

Mechanical Equipment Reliability Analysis: Case Study

Ali Nouri Qarahasanlou^{*1}, Reza ShakorShahabi¹, Reza Barabadi², Negar Fallahnejad¹

1. Mining Engineering, Faculty of Technical & Engineering, Imam Khomeini International University, Qazvin, Iran

2. Faculty of Mining, Petroleum & Geophysics Engineering, Shahrood University of technology, Shahrood, Iran

* ali_nouri@eng.ikiu.ac.ir

Abstract

In civil and mining industries, Wheel loaders are an important component and their cost capability at effective operation. The environmental and operational factors dramatically affect the performance of loaders. In many cases, failure data are often collected from multiple and distributed units in different operational conditions, which can introduce heterogeneity into the data. Part of such heterogeneity can be explained and isolated by the observable covariates, whose values and the way they can affect the item's reliability are known. However, some factors that may affect the item's reliability are typically unknown and lead to unobserved heterogeneity. These factors are categorized as unobserved covariates. In most reliability studies, the effect of unobserved covariates is neglected. This may lead to erroneous model selection for the time to failure of the item, as well as wrong conclusions and decisions. There is a lack of sufficient knowledge, theoretical background, and a systematic approach to model the unobserved covariate in reliability analysis. This paper aims to present a framework for reliability analysis in the presence of unobserved and observed covariates. The unobserved covariates will be analyzed using frailty models (Such as Mixed Proportional Hazard). A case will illustrate the application of the framework.

Keywords: Reliability; Observed Covariate; Unobserved Covariate; Mixed Proportional Hazard; Wheel loaders.

1. Nomenclature

| | | | |
|---------------|-----------------------------------------|-----------------|-------------------------------------------------------------------|
| AIC | Akaike information criterion | $NHPP$ | Non-Homogeneous Poisson Process |
| $h_0(t)$ | Baseline hazard rate | z_p | Precipitation (mm) |
| BIC | Bayesian information criterion | PH assumption | Proportional hazards |
| Q test | Cochran's Q test | PHM | Proportional hazards model |
| $H_0(t)$ | Cumulative hazard rate | z_{pt} | Proportionality of Truck |
| k | Degree of freedom | $R(t)$ | Reliability |
| $EMPHM$ | Extension mix-proportional hazard model | z | Risk factor |
| Γ | Gamma function | z_{rc} | Road condition |
| $h(t)$ | Hazard rate | z_{rf} | Rock fragmentation |
| iid | Independent and identically distributed | z_{rk} | Rock Kind |
| L | L likelihood function | $\hat{\beta}$ | Scale parameter of Weibull distribution |
| $MPHM$ | Mix Proportional Hazard Model | $\hat{\lambda}$ | Shape parameter of Weibull distribution |
| ln | Natural logarithm | z_t | Temperature (c) |
| $R_\theta(t)$ | Unconditional reliability function | θ | the degree of heterogeneity or The variance of Gamma distribution |
| z_{ns} | No. of service | $\hat{\eta}$ | the regression coefficient |
| | | t (hr) | Time |
| | | $z(t)$ | Time-dependent covariate |

| | |
|------------|-----------------------------------|
| α_j | Time-independent frailty function |
| TRP | Trend renewal process |
| z_{id} | Wheel loaders Code |
| z_{wc} | Whether Condition |
| z_{wp} | Working Place |
| z_{wf} | Working Shift |

2. Introduction

Wheel loaders are widely used in mines, construction projects, and waste processing fields due to their merits of high mobility, remarkable operational flexibility, and relatively low capital cost. Wheel loaders on construction sites are used for digging shallow trenches, laying pile lines, and filling pits. These activities are tedious, and the weather conditions, which depend on geographical locations, are often harsh [1]. In surface mines, wheel loaders are mainly used for removing topsoil, loading fragmented rock and ore from the production bench, and cleaning operation sites. Loading and hauling constitute 35-46% of the daily operation cost of a surface mine [2]. Since there is a huge volume of rock/soil to be moved, wheel loaders used in surface mines are very big and cost quite a large percentage of investment. This dictates that this equipment must be used efficiently to reduce the unit cost of the product. Finding enough experienced operators for this type of complex equipment has become a problem for mining companies. A wheel loader typically undertakes a typical job of loading fragmented rock from a pile into a truck or ore chute. The wheel loader completes this job with several cycles that each consist of three activities: loading, hauling or maneuvering, and dumping. But the failures will be stopped in the process, and the management will try to solve them. The question is, "How can I do it?". Reliability analysis is one of the most popular ways. In this way, engineers and managements use the previous system operation data to estimate future conditions. In many reliability studies, data sets are assumed to be homogeneous, where the failure data are independent and identically distributed [3]–[5]. However, in practice, homogeneous failure data can scarcely be found. In reality, failure data are often collected from multiple items, distributed at different locations, and working under different operational conditions (e.g., operator skill, maintenance strategies, etc.) [6]. This may introduce heterogeneity into the data [7], [8]. Differences in failure intensity are generally called heterogeneities and can be due to either observed or unobserved influence risk factors, which are called covariates [9]–[11]. Covariates describe the item's characteristics or the environment in which it operates [12], which may have different levels. For example, as a covariate on the reliability of a pump, vibration may be of high, low, or medium levels [13]. Observed covariates are factors whose effects on the failure process are known, and their associated levels are recorded with the failure data; they can be time-

dependent or time-independent. Time-dependent observed covariates vary continuously with time. Unobserved covariates are covariates whose effects on the failure process are typically unknown, or their associated levels during the operating time or at the time of the failure are not available in the failure database [14]. Unobserved covariates may lead to unobserved heterogeneity [13], [15], [16]. For example, in a production process, some pumps may have a soft foot problem due to a defect in the installation process. The soft foot problem will put the bearing in an over-stressed situation; this should be considered a covariate for reliability analysis. Suppose there is no information regarding soft foot in the failure database of the bearing. In that case, an unobserved covariate should be defined to capture the effect of the soft foot on the reliability of the bearing.

In general, due to the quality of manufacturing, installation, operation, and maintenance procedures, some items may become frailer, while others are more robust. In the presence of unobserved covariates, different items may have different levels of frailty. Unobserved covariates are typically unknown or unavailable for each item; hence, they cannot be explicitly included in the analysis. The result of our literature review revealed that, in many cases, for reasons of practicality and simplicity, unobserved covariates are eliminated during the failure data analysis [9], [11], [13], [14], [17]. However, if unobserved covariates are neglected, the reliability analysis only represents the reliability of items with an average level of frailty and not that of the individual items. In general, high-risk items (high frailty) tend to fail earlier than low-risk items (low frailty) for unobserved reasons; thus, the population composition changes over time. Hence, in time, the analysis represents the item with low frailty, and the estimated reliability increases more with time than the reliability of a randomly selected item of the population [18], [19].

The Cox regression model family, such as the proportional hazards model (PHM) and its extension, for example, the stratification approach, is the most dominant statistical approach for capturing the effect of covariates on the reliability performance of an item [6], [13], [17], [20]–[24]. In PHM, the hazard rate of an item is the product of a baseline hazard rate and a positive functional term that describes how the hazard rate changes as a function of covariates. However, the PHM is very sensitive to the omission of the covariates and cannot isolate the effect of unobserved covariates [13]. In survival analysis in medical science, the frailty model introduced by Clayton [25] and Vaupel et al. [16] describes the influence of unobserved covariates in a proportional hazards model. A frailty model is a random effects model for time variables, where the random effect (the frailty) has a multiplicative effect on the hazard [15], [23]. Gamma distribution, inverse Gaussian or exponential distribution can be used to model the frailty [6], [14], [15], [26].

Recently, some studies in the reliability field have used the frailty model to model the effect of missing covariates on the reliability of an item in making maintenance decisions [6], [9], [24], [27]. To get a glimpse of the status of the applications of the frailty model in recent years, we conducted an online search of Scopus in December 2018 with the relevant keywords. We found that 33 papers with the keyword 'frailty' and 'reliability' have been published since 2008. Of these, only seven are related to applying a frailty model in reliability engineering with a focus on maintenance purposes. Asha [28] incorporated the frailty model into the load share systems and showed that reliability analysis for a heterogeneous case could differ dramatically from that for a homogeneous setting. Xu and Li [29] obtained stochastic properties of univariate frailty models, which are a special case of multivariate frailty models, and Misra [30] used stochastic orders to compare frailty models arising from different choices of frailty distribution. Asfaw and Lindqvist [23], Finkelstein [31], Slimacek and Lindqvist [32] and Giorgio [11] modeled the frailty by the introduction of Non-Homogeneous Poisson Process (NHPP). For example, Slimacek and Lindqvist [32] used frailty to model the effect of unobservable differences between turbines, as unobserved covariates, on the reliability of wind turbines using the Poisson process. It must be mentioned that most of the available studies do not discuss how the time-dependent covariates should be handled in the frailty model; nevertheless, ignoring the time dependency of covariates is the most practical approach in these studies. Moreover, the required statistical tests for the investigation of observed and unobserved heterogeneity among the failure data are not discussed. It seems that simplistic assumptions, the lack of a systematic approach and inadequate theoretical background are the main barriers to the proper application of the frailty model in the reliability field. To overcome these challenges, the main contribution of this paper is to present a framework for failure data analysis in the presence of unobserved and observed covariates. The framework is based on the mixed proportional hazards model, originally developed by Lancaster [33] to determine the causes of variation among unemployed persons in the length of time they are out of work. Recently, researchers such as Ghomghale et al., Barabady et al., Zaki et al., and Bjarte et al. have used the mixed proportional hazards in the field of remaining useful life, spare parts estimation, recoverability, and reliability [34]–[40].

3. Reliability Analysis of wheel loaders

Jajarm Bauxite Mine, located in Iran, has 19 main open mines in the city of Jajarm. The longitudinal area of the mine from west to east (namely: Golbini 1-8, Zou 1-4, Tagouei 1-6, and Sangtarash) accounts for 16 kilometers. The length of each section is as follows:

Golbini: A total of 4.7 km, Zou mines: A total of 3.3 km, Tagouei mines: A total of 5 km, and Sangtarash mine: About 3 kilometers in length. The Jajarm bauxite falls in the lens-like layer category. The expanse of bauxite is mostly in the form of layers. The minerals lying on the karstic-dolomites form the Elika formation, which lies under the shales and sandstones of the Shemshak formation. The bauxite layer is not made of uniform thickness and consistent quality. The bauxite layer generally ranges from less than 1 meter to about 40 meters in thickness. The case study was focused on the failure data of two-wheel loaders ($m = 2$) from Kaj-Mahya Company, which were put into service in the Jajarm Bauxite Mine. The wheel loaders' major design characteristics (weight, size, maximum load capacity, etc.) were almost the same.

The systematic framework for reliability analysis of reliability data in the presence of observed and unobserved covariates (observed and unobserved heterogeneity) is described in Figure 1. This methodology is based on four important steps:

- Establishing the context and data collection
- Identifying the baseline hazard rate based on maintenance nature
- Modeling the effect of the covariates
- Parameter estimation

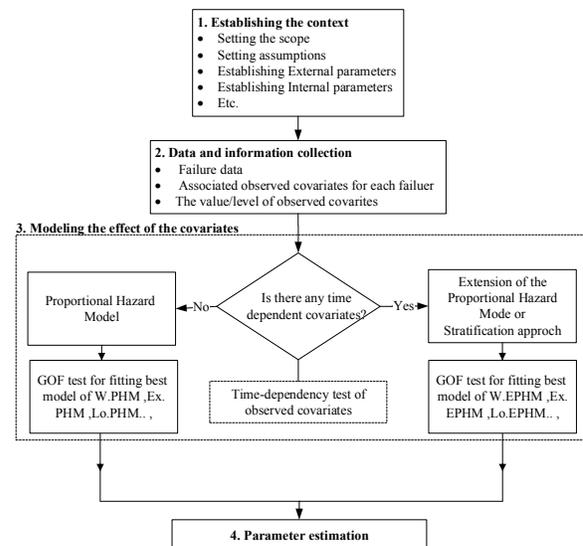


Figure 1. A framework for reliability model selection for Civil and mining Wheel loaders [34], [35], [39], [41]

As this figure shows, the context should be established in the first step. In this step, all external and internal parameters to be considered when analyzing failure data and setting the scope and assumptions for the reliability analysis should be defined. External context is the external environment in which the item is going to work, such as ambient temperature, pressure,

humidity, etc. Internal context is the internal conditions related to the item itself and the company running and maintaining the item, including the repair and physics of failure, operator condition, maintenance crew, etc. Understanding the external and internal context is important to identify the observed covariates. For example, based on the physics of failure, road conditions can contribute to the failure of a truck in a mine; hence, it should be considered as a covariate in the reliability analysis of the truck. In this step, the possible relationships between different covariates should be investigated, as well as the possible level for each of them.

In the next step, failure data and all possible observed covariates associated with each failure should be collected. After that, based on the nature of the failure data (e.g., trend behavior of the data) and the type of repair strategy, the appropriate baseline hazard should be selected for the data. For example, the common assumption for a repairable system can be i) perfect repair or good-as-new condition, ii) minimal repair or bad-as-old condition, or iii) jumps in the hazard rate after repair or different baseline hazard rate. Under the perfect repair strategy, the item is restored to a 'good-as-new' condition, and the main assumption is that the hazard rate is reset to that of a new system after maintenance. If the times between failures are independent and identically distributed (iid), it can be concluded that the item went through perfect repair [23]. In such cases, classical distribution, such as the Weibull distribution, can be used to model the baseline hazard rate.

In the case of minimal repair (bad-as-old), an item has the same intensity function after repair as before the failure. The failure times when the minimal repair is carried out can be considered a non-homogeneous Poisson process. In other words, the baseline hazard rate will be modeled using a non-homogeneous Poisson model. However, it should be mentioned that, on some occasions, such as overhaul, the system may return to a 'good-as-new' condition. Under this condition, it is assumed that the NHPP is cyclic, with each cycle starting as a renewal process and, within the cycle, failure times follow the NHPP. In this case, the failure data will then be categorized by these occasions (for example, overhaul). Then a stratification approach is used to estimate the effect of each covariate, while the NHPP model models the baseline hazard rate. However, the baseline hazard rate will change when a fleet of items is analyzed after some time and undergoing several repairs. For example, in some cases, as the number of failures increases, the average failure time decreases; hence, the baseline hazard rate will not be identical for a particular failure number. Here, the failure data can be categorized based on the failure number; it can be used to define strata, and then the stratification approach can be used to model the fleet failure data.

In general, the first step in analyzing the collected failure data of a repairable system is to check the trend of the failure data. If the data shows a trend, the NHPP, such as the power law process model or trend renewal process (TRP), can be used to model the baseline hazard rate. However, when there is no trend in the data, classical distribution, such as the Weibull distribution, can be used to model the baseline hazard rate. However, some goodness-of-fit tests, such as residual tests, should be used to find the best fit distribution for failure data. For more information regarding the trend test, see [9], [42].

In the next step, the time dependency of observed covariates should be checked. Later, the failure data need to be investigated for unobserved covariates. Data sets without unobserved heterogeneity will be analyzed using the classical proportional hazards model, including the proportional hazards model (when all observed covariates are time-independent) and the extension of the proportional hazards model (in the presence of time-dependent covariates). Moreover, data sets with unobserved heterogeneity will be analyzed using the mixed proportional hazards model (MPHM) family.

We had to collect the failure data along with related observed risk factors in the first step. To do so, we needed to identify the observed risk factors, as seen in Table 1. According to the Table, we identified 11 observed risk factors with a possible effect on the reliability of the wheel loaders. The training processes were different in these companies, leading to operators' different levels of skill; as a result, wheel loaders would experience different levels of stress. Altogether, these can cause various failure rates for wheel loaders identified in these companies. The numbers in square brackets in Table 4. were used to nominate (formulate) the risk factors. For example, wheel loaders work in three shifts, namely, Morning, Afternoon, and Night shifts; thus, we used 0, 1, and 2 here to refer to these shifts, respectively. A sample of data is shown in Table 1.

4. Heterogeneity test for unobserved covariates

Several statistical tests are available in the literature for identifying and quantifying the effects of unobserved heterogeneity. For example, Kimber [43] developed a Weibull-based score test for heterogeneity and then demonstrated its application in two case studies on infant nutrition. Under the assumption that the data follow a stratified proportional hazards model, where the hazard rate can be different within different strata, Gray [44] used the martingale residuals to test for variation over groups in survival data. Commenges and Andersen [45] used marginal, partial likelihood to develop a score test of homogeneity for survival data when the frailty model is used to model the covariates. The score test is valid for general distributions of the frailty variable, not only for the frequently used gamma distribution. In the

meta-analysis, Cochran's Q test (Q test) is normally used to check the homogeneity among data sets. However, the Q test only checks the presence versus the absence of heterogeneity; it does not report on the extent of such heterogeneity. The Q test is computed by summing the squared deviations of each study's effect, estimated from the overall effect estimate, weighting the contribution of each study by its inverse variance.

Table 1. The identified observed risk factors for the wheel loaders

| Risk factor | Level | Risk factor | level |
|---------------------------------|---------------------|--------------------------------------------------|---------------------|
| Working Place (z_{wp}) | Working face [1] | No. of service (z_{ns}) | Good [1] |
| | Working dump [2] | | Medium [2] |
| | | | Bad [3] |
| Working Shift (z_{wf}) | Morning shift [0] | Proportionality of Truck (z_{pt}) | Suitable [1] |
| | Afternoon shift [1] | | Partly suitable [2] |
| | Night shift [2] | | Unsuitable [3] |
| Road condition (z_{rc}) | Normal [1] | | |
| | Abnormal [2] | | |
| Rock fragmentation (z_{rf}) | Hard bauxite [1] | wheel loaders Code (Dummy Variable (z_{id})) | DT.1-DT2 |
| | Kaolin bauxite [2] | Rock Kind (z_{rk}) | bauxite [1] |
| | Chile boutique [3] | | Tailings [2] |
| | Tailings [4] | | Dolomite [3] |
| | Dolomite [5] | Precipitation (z_p) | Continuous |
| Temp. (z_t) | | | Continuous |
| Whether Condition (z_{wc}) | Sunny [1] | | |
| | Cloudy [2] | | |
| | Overcast [3] | | |
| | Dense fog [4] | | |

Under the hypothesis of homogeneity among the effect sizes, the Q statistic follows a chi-square distribution with $k - 1$ degree of freedom, k being the number of studies. However, these statistical tests and their applications are limited, mainly due to their requirements, in terms of data and assumptions. Each test is optimum to detect the heterogeneity of a specific form [46], [47]. For example, a shortcoming of the Q statistic is that it has poor power to detect true heterogeneity among studies when the meta-analysis includes a small number of studies and excessive power to detect negligible variability with a high number of studies. Recently, the I^2 index has been proposed to quantify the degree of heterogeneity in a meta-analysis [39]. A likelihood ratio test, the Akaike information criterion (AIC), and Bayesian information criterion (BIC) are common tests for checking the hypothesis of the presence of heterogeneity against the null hypothesis of non-heterogeneity ($\theta = 0$). In general, the AIC performs well when heterogeneity is small, but if heterogeneity is large, the BIC will often perform better

[9], [12], [49]. For example, in the case of Weibull distribution for the baseline hazard rate, the likelihood ratio can be written as:

$$R = 2 \left(\ln L(\hat{\lambda}, \hat{\beta}, \hat{\eta}, \hat{\theta}) - \ln L(\hat{\lambda}_0, \hat{\beta}_0, \hat{\eta}_0, 0) \right) \quad (1)$$

Here, $\hat{\lambda}$ and $\hat{\beta}$ are estimated parameters for Weibull distribution, $\hat{\eta}$ is the regression coefficient for observed covariates, and $\hat{\theta}$ can be interpreted as the degree of heterogeneity [9]. These parameters can be estimated by maximizing the full likelihood function. On a 5% significance level, the null hypothesis (no heterogeneity) will be rejected if $R \geq 2.706$. Moreover, under the minimal repair strategy, a power law can be used to represent the intensity function. Under the assumption of the power law intensity function, a three-step likelihood ratio test procedure can be performed to check whether a significant amount of heterogeneity among units exists [9]. As the first step, the null hypothesis, say $H_0: \lambda_1 = \lambda_2, \lambda_m = \lambda_0, \beta_1 = \beta_2, \beta_m = \beta_0$, should be tested against the alternative hypothesis, $H_1: \lambda_1 \neq \lambda_2, \lambda_m \neq \lambda_0, \beta_1 \neq \beta_2, \beta_m \neq \beta_0$. In the second and third steps, common λ , uncommon β and uncommon λ , common β should be carried out, respectively [11].

We employed the ratio test to assess the heterogeneity of data. Based on our assumption, the baseline hazard rate was represented by the Weibull distribution. Therefore, using the Weibull distribution, we wrote the failure rate of the wheel loaders as the baseline hazard, and the frailty model was represented by the gamma distribution with the following features (mean value = 1 & variance = θ) as follows:

$$h(t; z; \alpha) = \frac{\alpha^{\frac{1}{\theta}} t^{\frac{\alpha}{\theta} - 1}}{\Gamma(\frac{1}{\theta}) \theta^{\frac{1}{\theta}}} (\lambda \beta^\lambda t^{\lambda-1}) \exp[\sum_{i=1}^{p_1} \eta_i z_i] \quad (1)$$

Then, we did the likelihood ratio tests as below:

$$R = 2 \left(\left(\ln L(\hat{\lambda}, \hat{\beta}, \hat{\eta}, \hat{\theta}) - \ln L(\hat{\lambda}_0, \hat{\beta}_0, \hat{\eta}_0, 0) \right) \right) = 2(-247.467 - (-264.144)) = 33.27 \quad (2)$$

The P-Value for the $R=33.27$ was obtained as 0.000, suggesting the effect of unobserved risk factors (unobserved heterogeneity) on the reliability of wheel loaders' Time dependency test of observed covariates. Hence, as in Figure 1, the Mix Proportional Hazard Model (MPHM) or Extension Mix Proportional Hazard Model (EMPHM) should be used to analyze the data.

5. Time dependency test of observed covariates

There are two general approaches for checking the time dependency of covariates: *i*) the graphical procedure and *ii*) the goodness-of-fit testing procedure [17]. The developed graphical procedure can generally be categorized into three main groups: *i*) cumulative hazards plots, *ii*) average hazards plots, and *iii*) residual plots [13]. For example, in cumulative hazards plots, the data will be categorized based on the different levels of the covariate that is to be checked for time dependency. Consider that a covariate can be categorized into r

levels, in which the covariate is equal to z_r . After that, the hazard rate can be written as:

$$h_r(t; z; \alpha) = \frac{\alpha^{\frac{1}{\theta}-1} e^{-\frac{\alpha}{\theta}}}{\Gamma(\frac{1}{\theta})\theta^{\frac{1}{\theta}}} \lambda_{0r}(t) \exp\left[\sum_{i=1}^{p_1} \eta_i z_i\right] \quad (3)$$

Where $\eta_i z_i$ is the same as before, with $\eta_r z_r$ omitted, with $i=1,2,\dots,p_1$ and $j \neq r$. If the PH assumption is justified, then we will end up with:

$$h_{0r}(t) = C_r h_0(t), \text{ and } C_r = \frac{\alpha^{\frac{1}{\theta}-1} e^{-\frac{\alpha}{\theta}}}{\Gamma(\frac{1}{\theta})\theta^{\frac{1}{\theta}}} \exp(\eta_r z_r) \quad (4)$$

A similar relation can be concluded for the cumulative baseline hazard rate. Hence, if the assumption of PH is justified, then the plots of the logarithm of the estimated cumulative baseline hazard rates against time for defined categories should simply be shifted by an additive constant, η_r . In other words, they should be approximately parallel and separated, corresponding to the different values of the covariates. Departure from parallelism of the above plots for different categories may suggest that z_r is a time-dependent covariate. For a review of other graphical approaches, see [13], [17], [50]–[52].

In the same way as the cumulative baseline hazard rate, a log-log Kaplan-Meier curve over different (combinations of) categories of variables can be used to check the assumption of PH. A log-log reliability curve is simply a transformation of an estimated reliability curve that results from taking the natural log of an estimated survival probability twice. If we use a PHM or MPHM and plot the estimated log-log reliability curves for defined categories on the same graph, the two plots would be approximately parallel [13]. In the residuals plot in the first step, the residual should be estimated by using the estimated values of the cumulative hazard rate, $H_0(t_i)$, and the regression vector η as:

$$e_i = -H_0(t_i) \exp(\eta_r z_r) \quad (5)$$

If the PH assumption is justified, then the logarithm of the estimated reliability function of e_i against the residuals should lie approximately on a straight line with slope -1 [13], [53]. A transformed plot of the partial residual suggested by Schoenfeld can also be used as an exploratory tool to detect the time-varying effects of a covariate, even when the a priori form of time dependence is unknown [54]–[56]. The Schoenfeld Residuals Test is analogous to testing whether the slope of the scaled residuals on time is zero or not. If the slope is not zero, then the proportional hazard assumption has been violated [47]. When the covariates are quantitative, using graphical approaches is challenging, as it is difficult to define different levels for quantitative covariates and to decide whether the plots are parallel. In such cases, it is better to use a goodness-of-fit testing procedure such as the chi-squared goodness-of-fit test [5], [57], [58], the log-rank test [5], [57], the likelihood ratio test [5], [57], score tests [57], [59], the doubly cumulative hazard function [60], the Wilcoxon test [61]

and generalized moments specification tests [62]. For example, if the PH assumption is justified, the different two-sample tests, e.g., generalized Wilcoxon and log-rank tests, should have the same results [13].

As seen in Table 2 we had to examine the time dependency of the risk factors following the collection of data and observed risk factors. We applied the graphical approach (A ln–ln reliability curve) to evaluate the time dependency of all risk factors.

Table 2. A sample of failure data and their related observed risk factors

| TB F | z_{id} | z_{wp} | z_{wf} | z_{ns} | z_{pt} | z_{rf} | z_{wc} | z_p | z_t | z_{rc} | z_{rk} |
|------|----------|----------|----------|----------|----------|----------|----------|-------|-------|----------|----------|
| 10 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 0 | 5 | 1 | 2 |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 5 | 1 | 1 |
| 57 | 2 | 1 | 3 | 3 | 2 | 4 | 2 | 0 | 1 | 1 | 3 |
| 6/5 | 2 | 2 | 3 | 3 | 1 | 5 | 1 | 0 | 7 | 1 | 2 |

The $-\ln(-\ln \text{reliability})$ against the $\ln(\text{analysis time})$ for two observed risk factors is shown in

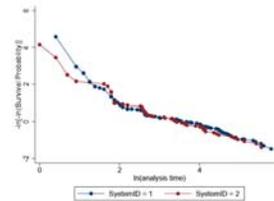


Figure 2, which were named Rock kind (z_{rk}) and System ID (z_{id}). Based on these two graphs, the curves are approximately parallel in both; thus, the proportionality assumption applies to the data sets; therefore, the risk factors are concluded to be time-independent. The $-\ln(-\ln \text{reliability})$ for other risk factors confirmed the same result. Hence, as in Figure 1, the MPHM should be used to analyze the data. Software such as STATA, R, SAS, and SYSTAT can be estimated the MPHM parameters.

6. Parameter estimation

In the MPHM, the hazard rate of an item is the product of a baseline hazard rate multiplied by two positive functions: *i*) observed covariate function and *ii*) an unobserved covariate function (frailty function). Suppose we have a fleet of j items, the hazard function for an item at time $t > 0$ is:

$$h_j(t; z; \alpha) = \alpha_j h_0(t) \psi(z; \eta) \quad (6)$$

Where $h_0(t)$ is an arbitrary baseline hazard rate, dependent on time alone, z is a row vector consisting of the observed covariates associated with the item, η is a column vector consisting of the regression parameters for identified observed covariates, and α_j is a time-independent frailty function for item j and represents the cumulative effect of one or more unobserved covariates.

In general, the baseline hazard rate ($h_0(t)$) may either be left unspecified or can be modeled using a specific parametric form such as Weibull distribution or NHPP.

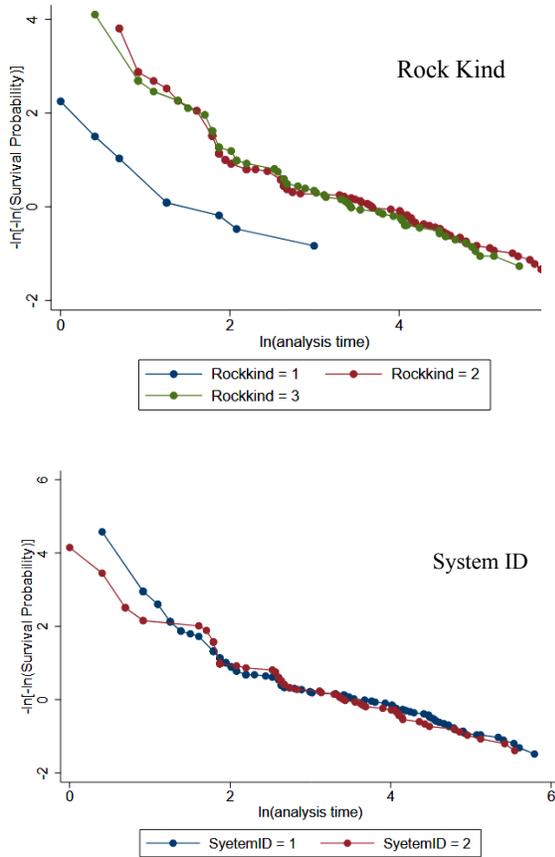


Figure 2. The log-log graph of the time between wheel loaders based on road conditions and the company

According to the MPHMM, the fleet of items (the population) is represented as a mixture in which the $h_0(t)$ and $\psi(z; \eta)$ are common to all items, although each has its frailty. The observed and unobserved covariates can affect the hazard rate so that the actual hazard rate ($h_j(t; z; \alpha)$) is either greater (e.g., in the case of higher vibration level or poor maintenance) or smaller (e.g., better training for operators, installation of a new ventilation system) than the baseline hazard rate. Moreover, the items with $\alpha_j > 1$ are said to be frailer for reasons left unexplained by the observed covariates and will have an increased risk of failure. The items for whom $\alpha_j < 1$ are less frail; hence, given a certain observed covariate pattern, they tend to be more reliable.

For MPHMM, given the relationship between the hazard rate and the reliability functions, it can be shown that the conditional (item) reliability function, $R(t; z; z(t)|\alpha)$, conditional on the frailty, α , is [14]:

$$R(t; z; z(t)|\alpha) = \{R(t; z; z(t))\}^\alpha \tag{7}$$

The unconditional (population) reliability function can then be estimated by integrating out the unobserved α . If α has probability density function $g(\alpha)$, then the population or unconditional reliability function is given by:

$$R_\theta(t; z; z(t)) = \int_0^\infty \{R(t; z; z(t))\}^\alpha g(\alpha) d\alpha \tag{8}$$

Where we use the subscript θ to emphasize the dependence on the frailty variance θ , the relationship between the reliability function and the hazard function still holds unconditional on α , and, thus, we can obtain the population hazard function using [14]:

$$\lambda_\theta(t; z; z(t)) = -\frac{d}{dt} R_\theta(t; z; z(t)) [R_\theta(t; z; z(t))]^{-1} \tag{9}$$

Having the gamma distribution with unobserved covariates [14]:

$$R_\theta(t; z; z(t)) = [1 - \theta \ln\{R(t; z; z(t))\}]^{-1/\theta} \tag{10}$$

Having the event times (t_{0i}, t_i, d_i), for $i = 1, \dots, n$ with the i th observation corresponding to the time (t_{0i}, t_i], with either failure occurring at the time $t_i (d_i = 1)$ or the failure time being right-censored at the time $t_i (d_i = 0)$, the likelihood function for survival data is given by:

$$LnL = \ln \prod_{i=1}^n \frac{\{R_{\theta_i}(t_i, z_i, z_i(t))\}^{1-d_i} f_{\theta_i}(t_i, z_i, z_i(t))^{d_i}}{R_{\theta_i}(t_i, z_i, z_i(t))} \tag{11}$$

Where f_{θ_i} is the probability density function. In a shared frailty model, suppose we have data for $i = 1, \dots, n$ groups, with $j = 1, \dots, n_i$ observations per group, consisting of the trivariate response (t_{0ij}, t_{ij}, d_{ij}), which indicates the start time, end time, and failure/censoring for the j th item from the i th group, while the shared frailties follow a gamma distribution, L_i can be expressed compactly as [14]:

$$L_i = \left[\prod_{j=1}^{n_i} \{\lambda_{ij}(t_{ij})\}^{d_{ij}} \right] \frac{\Gamma(1/\theta + D_i)}{\Gamma(1/\theta)} \theta^{D_i} \left\{ 1 - \theta \sum_{j=1}^{n_i} \ln \frac{R_{ij}(t_{ij})}{R_{ij}(t_{0ij})} \right\}^{-1/\theta + D_i} \tag{12}$$

Where $D_i = \sum_{j=1}^{n_i} d_{ij}$. Given the unconditional group likelihoods, we can estimate the regression parameters and frailty variance θ , by maximizing the overall log-likelihood $LnL = \sum_{i=1}^n \ln L_i$. In shared-frailty Cox models, the estimation consists of two steps. In the first step, the optimization is in terms of θ alone. For fixed θ , the second step consists of fitting a standard Cox model via penalized log-likelihood, with the v_i introduced as estimable coefficients of dummy variables identifying the groups. The same approach can be used to estimate the likelihood functions for EPHM, MPHMM, and PHM. For more information, see [9], [13], [14].

Table 3 and Table 4 represent the analysis results obtained by STATA software. Based on the analysis results, the Working Place (\mathbf{z}_{wp}), Temperature (\mathbf{z}_t), Weather Condition (\mathbf{z}_{wc}), No. of service (\mathbf{z}_{ns}), Rock fragmentation (\mathbf{z}_{rf}), Precipitation (\mathbf{z}_p), and Rock kind (\mathbf{z}_{rk}) showed significant impacts on the reliability of

wheel loaders. Applying the regression coefficient to the risk factors, we obtained the unconditional survival of wheel loaders as follows:

$$R_0(t) = \left[1 - 212.37 \ln \left(R_0(t) \exp \left(\begin{matrix} 0.27z_p - 0.320z_{wp} \\ + 0.204z_{wc} + 0.312z_t \\ + 0.566z_{ns} - 2.61z_{pl} \\ - 0.114z_{rj} + 0.44z_{rk} \end{matrix} \right) \right) \right]^{-1/212.37} \quad (14)$$

Table 3. The analysis result by the assumption of MPHM

| | | | | |
|-----------------------------|-------|------|-------|-------|
| Temperature | -.046 | .018 | 2.49 | 0.013 |
| Road condition | -.453 | .293 | 1.54 | 0.124 |
| No of Service | .866 | .290 | -2.94 | 0.003 |
| Rock Fragmentation Blasting | -.338 | .145 | 2.30 | 0.021 |
| Rock kind | .266 | .232 | -1.14 | 0.253 |
| System ID | -.192 | .197 | 0.98 | 0.329 |

| Risk factors | Coef. | Std. Err. | z | P-value |
|-----------------------------|-------|-----------|-------|---------|
| Shift | .120 | .045 | -2.50 | 0.012 |
| Working Place | -.320 | .086 | 2.98 | 0.003 |
| Proportionality of truck | .044 | .055 | -0.84 | 0.403 |
| Weather Condition | .205 | .034 | -3.91 | 0.000 |
| Precipitation mm | .040 | .006 | -4.03 | 0.000 |
| Temperature o C | .031 | .004 | -4.38 | 0.000 |
| Road condition | .156 | .077 | -1.98 | 0.048 |
| No Service | .566 | .064 | -4.58 | 0.000 |
| Rock Fragmentation Blasting | -.114 | .033 | 2.99 | 0.003 |
| Rock kind | .439 | .051 | -4.49 | 0.000 |
| System ID | -.022 | .070 | 0.35 | 0.725 |

Table 4. The baseline and unobserved parameters

| The variance of Gamma distribution (θ) | Weibull (p) |
|----------------------------------------|-------------|
| 212.37 | 121.24 |

Where $R_0(t) = e^{-t^{121.24}}$ is the baseline survival of the wheel loaders, unconditional and conditional hazard functions of wheel loaders are shown in Figure 3. **Figure 3: a)** The population hazard function of wheel loader based on the risk factor mean, **b)** The population reliability function of a wheel loader on the risk factor mean

In the next step, we did the analysis assuming no effect for unobserved risk factors to examine the possible bias of analysis in the case of ignoring the effect of unobserved risk factors. The analysis results by selecting the PHM as the model are given in Table 5. The figures show a big difference between the wheel loaders' population hazard rate and the individual wheel loaders. Based on the findings, the regression coefficients of observed risk factors with a significant impact on the hazard rate are generally higher in the PHM compared to the MPHM, and $\beta = .90$.

Table 5: The parameter estimation for the failure rate of Turks in PHM

| Risk factors | Coef. | Std. Err. | z | p-value |
|--------------------------|-------|-----------|-------|---------|
| Shift | 1.402 | .250 | -5.19 | 0.000 |
| Working Place | .885 | .307 | -2.80 | 0.005 |
| Proportionality of truck | -.785 | .265 | 2.90 | 0.004 |
| Weather Condition | -.008 | .144 | 0.05 | 0.957 |
| Precipitation | .058 | .047 | -1.25 | 0.210 |

Unconditional and conditional hazard functions of wheel loaders are shown in Figure 4. The comparison results of the hazard rates of wheel loaders in both models are shown in Figure 5. As seen, unobserved risk factors significantly impacted the hazard rate of wheel loaders; thus, failure to consider this factor can mislead further decisions on operation and maintenance strategies.

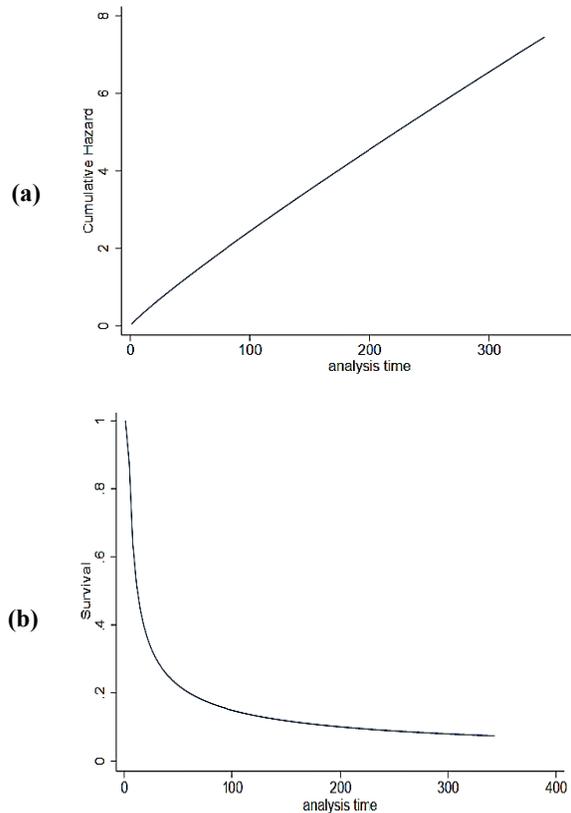


Figure 4. a) The population hazard function of wheel loaders based on the risk factor mean, **b)** The population reliability function of wheel loaders based on the risk factor mean

7. Conclusion

In reality, heterogeneity among failure data may occur in many cases. Heterogeneity can be due to observed or unobserved covariates, resulting in observed or

unobserved heterogeneity. The reliability analysis results for heterogeneous data can differ substantially from those in a homogeneous case. In most cases, failing to account for heterogeneity would lead to significant differences in estimating the effects of covariates. As a recommendation, all data sets should be checked for unobserved heterogeneity using an appropriate statistical test. The literature review results revealed the lack of a systematic approach to model unobserved covariates in the area of reliability analysis. In this study, the required statistical tests and available models for observed or unobserved heterogeneity in the reliability analysis of failure data are reviewed. Then a systematic framework is developed to facilitate the application of these models. The framework is based on the MPHMM, and its extension, which provides an appropriate tool for modeling observed and unobserved heterogeneity under the different types of maintenance strategies. In the analysis of the data sets with observed and unobserved heterogeneity, the time dependency test of the observed covariates needs to be performed in the first step. After that, the presence of unobserved covariates should be checked using an appropriate statistical test. Finally, considering the type of repair strategy carried out on the item, the most appropriate model among the MPHMM family should be selected.

In the second part of the paper, the application of the developed framework is illustrated by investigating the amount of observed and unobserved heterogeneity in the failure data of 2 wheel loaders from one company put into service in a Bauxite mine. The large variability in failure data and the differences in failure intensity of the wheel loaders indicate heterogeneity among the collected data, which observed and unobserved covariates can explain. The graphical approach is used to check the trend and correlation of failure data. The result showed no trend and correlation among the data, which can justify the *iid* assumption. Hence, the renewal process can represent the baseline hazard of wheel loaders. The result of time-dependency and heterogeneity tests (ratio test) indicated that all identified observed covariates are time-independent and that there is an unobserved heterogeneity among the failure data. This means that some other factors, which were not included in this study, might affect the reliability of the wheel loaders. Therefore, we need to explore further and model the effect of the unobserved factors, to enhance the accuracy of the estimation.

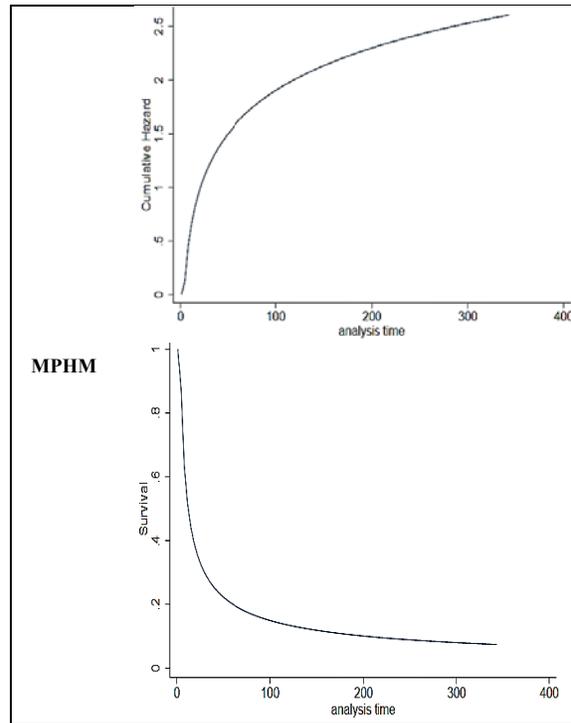
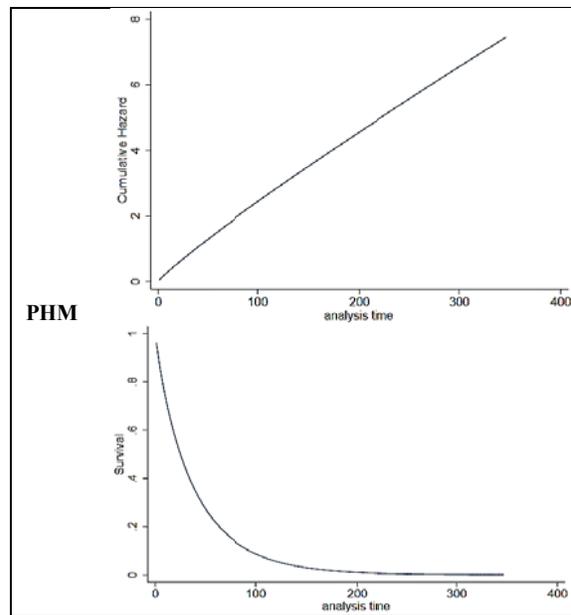


Figure 5. The comparison of hazard functions using MPHMM and PH



(Continue) Figure 6. The comparison of hazard functions using MPHMM and PH

Having these results and the developed framework (Figure 1), the MPHMM should be used to analyze the data. The analysis showed that six of the identified observed covariates significantly affect the wheel loaders' hazard rate. Ignoring the effect of unobserved covariates, and using PHM instead of MPHMM, will

underestimate the effect of Company, Working shift, Rock fragmentation, and slope of the road and overestimate the effect of Proportionality of wheel loaders and Number of services. Moreover, under the assumption of PHM, the Rock condition significantly affects the hazard rate of wheel loaders. In contrast, its effect is insignificant when MPHMM is used to model the failure data. Hence, for any decisions on the operation and maintenance strategy, the effect of unobserved covariates should be considered. Finally, although we have modeled the effect of unobserved covariates on the reliability of wheel loaders, to enhance the accuracy of the estimation model in future work, we need to explore the unobserved covariates further.

8. References

- [1] Y. Kikuchi, M. Tochizawa, S. Takeda, S. Kamada, and K. Hirose, "Automatic excavator," 1991, pp. 277–284.
- [2] E. Topal and S. Ramazan, "A new MIP model for mine equipment scheduling by minimizing maintenance cost," *Eur. J. Oper. Res.*, vol. 207, no. 2, pp. 1065–1071, 2010.
- [3] D. M. Loutit, R. Pascual, and A. K. Jardine, "A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data," *Reliab. Eng. Syst. Saf.*, vol. 94, no. 10, pp. 1618–1628, 2009.
- [4] X. Liu, "Survival models on unobserved heterogeneity and their applications in analyzing large-scale survey data," *J. Biom. Biostat.*, vol. 5, 2014.
- [5] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, 2011.
- [6] M. Finkelstein, "Imperfect repair and lifesaving in heterogeneous populations," *Reliab. Eng. Syst. Saf.*, vol. 92, no. 12, pp. 1671–1676, Dec. 2007, doi: 10.1016/j.res.2006.09.018.
- [7] S. G. Thompson and S. J. Sharp, "Explaining heterogeneity in meta-analysis: a comparison of methods," *Stat. Med.*, vol. 18, no. 20, pp. 2693–2708, 1999.
- [8] J. P. Higgins and S. G. Thompson, "Quantifying heterogeneity in a meta-analysis," *Stat. Med.*, vol. 21, no. 11, pp. 1539–1558, 2002.
- [9] A. Garmabaki, A. Ahmadi, J. Block, H. Pham, and U. Kumar, "A reliability decision framework for multiple repairable units," *Reliab. Eng. Syst. Saf.*, vol. 150, pp. 78–88, 2016, doi: 10.1016/j.res.2016.01.020.
- [10] M. Guida and M. Giorgio, "Reliability analysis of accelerated life-test data from a repairable system," *IEEE Trans. Reliab.*, vol. 44, no. 2, pp. 337–346, 1995.
- [11] M. Giorgio, M. Guida, and G. Pulcini, "Repairable system analysis in presence of covariates and random effects," *Reliab. Eng. Syst. Saf.*, vol. 131, pp. 271–281, Nov. 2014, doi: 10.1016/j.res.2014.04.009.
- [12] J. F. Lawless, "Regression methods for Poisson process data," *J. Am. Stat. Assoc.*, vol. 82, no. 399, pp. 808–815, 1987.
- [13] D. Kumar and B. Klefsjö, "Proportional hazards model: a review," *Reliab. Eng. Syst. Saf.*, vol. 44, no. 2, pp. 177–188, 1994.
- [14] R. G. Gutierrez, "Parametric frailty and shared frailty survival models," *Stata J.*, vol. 2, no. 1, pp. 22–44, 2002.
- [15] P. Hougaard, "Frailty models for survival data," *Lifetime Data Anal.*, vol. 1, no. 3, pp. 255–273, Sep. 1995, doi: 10.1007/BF00985760.
- [16] J. W. Vaupel, K. G. Manton, and E. Stallard, "The impact of heterogeneity in individual frailty on the dynamics of mortality," *Demography*, vol. 16, no. 3, pp. 439–454, 1979.
- [17] A. Barabadi, J. Barabady, and T. Markeset, "Maintainability analysis considering time-dependent and time-independent covariates," *Reliab. Eng. Syst. Saf.*, vol. 96, no. 1, pp. 210–217, 2011.
- [18] J. H. Abbring and G. J. Van Den Berg, "The unobserved heterogeneity distribution in duration analysis," *Biometrika*, vol. 94, no. 1, pp. 87–99, 2007.
- [19] P. T. P. Leszczyc and F. M. Bass, "Determining the effects of observed and unobserved heterogeneity on consumer brand choice," *Appl. Stoch. Models Data Anal.*, vol. 14, no. 2, pp. 95–115, 1998.
- [20] A. Barabadi, O. T. Gudmestad, and J. Barabady, "RAMS data collection under Arctic conditions," *Reliab. Eng. Syst. Saf.*, vol. 135, pp. 92–99, 2015, doi: 10.1016/j.res.2014.11.008.
- [21] X. Gao, J. Barabady, and T. Markeset, "An approach for prediction of petroleum production facility performance considering Arctic influence factors," *Reliab. Eng. Syst. Saf.*, vol. 95, no. 8, pp. 837–846, 2010.
- [22] A. H. S. Garmabaki, A. Ahmadi, Y. A. Mahmood, and A. Barabadi, "Reliability Modelling of Multiple Repairable Units," *Qual. Reliab. Eng. Int.*, vol. 32, no. 7, pp. 2329–2343, Nov. 2016, doi: 10.1002/qre.1938.
- [23] Z. G. Asfaw and B. H. Lindqvist, "Unobserved heterogeneity in the power law nonhomogeneous Poisson process," *Reliab. Eng. Syst. Saf.*, vol. 134, pp. 59–65, 2015.
- [24] [24] M. Li and J. Liu, "Bayesian hazard modeling based on lifetime data with latent heterogeneity," *Reliab. Eng. Syst. Saf.*, vol. 145, pp. 183–189, 2016.
- [25] D. G. Clayton, "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence," *Biometrika*, vol. 65, no. 1, pp. 141–151, 1978.
- [26] S. Asadzadeh, A. Aghaie, H. Shahriari, and S. T. A. Niaki, "The application of proportional hazards and frailty models to multistage processes surveillance," *Int. J. Adv. Manuf. Technol.*, vol. 74, no. 1, pp. 461–470, Sep. 2014, doi: 10.1007/s00170-014-5914-4.
- [27] V. Slimacek and B. H. Lindqvist, "Nonhomogeneous Poisson process with nonparametric frailty," *Reliab. Eng. Syst. Saf.*, vol. 149, pp. 14–23, May 2016, doi: 10.1016/j.res.2015.12.005.
- [28] G. Asha, A. V. Raja, and N. Ravishanker, "Reliability modelling incorporating load share and frailty," *Appl. Stoch. Models Bus. Ind.*, vol. 34, no. 2, pp. 206–223, 2018, doi: 10.1002/asmb.2294.
- [29] M. Xu and X. Li, "Negative dependence in frailty models," *J. Stat. Plan. Inference*, vol. 138, no. 5, pp. 1433–1441, 2008.
- [30] N. Misra, N. Gupta, and R. D. Gupta, "Stochastic comparisons of multivariate frailty models," *J. Stat. Plan. Inference*, vol. 139, no. 6, pp. 2084–2090, Jun. 2009, doi: 10.1016/j.jspi.2008.09.006.
- [31] M. Finkelstein, "Shocks in homogeneous and heterogeneous populations," *Reliab. Eng. Syst. Saf.*, vol. 92, no. 5, pp. 569–574, 2007.
- [32] V. Slimacek and B. H. Lindqvist, "Reliability of wind turbines modeled by a Poisson process with covariates, unobserved heterogeneity and seasonality," *Wind Energy*,

- vol. 19, no. 11, pp. 1991–2002, 2016, doi: 10.1002/we.1964.
- [33] T. Lancaster, “Econometric methods for the duration of unemployment,” *Econom. J. Econom. Soc.*, pp. 939–956, 1979.
- [34] A. Ghomghaleh et al., “Prediction of remaining useful life (RUL) of Komatsu excavator under reliability analysis in the Weibull-frailty model,” *Plos One*, vol. 15, no. 7, p. e0236128, 2020.
- [35] A. Ghomghale, M. Ataei, R. Khalokakaie, A. Barabadi, and A. Nouri Qarahasanlou, “The Application of Frailty Model in Remaining Useful Life Estimation (Case Study: Sungun Copper Mine’s Loading System),” *J. Model. Eng.*, vol. 18, no. 62, Oct. 2020, doi: 10.22075/jme.2020.19249.1817.
- [36] R. Barabadi, M. Ataei, R. Khalokakaie, and A. Nouri Qarahasanlou, “Observed and Un-Observed Covariate Effects on Baseline Hazard Rate - Case study: Jajarm Bauxite Mine,” *J. Model. Eng.*, vol. 0, Nov. 2019, doi: 10.22075/jme.2019.17837.1721.
- [37] R. Barabadi, M. Ataei, R. Khalokakaie, A. Barabadi, and A. N. Qarahasanlou, “Spare Part Management Considering Risk Factors,” in *International Congress and Workshop on Industrial AI*, 2021, pp. 24–39.
- [38] R. Barabadi, M. Ataei, R. Khalokakaie, and A. Nouri Qarahasanlou, “Spare-part management in a heterogeneous environment,” *Plos One*, vol. 16, no. 3, p. e0247650, 2021.
- [39] R. Zaki, A. Barabadi, J. Barabadi, and A. Nouri Qarahasanlou, “Observed and unobserved heterogeneity in failure data analysis,” *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.*, p. 1748006X211022538, 2021.
- [40] B. Rod, A. Barabadi, and M. Naseri, “Recoverability Modeling of Power Distribution Systems Using Accelerated Life Models: Case of Power Cut due to Extreme Weather Events in Norway,” *J. Manag. Eng.*, vol. 36, no. 5, p. 05020012, 2020.
- [41] A. Nouri Qarahasanlou, M. Ataei, R. Khalokakaie, S. Fatoorachi, and R. Barabadi, “Operating Environment Based Reliability Analysis of Mining Equipment Case Study: Molybdenum-Copper Mine (Sungun Copper Mine),” *J. Anal. Numer. Methods Min. Eng.*, vol. 9, no. 18, pp. 129–141, Apr. 2019.
- [42] A. Barabadi, “Reliability model selection and validation using Weibull probability plot—A case study,” *Electr. Power Syst. Res.*, vol. 101, pp. 96–101, Aug. 2013, doi: 10.1016/j.epr.2013.03.010.
- [43] A. C. Kimber, “A Weibull-based score test for heterogeneity,” *Lifetime Data Anal.*, vol. 2, no. 1, pp. 63–71, Mar. 1996, doi: 10.1007/BF00128471.
- [44] R. J. Gray, “Tests for Variation over Groups in Survival Data,” *J. Am. Stat. Assoc.*, Feb. 2012, Accessed: Jan. 30, 2019. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476502>
- [45] D. Commenges and P. K. Andersen, “Score test of homogeneity for survival data,” *Lifetime Data Anal.*, vol. 1, no. 2, pp. 145–156, Jun. 1995, doi: 10.1007/BF00985764.
- [46] M. Mittlböck and H. Heinzl, “A simulation study comparing properties of heterogeneity measures in meta-analyses,” *Stat. Med.*, vol. 25, no. 24, pp. 4321–4333, 2006, doi: 10.1002/sim.2692.
- [47] J. P. A. Ioannidis, “Interpretation of tests of heterogeneity and bias in meta-analysis,” *J. Eval. Clin. Pract.*, vol. 14, no. 5, pp. 951–957, 2008, doi: 10.1111/j.1365-2753.2008.00986.x.
- [48] T. B. Huedo-Medina, J. Sánchez-Meca, F. Marin-Martínez, and J. Botella, “Assessing heterogeneity in meta-analysis: Q statistic or I² index?,” *Psychol. Methods*, vol. 11, no. 2, p. 193, 2006.
- [49] M. J. Brewer, A. Butler, and S. L. Cooksley, “The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity,” *Methods Ecol. Evol.*, vol. 7, no. 6, pp. 679–692, 2016.
- [50] S. W. Lagakos, “The graphical evaluation of explanatory variables in proportional hazard regression models,” *Biometrika*, vol. 68, no. 1, pp. 93–98, Apr. 1981, doi: 10.1093/biomet/68.1.93.
- [51] E. Arjas, “A graphical method for assessing goodness of fit in Cox’s proportional hazards model,” *J. Am. Stat. Assoc.*, vol. 83, no. 401, pp. 204–212, 1988.
- [52] R. J. Gray, “Some Diagnostic Methods for Cox Regression Models Through Hazard Smoothing,” *Biometrics*, vol. 46, no. 1, pp. 93–102, 1990, doi: 10.2307/2531633.
- [53] D. R. Cox, “A note on the graphical analysis of survival data,” *Biometrika*, vol. 66, no. 1, pp. 188–190, Apr. 1979, doi: 10.1093/biomet/66.1.188.
- [54] D. Schoenfeld, “Partial residuals for the proportional hazards regression model,” *Biometrika*, vol. 69, no. 1, pp. 239–241, Apr. 1982, doi: 10.1093/biomet/69.1.239.
- [55] A. N. Pettitt and I. B. Daud, “Investigating Time Dependence in Cox’s Proportional Hazards Model,” *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 39, no. 3, pp. 313–329, 1990, doi: 10.2307/2347382.
- [56] S. Park and D. J. Hendry, “Reassessing Schoenfeld Residual Tests of Proportional Hazards in Political Science Event History Analyses,” *Am. J. Polit. Sci.*, vol. 59, no. 4, pp. 1072–1087, 2015, doi: 10.1111/ajps.12176.
- [57] J. D. Kalbfleisch and R. L. Prentice, *The statistical analysis of failure time data*. Wiley-Interscience, 2011.
- [58] D. Schoenfeld, “Chi-squared goodness-of-fit tests for the proportional hazards regression model,” *Biometrika*, vol. 67, no. 1, pp. 145–153, Jan. 1980, doi: 10.1093/biomet/67.1.145.
- [59] J. O’Quigley and F. Pessione, “Score Tests for Homogeneity of Regression Effect in the Proportional Hazards Model,” *Biometrics*, vol. 45, no. 1, pp. 135–144, 1989, doi: 10.2307/2532040.
- [60] I. W. McKeague and K. J. Utikal, “Goodness-of-Fit Tests for Additive Hazards and Proportional Hazards Models,” *Scand. J. Stat.*, vol. 18, no. 3, pp. 177–195, 1991.
- [61] R. Gill and M. Schumacher, “A simple test of the proportional hazards assumption,” *Biometrika*, vol. 74, no. 2, pp. 289–300, Jun. 1987, doi: 10.1093/biomet/74.2.289.
- [62] G. R. Neumann, “A Generalized Moments Specification Test of the Proportional Hazards Model AU - Horowitz, Joel L.,” *J. Am. Stat. Assoc.*, vol. 87, no. 417, pp. 234–240, Mar. 1992, doi: 10.1080/01621459.1992.10475197.