Covariate Selection Considering Measurement Error with Application in Accelerated Life Testing

Samira Karimi, PhD Student, University of Arkansas

Haitao Liao, PhD, University of Arkansas

Edward Pohl, PhD, University of Arkansas

Key Words: Measurement error, Corrected likelihood, Score function, Weibull distribution, Lognormal distribution

SUMMARY & CONCLUSIONS

Covariates (e.g., temperature, humidity, and electric current) are those factors affecting the outcome under study. Practitioners, such as reliability engineers, are often faced with measurement errors in covariates and/or unimportant covariates when collecting data on important variables. Such errors and unimportant covariates usually lead to low-quality estimation results and significantly increase computational efforts. To make the best use of data, it is essential to reduce the negative impact of measurement errors in covariates and eliminate those unimportant covariates, so that an adequate model with accurate and precise model parameter estimates can be obtained. A typical example involving measurement errors in covariates is accelerated life testing (ALT). Even in a laboratory testing environment, the exact measurements of covariates cannot be guaranteed. Moreover, the test conditions may not be perfectly controlled, and bringing the conditions back to the required levels may take some time. Consequently, these affect the reliability estimation for a product under investigation. Considering failure-time data collected from the field, the negative impact of measurement errors in covariates is even more significant. Besides a number of known accelerating variables, it is beneficial to monitor some other conditions that might also be influential on the product's reliability. However, it is often difficult to tell which variables are actually important prior to data analysis. To overcome this challenge, variable selection needs to be considered in order to reduce the model complexity. In this work, both Weibull and Lognormal regression models are studied for modeling ALT data with measurement errors in covariates. The numerical results validate the proposed method for handling measurement errors in covariates and for eliminating unimportant covariates.

1 INTRODUCTION

Covariates are those factors affecting the outcome under study. In the analysis of ALT data, such covariates (e.g., temperature, humidity, and electric current) affect the lifetime distribution of a product. When collecting ALT data, engineers often encounter measurement errors in covariates. Extensive research has been conducted on the analysis of different types of ALT data, but reliability estimation in the presence of measurement errors in multiple covariates has not been well studied. Moreover, it may be beneficial to monitor some other conditions that might also be influential on the product's reliability. However, it is often difficult to tell which variables are actually important prior to data analysis, especially when measurement errors are unavoidable. Indeed, such errors and unimportant covariates may lead to low-quality estimation results and complex models. To facilitate data analysis in ALT, it is essential to reduce the negative impact of measurement errors in covariates and eliminate those unimportant covariates.

Detection of measurement error has been investigated for other statistical purposes. Important basics of modeling measurement error are well provided by Fuller [1]. So far, only a few statistical methods have been employed for variable selection in the presence of measurement error. For example, a variant of Lasso introduced by Sørensen et al. [2] considers independent, identically and normally distributed additive measurement errors. Another method called CoCoLasso by Datta and Zou [3] considers additive and multiplicative measurement errors. However, both methods focus only on linear models rather than general regression models. Another group of studies analyzes cases for regression models with measurement errors without variable selection [4][5]. The Measurement Error Boosting (MEBoost) algorithm introduced by Brown et al. [6] usually outperforms the variant of Lasso [2] and CoCoLasso [3] in terms of prediction error and coefficient bias. Technically, the MEBoost algorithm is developed for Normal, Poisson, Gamma and Wald regression models.

The first step in considering measurement error for generalized linear regression models is to create suitable corrected likelihood and score functions. Nakamura [7] proposes corrected score functions based on generalized linear models (i.e., Normal, Poisson, Gamma and Wald regression models). These corrected score functions are applied by Brown et al. [6]. Moreover, Novick and Stefanski [8] propose a corrected score estimation method using complex variable simulation extrapolation, and Agustin [9] proposes an exact corrected log-likelihood function for a Cox's proportional hazards model. However, the exact corrected score functions for some probability distributions widely used in reliability, such as Weibull and Lognormal, have not been well studied.

This paper studies regression models for reliability evaluation based on the Weibull and Lognormal distributions. The corresponding statistical methods analyze ALT data involving multiple candidate accelerating variables subject to measurement errors. The statistical models and estimation procedure are tested using an ALT data from the literature [10].

The remainder of this paper is organized as follows. In Section 2, the basic regression technique in the presence of covariate measurement error is provided, and the MEBoost method for variable selection is introduced. In Section 3, the Weibull and Lognormal regression models are presented, and the corresponding statistical inference methods are addressed. In Section 4, an ALT data set from the literature is modified and analyzed to demonstrate the capability of the proposed models and estimation methods in improving parameter estimation and covariate selection. Finally, Section 5 draws conclusions.

2 BACKGROUND

2.1 Regression in the presence of covariate measurement error

Modeling covariate error has been treated in two ways in the literature. The first approach focuses on functional models by assuming the covariates are fixed [11]. The second method is to use structural models by assuming the covariates are random quantities. This method is utilized in this study.

The usual regression model is $Y_{n\times1} = \boldsymbol{X}_{n\times p}\beta_{p\times1} + \epsilon_{n\times1}$, where $Y_{n\times1}$ is the response vector, $\boldsymbol{X}_{n\times p}$ is the matrix of covariates, $\beta_{p\times1}$ is the coefficient vector, and $\epsilon_{n\times1}$ is the vector of independent and identically distributed random errors. In the presence of covariate measurement errors, it is assumed that instead of observing $\boldsymbol{X}_{n\times p}$, an "error-prone" matrix $\boldsymbol{W}_{n\times p}$ is observed. Specifically, when the measurement errors are additive, $\boldsymbol{W}_{n\times p}$ can be expressed as:

$$\boldsymbol{W}_{n\times p} = \boldsymbol{X}_{n\times p} + \boldsymbol{U}_{n\times p}, \tag{1}$$

where $\boldsymbol{U}_{n\times p}$ is a random covariate error matrix. We define Λ as the covariance matrix of the measurement error matrix $\boldsymbol{U}_{n\times p}$. In this work, the elements of $\boldsymbol{U}_{n\times p}$ are assumed to be independent and normally distributed as $N(0,\Lambda)$ where Λ is a diagonal matrix. Clearly, a naïve model may be used by ignoring the measurement errors in covariates, but this could result in biased coefficient estimates and model parameters [6].

2.2 Variable selection in the presence of measurement error

In the literature, only a few models have been developed to handle variable selection with measurement errors. Ma and Li [12] study variable selection in the presence of measurement errors via a penalty function. While presenting good quality for general applications, the drawback of this method is mainly the computational inefficiency for high-dimensional problems. Faster methods such as the work by Sørensen et al. [2] analyze the presence of measurement errors using Lasso. In a more recent work by Datta and Zou [3], a correction of Lasso is proposed and referred to as CoCoLasso. It has been shown that CoCoLasso is less computationally intensive in comparison to the previous methods. The most recent work by Brown et al. [6] is called MEBoost algorithm, which is more computationally efficient and has a better performance than the previous methods. The method is developed based on the idea of ThrEEBoost [13] and uses corrected score functions [7]. The MEBoost algorithm is applied in this work due to its superior properties. Algorithm 1 shows the general form of the MEBoost algorithm, where the algorithm is stated for a regression model. The threshold parameter $\tau \in [0,1]$ provides possibility to update only coefficients with significant score function values. In particular, when $\tau = 0$ all the parameters will be updated. The set $\boldsymbol{J}_t$ determines the coefficients to be updated, and γ is the step size for updating the coefficients. As γ becomes larger, some precision will be lost. On the other hand, choosing a small value for γ will increase the computational time but can guarantee more accurate estimation results. θ represents the vector of model parameters (e.g., the scale parameter in Weibull or Lognormal) excluding the regression coefficients, β is the vector of regression coefficients, and $S^*(Y,\boldsymbol{W},\beta)^*$ is the corrected score function.

Algorithm 1: MEBoost

Procedure MEBoost
Set $\beta^0 = 0$
Set initial value for model parameters
  for $t = 0,\dots,T$ do
    Compute $v = \boldsymbol{S}^*(Y,\boldsymbol{W},\beta)_{\boldsymbol{\beta}=\boldsymbol{\beta}^{t-1}}$
    Identify $J_t = \{j: |v_j| \geq \tau.\max_j |v_j|\}$
    for all $j_t \in \boldsymbol{J}_t$ do
      update $\beta_{j_t}^{(t)} = \beta_{j_t}^{(t-1)} + \gamma.sign(v_{j_t})$
    end for
    Compute $\hat{\theta} = \{\theta: \frac{\partial l^*}{\partial \theta} = 0\}$
  end for
end procedure

2.3 Corrected score functions

Nakamura [7] proposes corrected score functions based on a number of generalized linear models. The corrections were derived based on a rule that the expectation of the corrected score function should be equal to the naïve score function. In this work, corrected score functions are derived for the Weibull and Lognormal distributions.

3 PROPOSED MODELS

3.1 Weibull regression model

In this study, a corrected score function is provided for the Weibull regression model considering covariate measurement errors. In this work, the log-location-scale parameterization of Weibull distribution is used. The probability density function (pdf) denoted as $f(.)$ and cumulative distribution function (CDF) denoted as $F(.)$ are as follows:

$$f(y) = \frac{1}{\sigma y}\phi_{sev}\left(\frac{\log y - \mu}{\sigma}\right), \tag{2}$$

$$F(y) = \Phi_{sev}\left(\frac{\log y - \mu}{\sigma}\right), \tag{3}$$

where μ is the location parameter, σ is the scale parameter, and ϕ_{sev} and Φ_{sev} are the pdf and CDF of smallest extreme value distribution:

$$\phi_{sev}(y) = \exp(y - \exp(y)), \quad (4)$$

$$\Phi_{sev}(y) = 1 - \exp(-\exp(y)). \quad (5)$$

Our regression model considers a linear relationship $\mu_i = \beta^T w_i$, for data point y_i (failure/censoring time, $i = 1, 2, \dots, n$), where w_i is the vector of covariates with a first element equal to one, β is the vector of regression coefficients. In this formulation, the model parameter excluding the coefficients is $\theta = (\sigma)$. Throughout the paper, we denote $a_i = \log y_i - \beta^T w_i$. Then the pdf of failure time and the reliability function at y_i can be defined as:

$$f(y_i) = \frac{1}{\sigma y_i} \phi_{sev}\left(\frac{a_i}{\sigma}\right), \quad (6)$$

$$R(y_i) = 1 - \Phi_{sev}\left(\frac{a_i}{\sigma}\right). \quad (7)$$

Clearly, the naïve log-likelihood function for the regression model without considering measurement errors can be expressed as:

$$l(Y, \boldsymbol{W}, \beta) = \sum_{i=1}^{n} \nu_i(-\log y_i - \log\sigma + \frac{a_i}{\sigma} - \exp\frac{a_i}{\sigma}) + (1-\nu_i)(-\exp\frac{a_i}{\sigma}) \quad (8)$$

where $\nu_i = \{1, \text{if } y_i \text{ is a failure time}; 0, \text{otherwise}\}$ is an indicator function. To deal with measurement errors, the following corrected log-likelihood function is used (the detailed derivation is omitted):

$$l^*(Y, \boldsymbol{W}, \beta) = \sum_{i=1}^{n} \nu_i(-\log y_i - \log\sigma + \frac{a_i}{\sigma}) - \exp(\frac{a_i}{\sigma} - \frac{\beta^T\Lambda\beta}{2\sigma^2}). \quad (9)$$

Technically, the corrected score function can be obtained by differentiating l^* with respect to β as follows:

$$S^*(Y, \boldsymbol{W}, \beta) = \sum_{i=1}^{n} \nu_i(-\frac{w_i}{\sigma}) - \left(-\frac{w_i}{\sigma} - \frac{\beta^T\Lambda}{\sigma^2}\right)\exp(\frac{a_i}{\sigma} - \frac{\beta^T\Lambda\beta}{2\sigma^2}). \quad (10)$$

The corrected estimate of σ is calculated by solving $\frac{\partial l^*}{\partial\sigma} = 0$. Indeed, the following equation is solved numerically:

$$\sum_{i=1}^{n} \nu_i(-\frac{1}{\sigma} - \frac{a_i}{\sigma^2}) - \left(\frac{-(\log y_i - \beta^T w_i)}{\sigma^2} + \frac{\beta^T\Lambda\beta}{\sigma^3}\right) exp\left(\frac{\log y_i - \beta^T w_i}{\sigma} - \frac{\beta^T\Lambda\beta}{2\sigma^2}\right) = 0. \quad (11)$$

In particular, a bisection search method can be used for this purpose. Note that the derived formulas in equations (10)-(11) will be substituted in Algorithm 1.

3.2 Lognormal regression model

The pdf of Lognormal distribution is:

$$f(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right). \quad (12)$$

Let y_i ($i = 1, 2, \dots, n$) be the failure/censoring time. The pdf and reliability function for the regression model with $\mu_i = \beta^T w_i$ can be expressed as:

$$f(y_i) = \frac{1}{y_i\sigma\sqrt{2\pi}} \exp(-\frac{(a_i)^2}{2\sigma^2}), \quad (13)$$

$$R(y_i) = \frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{a_i}{\sqrt{2}\sigma}\right), \quad (14)$$

The log-likelihood function of the naive regression model is given by:

$$l(Y, \boldsymbol{W}, \beta) = \sum_{i=1}^{n} \nu_i\left[-\log y_i - \log\sigma - \frac{1}{2}\log(2\pi) - \frac{(a_i)^2}{2\sigma^2}\right] + (1-\nu_i)\left[\log\left(\frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{a_i}{\sqrt{2}\sigma}\right)\right)\right]. \quad (15)$$

Again, the vector of model parameters other than the coefficients is $\theta = (\sigma)$. To ease presentation, we denote $g_i = \frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{\log y_i - \beta^T w_i}{\sqrt{2}\sigma}\right)$ for the rest of this paper.

For the lognormal model, we can define the corrected log-likelihood function as (the detailed derivation and approximation are omitted):

$$l^*(Y, \boldsymbol{W}, \beta) = \sum_{i=1}^{n} \nu_i\left[-\log y_i - \log\sigma - \frac{1}{2}\log(2\pi) - \frac{(a_i)^2}{2\sigma^2} + \frac{\beta^T\Lambda\beta}{2\sigma^2}\right] + (1-\nu_i)\left[\log(g_i) + \frac{\beta^T\Lambda\beta}{2\sigma^2}\exp(-\frac{a_i^2}{2\sigma^2})\right]. \quad (16)$$

It is worth pointing out that the obstacle for finding the estimate of σ is caused by the presence of censored data. For cases with non-censored failure-time data, the equations are much simpler, and the estimate can be obtained without relying on a numerical method. For a case involving censoring times, the bisection search method is utilized to obtain the estimate $\hat{\sigma}$ in every iteration of the algorithm. The corrected score function and $\hat{\sigma}$ are obtained based on the following equations:

$$S^*(Y, \boldsymbol{W}, \beta) = \sum_{i=1}^{n} \nu_i\left[\frac{1}{\sigma^2}(w_i a_i) + \frac{\beta^T\Lambda}{\sigma^2}\right] + (1-\nu_i)[\frac{1}{g_i}\exp\left(-\frac{(a_i)^2}{2\sigma^2}\right)\frac{w_i}{\sqrt{2\pi}\sigma} + \frac{\beta^T\Lambda}{\sigma^2}\exp\left(-\frac{a_i^2}{2\sigma^2}\right) + \frac{a_i w_i \beta^T\Lambda\beta}{\sigma^4}\exp(-\frac{a_i^2}{2\sigma^2})]. \quad (17)$$

$$\frac{\partial l^*(Y, \boldsymbol{W}, \beta)}{\partial\sigma} = \nu_i\left[\frac{-1}{\sigma} + \frac{(a_i)^2}{\sigma^3} - \frac{\beta^T\Lambda\beta}{\sigma^3}\right] + (1-\nu_i)\left[\frac{1}{g_i}\exp\left(-\frac{(a_i)^2}{2\sigma^2}\right)\frac{a_i}{\sqrt{2\pi}\sigma} + -\frac{\beta^T\Lambda\beta}{\sigma^3}\exp\left(-\frac{(a_i)^2}{2\sigma^2}\right) + \frac{a_i^2\beta^T\Lambda\beta}{\sigma^4}\exp\left(-\frac{a_i^2}{2\sigma^2}\right)\right] = 0. \quad (18)$$

4 NUMERICAL STUDY

4.1 Real-world example

The reliability data set from an ALT experiment for tantalum electrolytic capacitors with temperature and voltage as the covariates [10] is modified and studied in this work. The original data consists of 2204 data points with 2164 censored data and only 40 failure times. To demonstrate the capability of the model in variable selection, a new covariate (called Cov 3) is randomly added to the data to mimic a real-world case. It is expected that the model can identify this covariate as being insignificant. Moreover, some random errors $\boldsymbol{\delta_v}$ and $\boldsymbol{\delta_t}$ are added to the existing covariates (i.e., voltage and temperature). The random errors are generated from $N(0,0.15)$ distribution, where 0.15 is the assumed variance of measurement error on each covariate. Table 1 shows the modified ALT data. Then, the proposed method is tested to find how well it is capable of eliminating the unimportant covariate(s) and to investigate the capability of the proposed method in dealing with random measurement errors in these covariates.

As the first step, by using the mean covariates (e.g., 35

Volts, 57 Volts, etc.) in the modified data, probability plots are used to roughly guide the selection of underlying failure-time distribution. Figures 1 and 2 illustrate the Weibull probability plot and the Lognormal probability plot, respectively. One can see that the Weibull distribution is more appropriate in explaining the data.

Table 1 ALT data for tantalum electrolytic capacitors modified from [10] by adding random errors to voltage and temperature levels and adding randomly an extra covariate (Cov 3) (note: in the table, δ_v and δ_t were randomly generated from $N(0, 0.15)$ for each data point).

Hours	Status	Numb. devices	Voltage (Volts)	Temp. (°C)	**Cov 3**
20	Failure	1	35+$\boldsymbol{\delta_v}$	85+$\boldsymbol{\delta_t}$	**0**
90	Failure	1	35+$\boldsymbol{\delta_v}$	85+$\boldsymbol{\delta_t}$	**5**
700	Failure	1	35+$\boldsymbol{\delta_v}$	85+$\boldsymbol{\delta_t}$	**10**
37000	Failure	1	35+$\boldsymbol{\delta_v}$	85+$\boldsymbol{\delta_t}$	**0**
37000	Censored	996	35+$\boldsymbol{\delta_v}$	85+$\boldsymbol{\delta_t}$	**5**
…	…	…	…	…	**…**
8900	Failure	1	57+$\boldsymbol{\delta_v}$	45+$\boldsymbol{\delta_t}$	**10**
8900	Censored	49	57+$\boldsymbol{\delta_v}$	45+$\boldsymbol{\delta_t}$	**5**

Note that the random errors were added to the covariate values before transformation to mimic real-world practice. The transformations of the covariates are done based on the ALT model addressed in [14] where $w_1 = \log(volt)$ and $w_2 = 11605/(temp + 273.15)$. Moreover, w_3 representing the third covariate is the normalized form of Cov 3. In this work, two different values of var_U (i.e., 0.15 [true] and 0.75) are considered for the variance of measurement errors when estimating the model parameters. The goal is to show the impact of misspecification of this important quantity on the overall estimation accuracy.

The hyperparameters in the proposed parameter estimation and variable selection procedure, including γ, T and τ, can have significant impact on the final ALT model. As a result, they have to be chosen wisely. In particular, γ is the step size that determines how much the value of the parameters can be updated in each iteration. It is conceivable that as the value of γ increases, some estimation precision will be lost, but a small value of γ may require more iterations. In practice, it is basically a trade-off between precision and computational time. In this work, an adaptive step size is suggested. This technical treatment is based on the idea of learning rate decay in neural networks. Let t be the iteration number, $t = 0, \dots, T$, and dr be the decay rate, then γ_t can be defined as:

$$\gamma_t = \frac{1}{1 + t \cdot dr}$$

The value of dr can be determined based on the requirements of a specific problem. To avoid overfitting and excess calculation time, T is determined based on the convergence of the model. In other words, a large value is considered for T, but the model stops when the solution converges. Moreover, a variety of values of τ (0.2 and 0.6) are used in this study for comparing the effect of this parameter on parameter estimation in different problem settings.

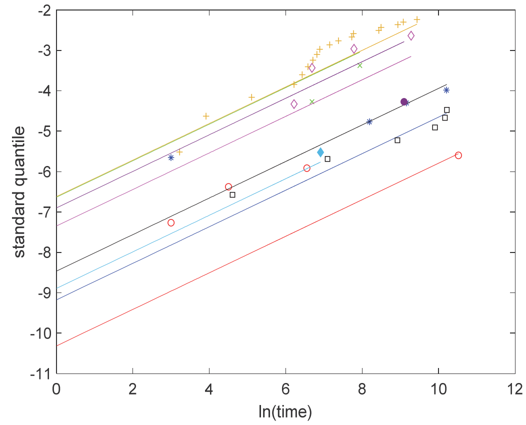

Figure 1- Weibull multiple probability plots for the tantalum capacitors dataset

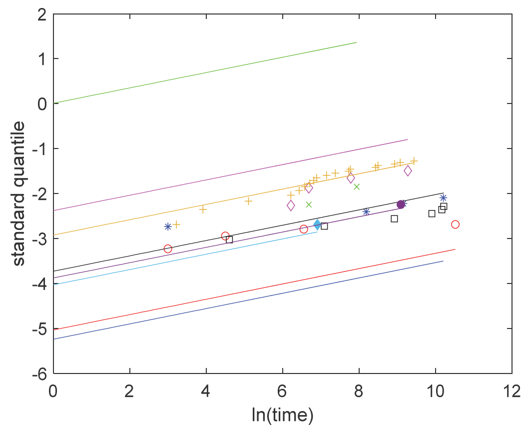

Figure 2- Lognormal multiple probability plots for the tantalum capacitors dataset

Tables 2 and 3 show the results of model estimation and variable selection considering measurement errors for the Weibull model and the Lognormal model, respectively. The estimated coefficients and parameter values for each solution are presented. For comparison, the estimation results of the Naïve models using the modified data and the models (called original model) using the original data from [10] without adding the extra covariate and measurement errors are also presented.

To select the resulting ALT model, the log-likelihood values are considered. The resulting model should have the higher log-likelihood value for each setting of var_U. In addition, to verify the effectiveness of the proposed method for model parameter estimation, the estimated parameters are compared with the ones of the original model. To this end, the combined absolute percentage error (CAPE) is calculated by $\prod_i |\frac{C_i - C_{orig-i}}{C_{orig-i}}|$, where C_i represents the estimated coefficient/parameter of the current model and C_{orig-i} is the corresponding coefficient/parameter of the original model.

From Table 2, one can see that the best Weibull model (highlighted in bold), when the variance of measurement errors is correctly specified (var_U = 0.15), is obtained with the log-likelihood of -522.78 using $\tau = 0.2$. The corresponding CAPE value is quite small. Clearly, the method identifies that the extra covariate Cov 3 with coefficient of 0.02 is insignificant. However, when the value of var_U is misspecified (var_U = 0.75), the estimation results are misleading. Therefore, it is extremely important to know the characteristics of measurement errors in covariates when dealing with such ALT data. More importantly, the Naïve model without considering measurement errors results in a lower log-likelihood and cannot identify the insignificant extra covariate. This shows the significant practical value of our proposed method in identifying unimportant covariates with measurement errors. In addition, the selection of τ can be clearly seen. Indeed, a higher value of τ reduces computational time, but in the case of Weibull model, a higher value of τ reduces our capability in covariate selection. In other words, τ is an extremely important hyperpaprameter, and thus should be selected carefully.

Table 3 shows that although the Lognormal model is less appropriate than the Weibull model for the original data (see the corresponding log-likelihood values), when the variance of measurement errors is correctly specified (var_U = 0.15), the Lognormal model to be selected becomes a better choice in terms of log-likelihood value. Moreover, when the variance of measurement errors is misspecified, the estimation results are misleading. The Naïve model results in a lower log-likelihood and cannot identify the insignificant extra covariate.

Table 2 Coefficients and parameters estimated for Weibull regression model with different parameters of the MEBoost algorithm as compared to the original and naïve models.

Variable	Original model	$var_U = 0.15$		$var_U = 0.75$		Naïve model
		$\tau = 0.2$	$\tau = 0.6$	$\tau = 0.2$	$\tau = 0.6$	
Intercept	77.37	81.20	80.10	80.00	80.00	80.00
Voltage	-16.99	-17.32	-19.44	-27.76	-27.72	-13.22
Temperature	0.18	0.04	0.36	-0.82	0.04	-0.02
Cov 3	-	0.02	-0.12	-0.18	0	0.22
$\hat{\sigma}$	2.21	1.91	2.02	3.00	4.41	4.68
CAPE	-	**0.00010**	0.00044	0.043	0.017	**0.0094**
Log-likelihood*	-538.83	**-522.78**	-527.17	518.78	-69.68	**-579.50**

Table 3 Coefficients and parameters estimated for Lognormal regression model with different parameters of the MEBoost algorithm as compared to the original and naïve models.

Variable	Original model	$var_U = 0.15$		$var_U = 0.75$		Naïve model
		$\tau = 0.2$	$\tau = 0.6$	$\tau = 0.2$	$\tau = 0.6$	
Intercept	82.46	80.57	80.46	80.42	79.21	80.64
Voltage	-17.25	-19.35	-17.53	-22.14	-21.96	-15.02
Temperature	0.16	0.35	0.13	0.53	0.57	0.24
Cov 3	-	-0.06	-0.04	0.07	0	0.2
$\hat{\sigma}$	5.87	3.54	3.36	10	10	10
CAPE	-	**0.0013**	3.12E-05	0.011	0.020	**0.0010**
Log-likelihood*	-540.81	**-426.87**	-456.23	705.87	683.18	**-588.69**

5 CONCLUSIONS

Although measurement errors and variable selection have been studied by statisticians and reliability engineers, the combination of the two problems has not been broadly investigated. In this work, the effects of measurement errors in covariates in ALT are considered. Our numerical study illustrates that appropriately dealing with measurement errors in covariates can result in more accurate reliability estimation for a product of interest. To use the MEBoost algorithm for Welbull and Lognormal regression models in the presence of measurement errors in covariates, the corrected log-likelihood and score functions are developed for analyzing ALT data. Indeed, the proposed method can be used for reliability estimation using field data where more covariates may be available with measurement errors and some of the covariates may not be significantly influential. From a practical point of view, this method, if appropriately applied, can assist a practitioner in eliminating those unimportant covariates and provide an accurate reliability estimate instead of resulting in a more complex model with detrimental power in data interpretation.

AKNOWLEDGMENTS

This research is sponsored by the U.S. National Science Foundation under the grant #1635379.

REFERENCES

[1] W. Fuller, Measurement error models, Hoboken, NJ: Wiley series in probability and statistics, 1987.

[2] Ø. Sørensen, A. Frigessi and M. Thoresen, "Measurement error in Lasso: Impact and likelihood bias correction," *Statistica sinica,* vol. 1, pp. 809-829, 2015.

[3] A. Datta and H. Zou, "Cocolasso for high-dimensional error-in-variables regression," *The Annals of Statistics,* vol. 45, no. 6, pp. 2400-2426, 2017.

[4] L. Stefanski and R. Carroll, "Covariate measurement error in logistic regression," *Ann Stat,* vol. 13, no. 4, pp. 1335-1351, 1985.

[5] B. Rosner and W. Willett, "Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error," *Am J Epidemiol,* vol. 136, no. 11, pp. 1400-1413, 1992.

[6] B. Brown, T. Weaver and J. Wolfson, "MEBoost: Variable selection in the presence of measurement error," *Statistics in medicine,* vol. 38, no. 15, pp. 2705-2718, 2019.

[7] T. Nakamura, "Corrected score function for errors-in-variables models: methodology and application to generalized linear models," *Biometrika,* vol. 77, no. 1, pp. 127-137, 1990.

[8] S. J. Novick and L. A. Stefanski, "Corrected score estimation via complex variable simulation extrapolation," *Journal of the American Statistical Association,* vol. 97, no. 458, pp. 472-481, 2002.

[9] T. Augustin, "An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions," *Scandinavian Journal of Statistics,* vol. 31, no. 1, pp. 43-50, 2004.

[10] N. D. Singpurwalla, V. C. Castellino and D. Y. & Goldschen, "Inference from accelerated life tests using eyring type re-parameterizations," *Naval Research Logistics Quarterly,* vol. 22, no. 2, pp. 289-296, 1975.

[11] J. Buonaccorsi, Measurement error: models, methods, and applications, Boca Raton, FL: Chapman and Hall/CRC, 2010.

[12] Y. Ma and R. Li, "Variable selection in measurement error models," *Bernoulli,* vol. 16, no. 1, pp. 274-300, 2010.

[13] J. Wolfson, "EEBoost: A general method for prediction and variable selection based on estimating equations," *Journal of the American Statistical Association,* vol. 106, no. 493, pp. 296-305, 2011.

[14] W. Q. Meeker and L. A. Escobar, Statistical Methods for Reliability Data, New York, N.Y.: John Wiley & Sons, Inc., 1998.

BIOGRAPHIES

Samira Karimi, PhD Student
Department of Industrial Engineering
University of Arkansas
Fayetteville, AR, 72701, USA

Email: sakarimi@email.uark.edu

Samira is a Ph.D. student in the Industrial Engineering Department at the University of Arkansas, Fayetteville, Arkansas. She received her B.Sc. (2014) and M.Sc. (2016) in Industrial Engineering from Sharif University of Technology, Tehran, Iran. Her research interests include reliability modeling and statistical analysis. She received 2017 PHM-Harbin Best Paper Award and the 2019 SRE Stan Ofsthun Best Paper Award. She is a student member of IISE and INFORMS.

Haitao Liao, PhD
Department of Industrial Engineering
University of Arkansas
Fayetteville, AR, 72701, USA

Email: liao@uark.edu

Dr. Haitao Liao is a Professor and John and Mary Lib White Systems Integration Chair in the Industrial Engineering Department at the University of Arkansas, Fayetteville, Arkansas. He received his Ph.D. degree in 2004 from the Department of Industrial and Systems Engineering at Rutgers University. His research interests focus on modeling of accelerated testing, probabilistic risk assessment, maintenance models and optimization, spare part inventory control, and prognostics. He serves as an Area Editor for the IISE Transactions on Quality and Reliability Engineering. He received a National Science Foundation CAREER Award in 2010, IISE William A. J. Golomski Award in 2011, 2014 and 2018, SRE Stan Ofsthun Best Paper Award in 2015 and 2019, and 2017 Alan O. Plait Award for Tutorial Excellence. He is a Fellow of IISE, a member of INFORMS, and a lifetime member of SRE.

Edward Pohl, PhD
Department of Industrial Engineering
University of Arkansas
Fayetteville, AR, 72701, USA

Email: epohl@uark.edu

Dr. Edward Pohl is a Professor and Head of the Department of Industrial Engineering at the University of Arkansas. He is the current holder of the 21st Century Professorship. Ed spent twenty years in the United States Air Force where he served in a variety of engineering, analysis and academic positions during his career. Ed received his Ph.D. in systems and industrial engineering from the University of Arizona, an M.S. in reliability engineering from the University of Arizona, an M.S. in systems engineering from AFIT, an M.S. in engineering management from the University of Dayton, and a B.S.E.E. from Boston University. His primary research interests are in risk, reliability, engineering optimization, healthcare and supply chain risk analysis, decision making, and quality. Ed is a Fellow of IISE, a Fellow of the Society of Reliability Engineers (SRE), a senior member of ASQ, and a senior member of IEEE. Ed serves as an Associate Editor for the Journal of Military Operations Research, the Journal of Risk and Reliability, and the IEEE Transactions on Reliability. He is a two-time winner of the Alan Plait award for Outstanding Tutorial at RAMS and the William A. J. Golomski Award.