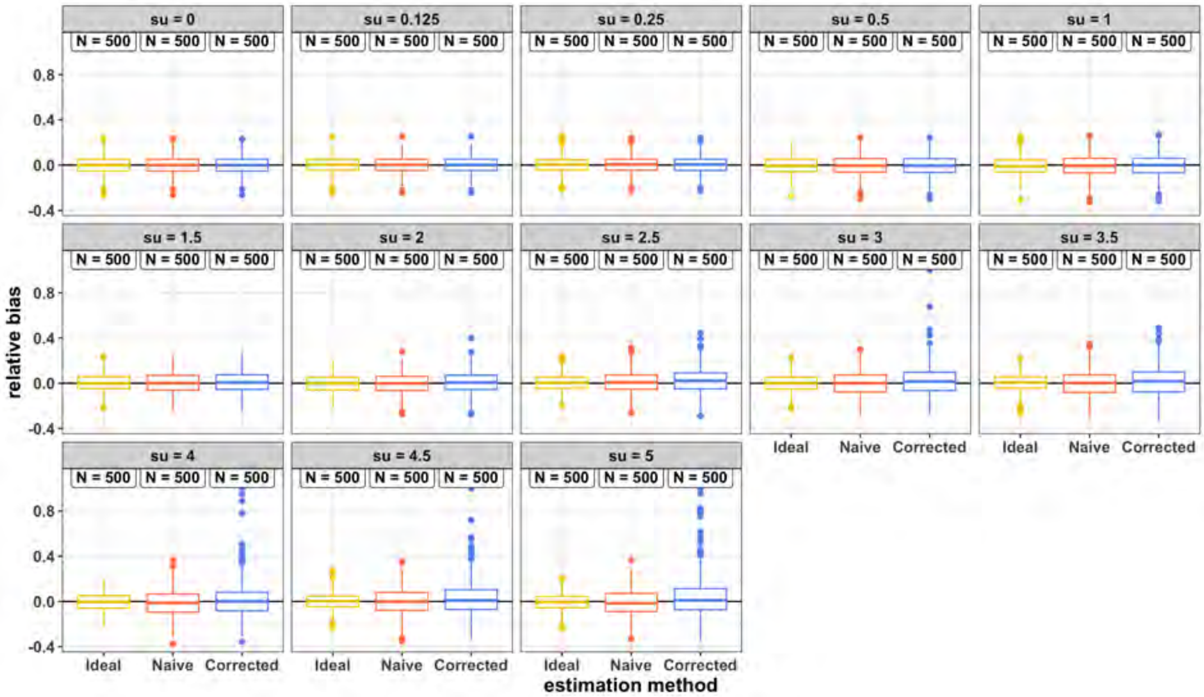

Figure 14: Estimates for β_0 with $n = 50$ and varying σ_u^2, linear model

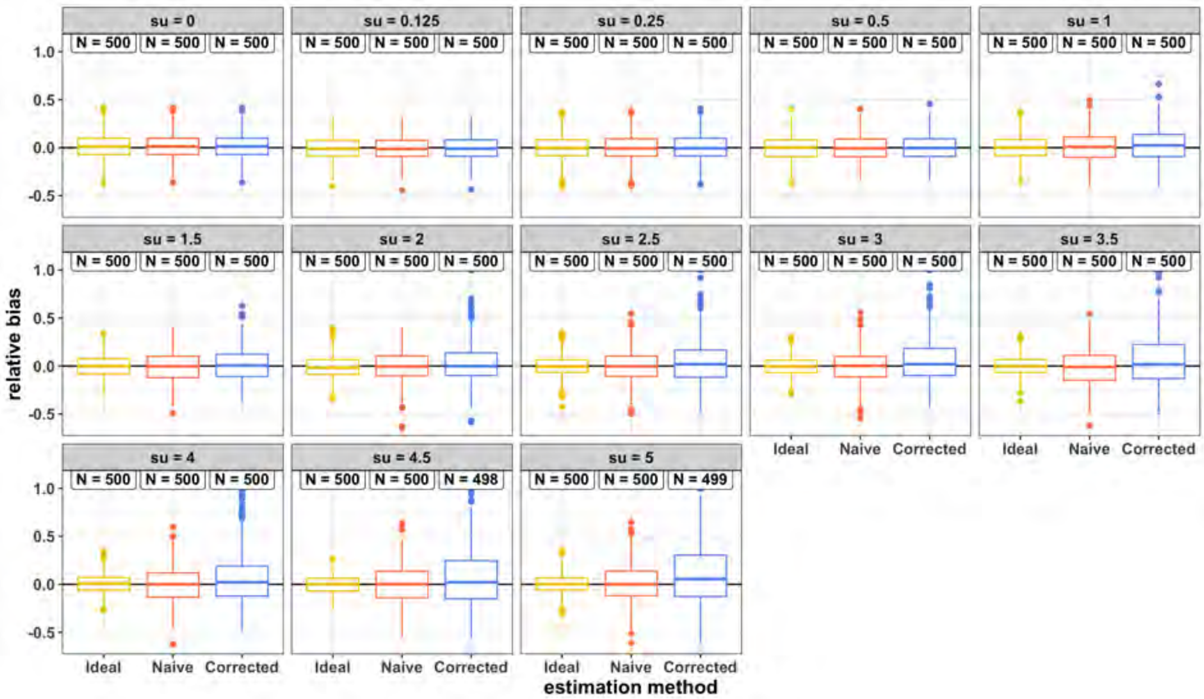

Figure 15: Estimates for β_1 with $n = 50$ and varying σ_u^2, linear model

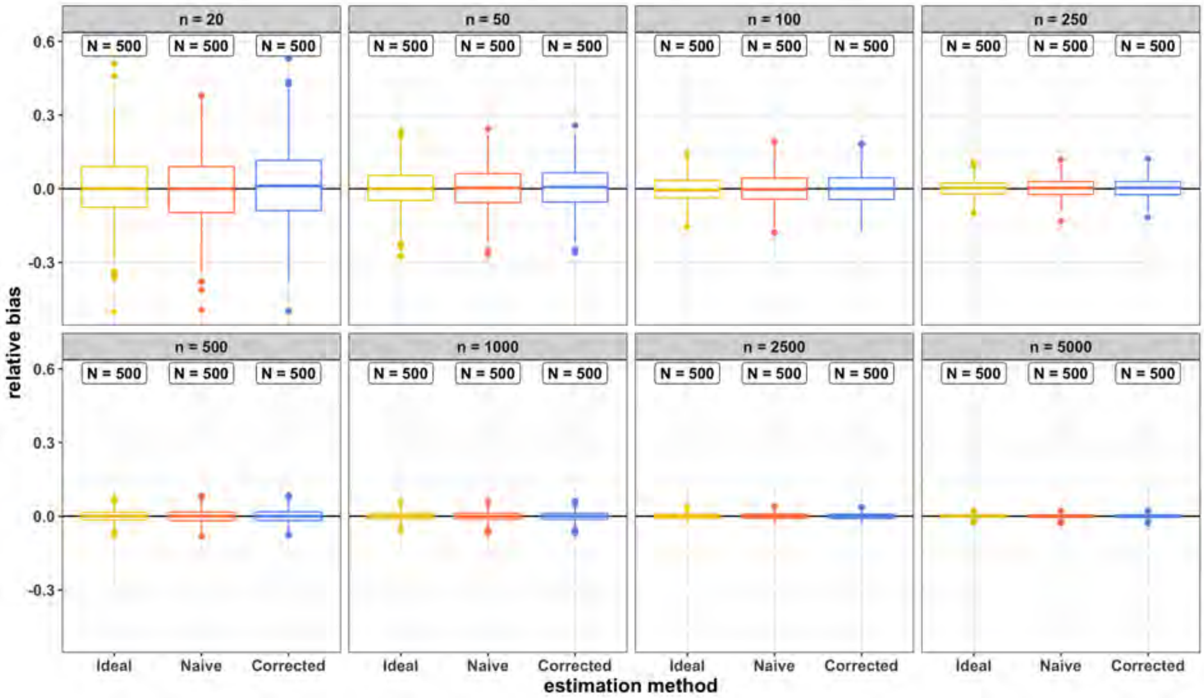

Figure 16: Estimates for β_0 with $\sigma_u^2 = 1$ and varying n, linear model

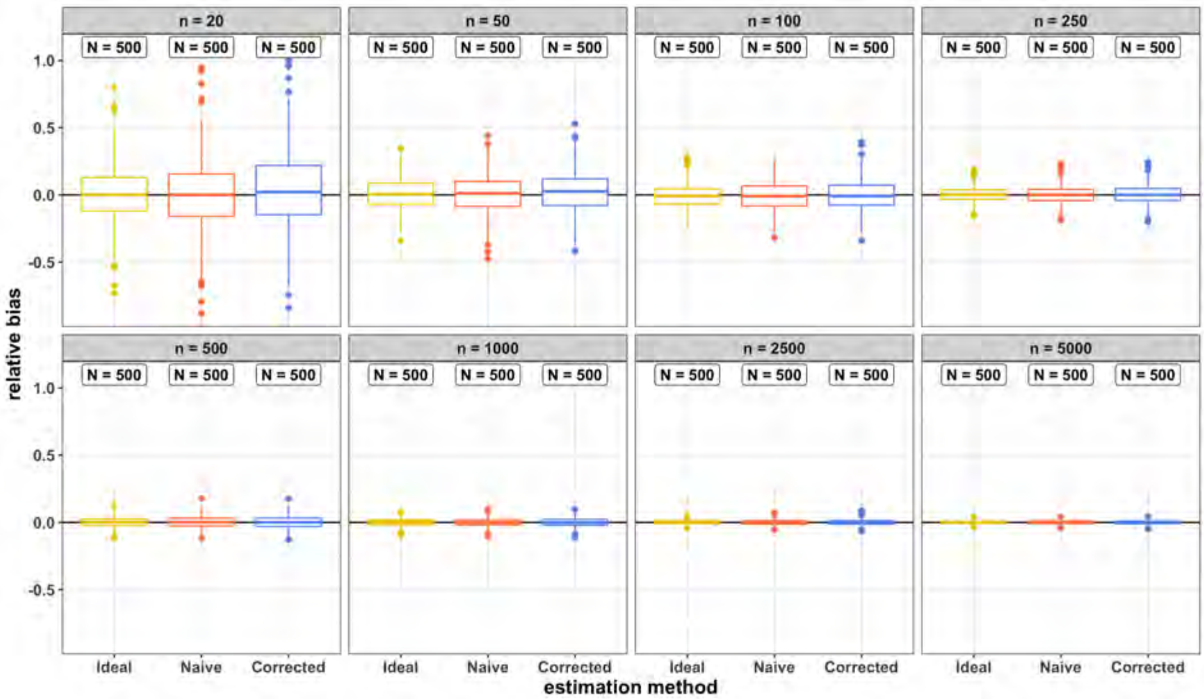

Figure 17: Estimates for β_1 with $\sigma_u^2 = 1$ and varying n, linear model

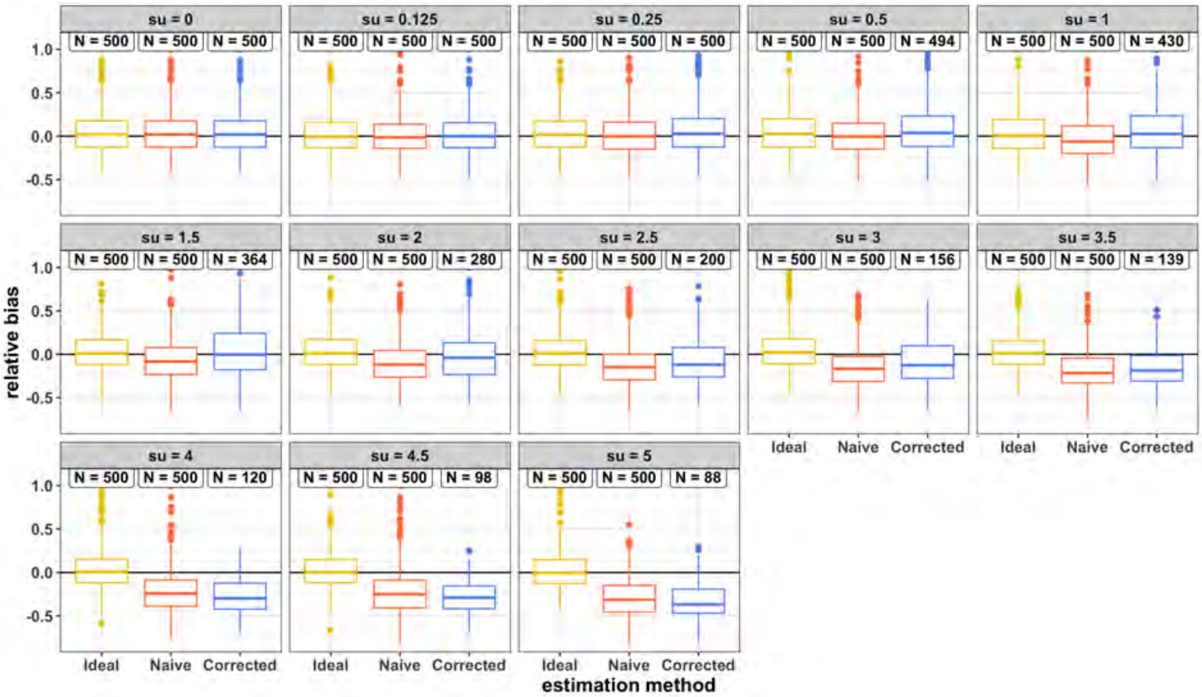

Figure 18: Estimates for β_0 with $n = 50$ and varying σ_u^2, Poisson model

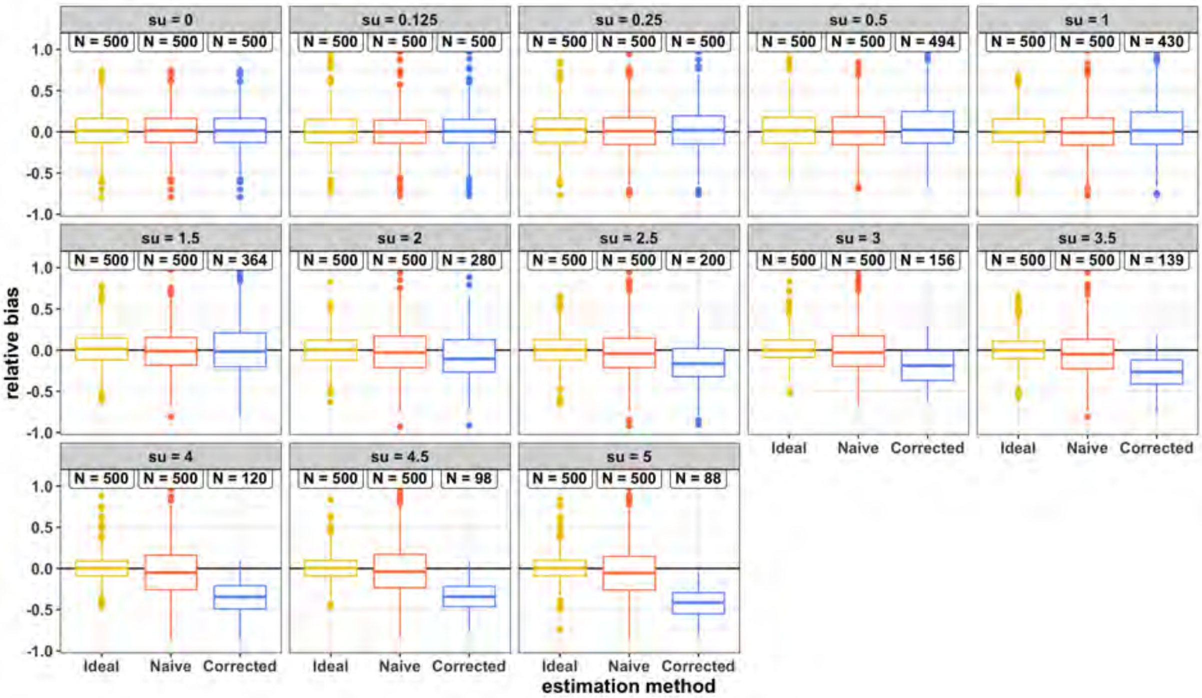

Figure 19: Estimates for β_1 with $n = 50$ and varying σ_u^2, Poisson model

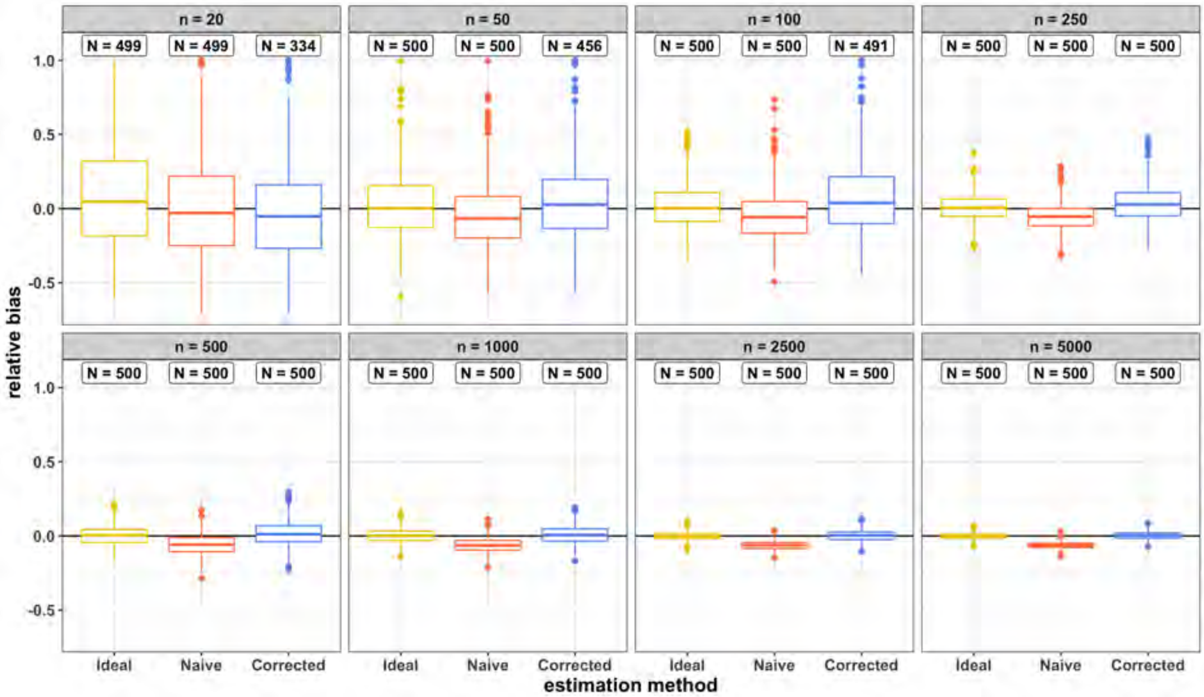

Figure 20: Estimates for β_0 with $\sigma_u^2 = 1$ and varying n, Poisson model

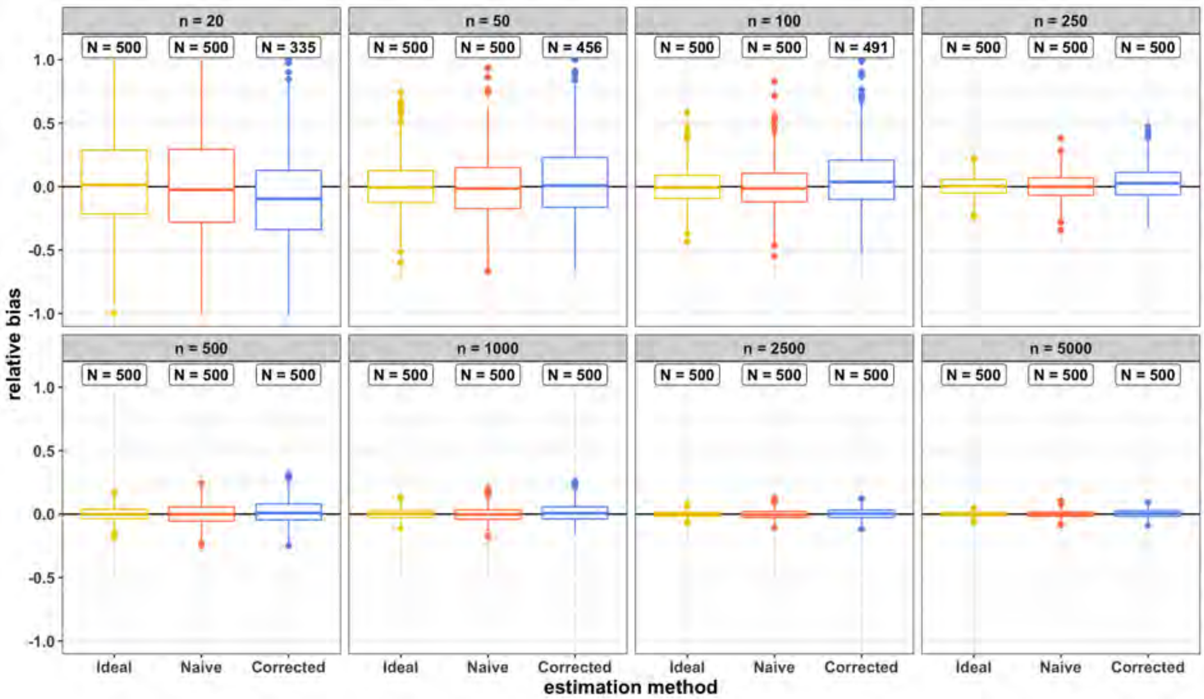

Figure 21: Estimates for β_1 with $\sigma_u^2 = 1$ and varying n, Poisson model

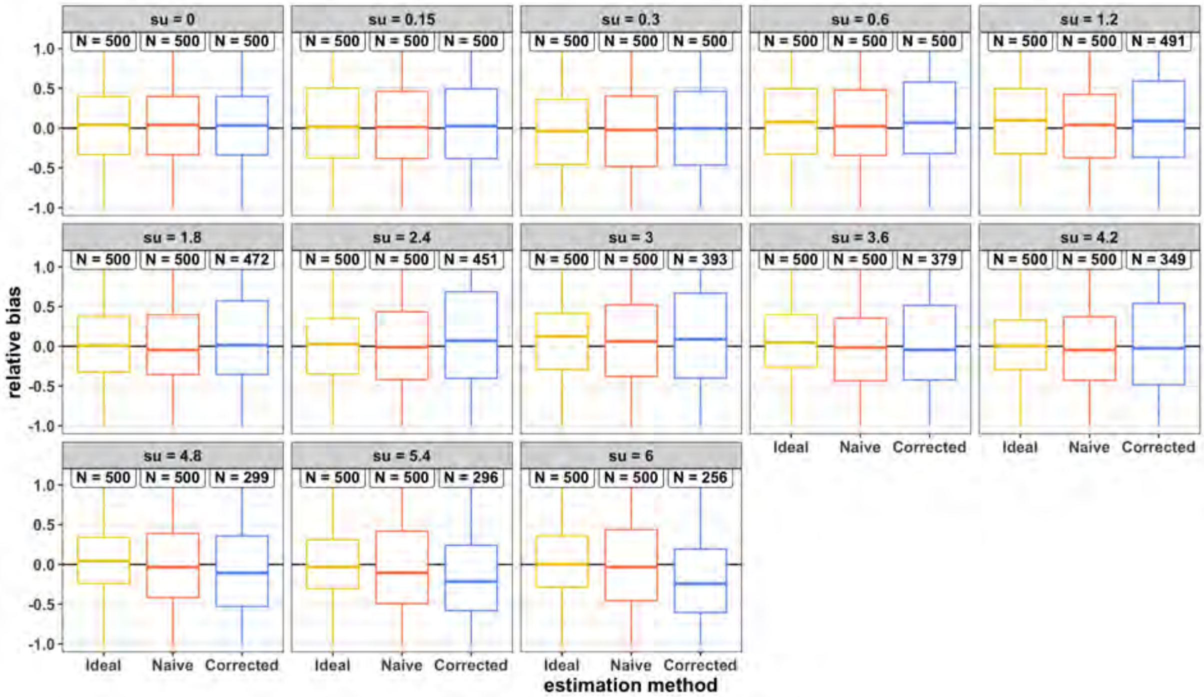

Figure 22: Estimates for β_1 with $n = 50$, $\lambda = 1/66.86$ and varying σ_u^2, Cox's PH model

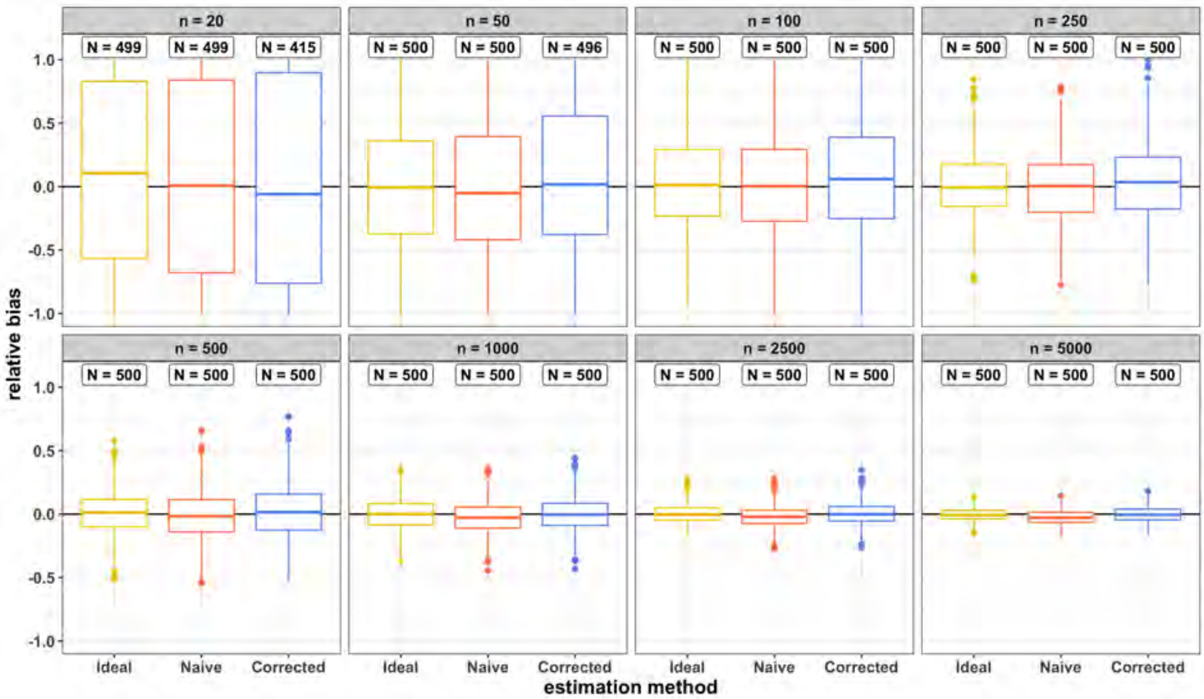

Figure 23: Estimates for β_1 with $\sigma_u^2 = 1.2$, $\lambda = 1/66.86$ and varying n, Cox's PH model

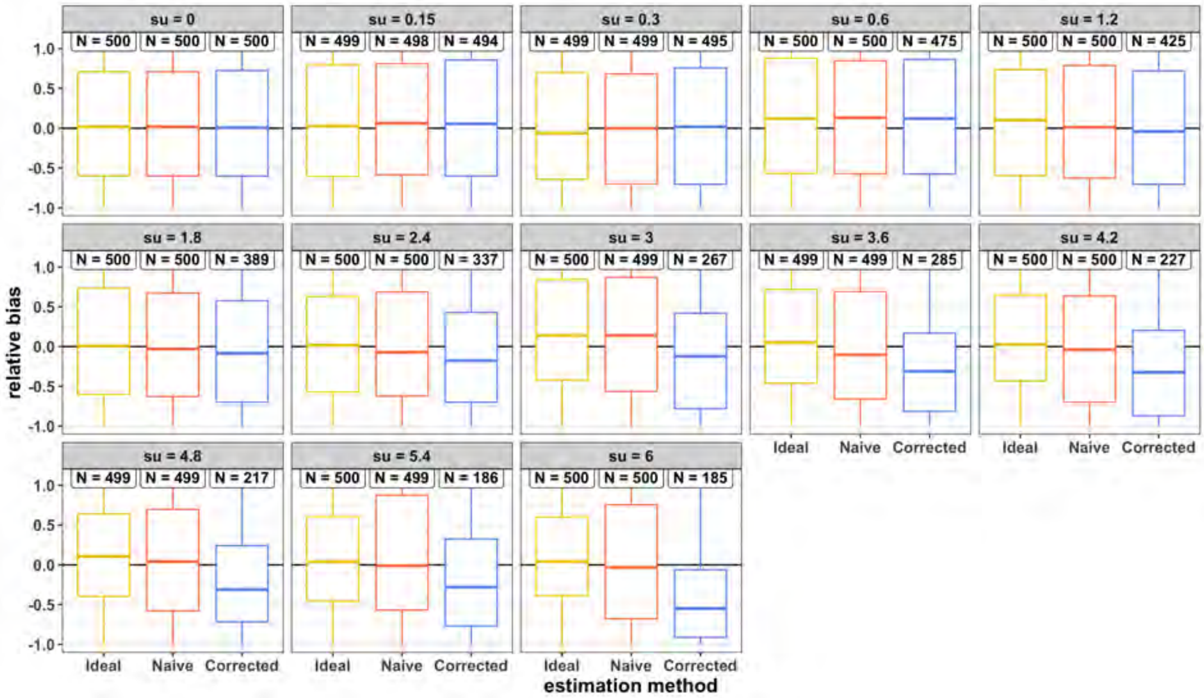

Figure 24: Estimates for β_1 with $n = 50$, $\lambda = 1/30$ and varying σ_u^2, Cox's PH model

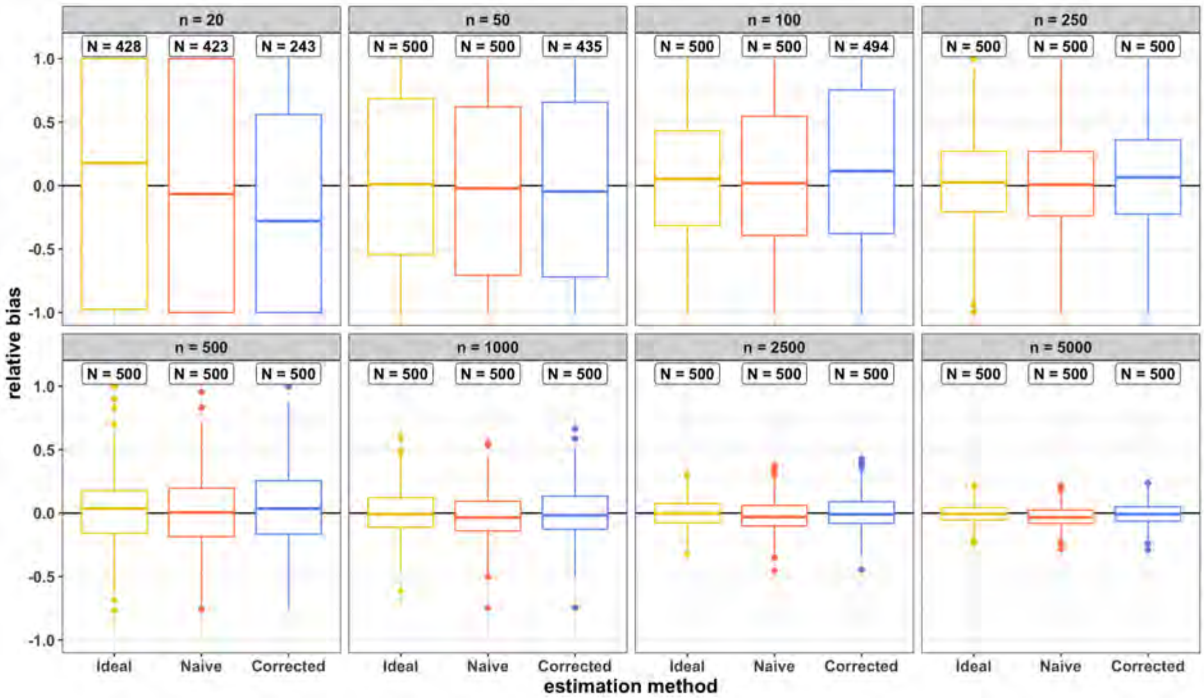

Figure 25: Estimates for β_1 with $\sigma_u^2 = 1.2$, $\lambda = 1/30$ and varying n, Cox's PH model

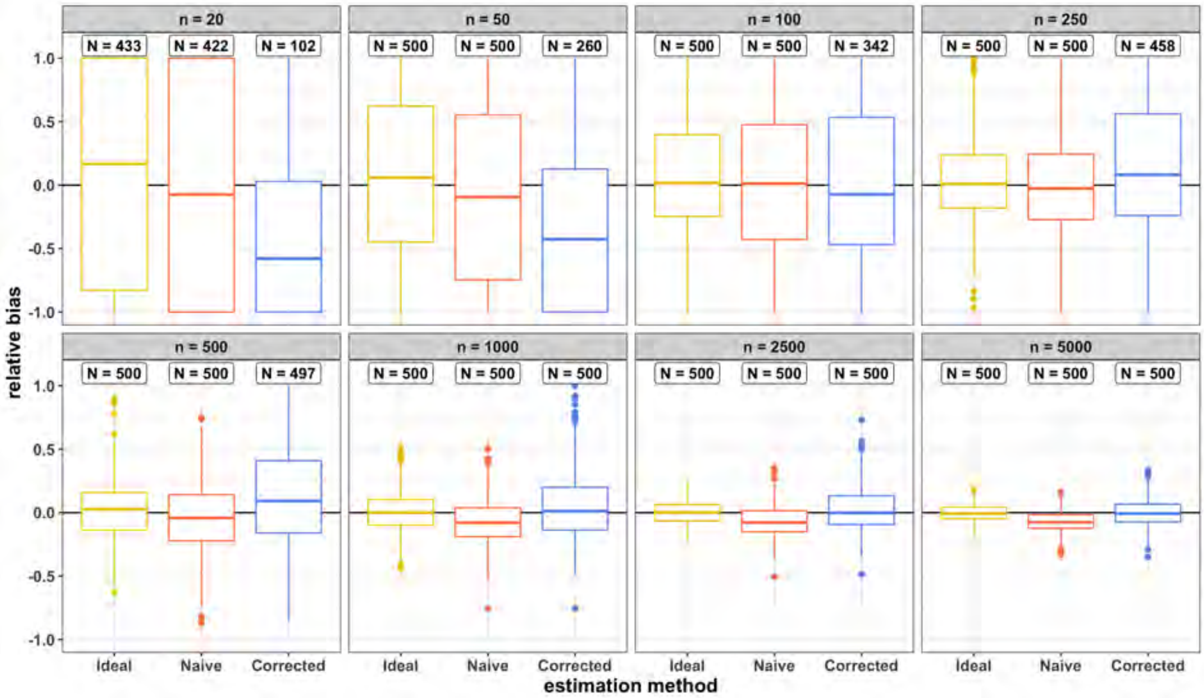

Figure 26: Estimates for β_1 with $\sigma_u^2 = 4.2$, $\lambda = 1/30$ and varying n, Cox's PH model

List of Figures

1 Estimates for β_0 with $\sigma_u^2 = 3.5$ and varying n, linear model . . . 22
2 Estimates for β_1 with $n = 500$ and varying σ_u^2, linear model . . . 22
3 Estimates for β_0 with $\sigma_u^2 = 3.5$ and varying n, Poisson model . . . 24
4 Estimates for β_1 with $\sigma_u^2 = 3.5$ and varying n, Poisson model . . . 24
5 Estimates for β_0 with $n = 500$ and varying σ_u^2, Poisson model . . . 25
6 Estimates for β_1 with $n = 500$ and varying σ_u^2, Poisson model . . . 25
7 Estimates for β_1 with $\sigma_u^2 = 4.2$, $\lambda = 1/66.86$ and varying n, Cox's PH model . . . 26
8 Estimates for β_1 with $n = 500$, $\lambda = 1/66.86$ and varying σ_u^2, Cox's PH model . . . 28
9 Estimates for β_1 with $n = 500$, $\lambda = 1/30$ and varying σ_u^2, Cox's PH model . . . 28
10 Estimates for β_0 under false assumptions and varying n, Poisson model . . . 30
11 Estimates for β_1 under false assumptions and varying n, Poisson model . . . 30
12 Estimates for β_1 with $\sigma_u^2 = 3.5$ and varying n, linear model . . . 32
13 Estimates for β_0 with $n = 500$ and varying σ_u^2, linear model . . . 32
14 Estimates for β_0 with $n = 50$ and varying σ_u^2, linear model . . . 33
15 Estimates for β_1 with $n = 50$ and varying σ_u^2, linear model . . . 33
16 Estimates for β_0 with $\sigma_u^2 = 1$ and varying n, linear model . . . 34
17 Estimates for β_1 with $\sigma_u^2 = 1$ and varying n, linear model . . . 34
18 Estimates for β_0 with $n = 50$ and varying σ_u^2, Poisson model . . . 35
19 Estimates for β_1 with $n = 50$ and varying σ_u^2, Poisson model . . . 35
20 Estimates for β_0 with $\sigma_u^2 = 1$ and varying n, Poisson model . . . 36
21 Estimates for β_1 with $\sigma_u^2 = 1$ and varying n, Poisson model . . . 36
22 Estimates for β_1 with $n = 50$, $\lambda = 1/66.86$ and varying σ_u^2, Cox's PH model . . . 37
23 Estimates for β_1 with $\sigma_u^2 = 1.2$, $\lambda = 1/66.86$ and varying n, Cox's PH model . . . 37
24 Estimates for β_1 with $n = 50$, $\lambda = 1/30$ and varying σ_u^2, Cox's PH model . . . 38
25 Estimates for β_1 with $\sigma_u^2 = 1.2$, $\lambda = 1/30$ and varying n, Cox's PH model . . . 38
26 Estimates for β_1 with $\sigma_u^2 = 4.2$, $\lambda = 1/30$ and varying n, Cox's PH model . . . 39

Bibliography

T. Augustin. An exact corrected log-likelihood function for cox's proportional hazards model under measurement error and some extensions. *Scandinavian Journal of Statistics*, 31(1):43–50, 2004. ISSN 03036898, 14679469. URL `http://www.jstor.org/stable/4616810`.

R. Bender, T. Augustin, and M. Blettner. Generating survival times to simulate cox proportional hazards models. *Statistics in Medicine*, 24(11):1713–1723, 2005. doi: https://doi.org/10.1002/sim.2059. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2059`.

R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. Measurement error in nonlinear models: A modern perspective, second edition. pages 2, 25–29, 42–46 169–174, 359–360, 364–373, 2006. URL `https://doi.org/10.1201/9781420010138`.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972. doi: https://doi.org/10.1111/j.2517-6161.1972.tb00899.x. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1972.tb00899.x`.

J. Gauß. Korrigierte Scorefunktionen unter nichtnormalverteiltem Messfehler. pages 11–12, 32–54, 2020. URL `https://epub.ub.uni-muenchen.de/75179/1/BA_Gauss.pdf`.

R. H. H. Groenwold and O. M. Dekkers. Measurement error in clinical research, yes it matters. *European Journal of Endocrinology*, 183(3):E3 – E5, 2020. doi: 10.1530/EJE-20-0550. URL `https://eje.bioscientifica.com/view/journals/eje/183/3/EJE-20-0550.xml`.

G. Kauermann, H. Küchenhoff, and C. Heumann. Statistical foundations, reasoning and inference. pages 13–14, 313–318, 2021. URL `https://doi.org/10.1007/978-3-030-69827-0`.

F. H. Kong and M. Gu. Consistent estimation in cox proportional hazards model with covariate measurement errors. *Statistica Sinica*, 9(4):953–969, 1999. ISSN 10170405, 19968507. URL `http://www.jstor.org/stable/24306629`.

M. Kreuzer, B. Grosche, A. Brachner, K. Martignoni, M. Schnelzer, H.-J. Schopka, I. Brüske-Hohlfeld, H.-E. Wichmann, and W. Burkart. The german uranium miners cohort study: Feasibility and first results. *Radiation Research*, 152(6):S56–S58, 1999. ISSN 00337587, 19385404. URL `http://www.jstor.org/stable/3580115`.

T. P. Morris, I. R. White, and M. J. Crowther. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11):2074–2102, 2019. doi: https://doi.org/10.1002/sim.8086. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8086`.

T. Nakamura. Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77(1):127–137, 1990. ISSN 00063444. URL `http://www.jstor.org/stable/2336055`.

J. Nocedal and S. J. Wright. Numerical optimization, second edition. pages 238–240, 2006. URL `https://doi.org/10.1007/978-0-387-40065-5`.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

L. A. Stefanski. Unbiased estimation of a nonlinear function a normal mean with application to measurement error models. *Communications in Statistics - Theory and Methods*, 18(12):4335–4358, 1989. URL `https://doi.org/10.1080/03610928908830159`.

A. Swaminathan. Implementing the cox model in r: Bridging the gap between math and code. *Medium [Online]*, 2021. URL `https://medium.com/analytics-vidhya/implementing-the-cox-model-in-r-b1292d6ab6d2`. Accessed on 2022-05-30.

Declaration of Originality

I confirm that the submitted thesis is original work and was written by me without further assistance. Appropriate credit has been given where reference has been made to the work of others. The thesis was not examined before, nor has it been published. The submitted electronic version of the thesis matches the printed version.