Yayın Geliş Tarihi (Submitted): **21/10/2024**

Yayın Kabul Tarihi (Accepted): **24/12/2024**

Makele Türü (Paper Type): **Araştırma Makalesi – Research Paper**

Please Cite As/Atıf için:
Yavuz, H. S. and Alayoğlu, İ. (2024), Robust lifetime estimation with small sample sized high speed railway ETCS component data, *Nicel Bilimler Dergisi*, 6(2), 181-207. *doi: 10.51541/nicel.1571005*

ROBUST LIFETIME ESTIMATION WITH SMALL SAMPLE SIZED HIGH SPEED RAILWAY ETCS COMPONENT DATA

Hasan Serhan YAVUZ[1] and İsmail ALAYOĞLU[2]

ABSTRACT

Railway maintenance and efficient operation has been an important issue for a safe rail traffic. When an unexpected malfunction occurs on various components on the train or on the railway system, it may result in unscheduled maintenance which may cause the rail traffic to stop. In this paper, we study the random failure model of some frequently malfunctioning high-speed railway equipment based on the statistical analysis of the real data of failure records in Turkey. Popular distribution functions and parameter estimation methods have been used considering that the data has a small sample size, and it may contain outliers. In this study, we showed that for the case of a few numbers of failure data, the L-moments method gives effective results when there exists no outlier and the robust-M method gives effective results when there exists an outlier or outliers.

Keywords: Lifetime estimation, Materials reliability, Small sample, Predictive maintenance, ETCS signaling components.

[1] Corresponding Author, Assoc. Prof. Dr., Electrical and Electronics Engineering Department, Faculty of Engineering and Architecture, Eskisehir Osmangazi University, Eskişehir, Türkiye. ORCID ID: https://orcid.org/0000-0002-4944-1013

[2] PhD student, Electrical and Electronics Engineering Department, Faculty of Engineering and Architecture, Eskisehir Osmangazi University, Eskişehir, Türkiye. ORCID ID: https://orcid.org/0000-0002-7291-7868

AZ ÖRNEKLEMLİ YÜKSEK HIZLI DEMİRYOLU ETCS BİLEŞEN VERİLERİYLE GÜRBÜZ ÖMÜR SÜRESİ TAHMİNİ

ÖZ

Demiryolu sistemlerinin bakımı ve işletme verimliliği, güvenli bir demiryolu trafiği için önemlidir. Tren veya demiryolu sistemindeki çeşitli bileşenlerde beklenmeyen bir arıza meydana geldiğinde, demiryolu trafiğinin tamamen durmasına yol açabilecek bir zorunlu bakım gerektirebilir. Bu makalede, Türkiye'deki arıza kayıtları verilerinin istatistiksel analizine dayalı olarak sık aralıklarla arızalanan bazı yüksek hızlı demiryolu ekipmanlarının rassal arıza modeli incelenmiştir. Verilerin küçük örneklem hacmine sahip olması ve aykırı değerler içerebilme durumları da dikkate alınarak yaygın olarak kullanılan dağılım fonksiyonları ve parametre tahmin teknikleri uygulanmıştır. Bu çalışmada, az sayıda arıza verisi durumunda, aykırı değer olmadığında L-momentler yönteminin ve aykırı değer veya aykırı değerler olduğunda ise gürbüz-M yönteminin etkili sonuçlar verdiği gösterilmiştir.

Anahtar Kelimeler: Ömür süresi tahmini, Bileşen güvenilirliği, Küçük örneklem, Tahmini bakım, ETCS sinyalizasyon bileşenleri.

1. INTRODUCTION

Railway transportation is important in the process of industrialization as it helps to transport the people and goods at a faster and cheaper rates (Jia et al., 2021; Zheng et al., 2021). Therefore, it is necessary to operate the railway traffic safely. Railway maintenance includes tracking the entire railway equipment for both trains and the infrastructure to protect the accidents and malfunctions. Maintenance can basically be categorized as scheduled, corrective and predictive. The scheduled maintenance is a routine for periodically inspection to notice small problems and fix them before the major ones exist. The corrective maintenance includes repairing the equipment when a malfunction or breakdown occurs. The scheduled maintenance can be planned, but corrective maintenance is usually the result of a sudden deterioration that is needed to be immediately fixed. On the other hand, the predictive maintenance is related to monitor a system and evaluate it against historical trends to predict failure before it occurs.

Reliability is the ability of a product to perform its expected function for a specified period of time and in accordance with certain objectives. Reliability analysis is widely used in many engineering activities and has always been an important topic in engineering studies. Considering the hazardous nature of high-speed train operation, the assessment of vehicle reliability is essential to guarantee its safe operation.

Many studies have been carried out with maintenance, repair and accident prevention related to the railways or railway components. Zhang et al. (2012) studied fault diagnosis methods of an absolute positioning sensor which is an important component for a high-speed maglev train. They used support vector machines to recognize the fault characters and the signal flow method to locate the faulty parts. Gomez et al. (2018) used artificial neural networks for vibration signal analysis to detect and diagnose cracks in real railway axles to avoid accidents. Bemment et al. (2017) used historical failure data to improve the reliability and availability of railway track switching. They state that rail switches are important components that provide flexibility to railway networks, but when they fail, they cause disproportionate delays, especially in heavy-traffic passenger rail systems.

The pursuit of better regularity aims to improve operational equipment availability and safety. Marc Antoni (2009) emphasized the importance of ageing of signalling equipment and its impact on maintenance strategy in railways. He used Weibull and Bertholon reliability model to describe an ageing phenomenon and stated that in most cases a systematic replacement strategy offers the best solution. Mokhtarian et al. (2013) proposed a Bayesian nonparametric reliability analysis for a railway system at component level. Shangguan et al. (2020) proposed a board level lifetime prediction for power board of balise transmission module which is an electronic transponder placed between the rails of a railway as part of an automatic train protection system. However, in such methods, the number of fault observations has a direct effect on the analysis results. The low number of observations reduces the effectiveness of the method used and the validity of the predictions accordingly. The lifetimes of seriously important parts used in expensive applications such as high-speed trains are quite long. Therefore, few failures occur during even long terms, but these malfunctions can cause serious consequences or delays at an unexpected times.

The railway system is a large-scale complex system having many components and interconnected subsystems. System reliability has been maintained by appropriate maintenance measures. However, real-life reliability data at the component level of a railway system, far from being complete, are not always available in practice. Manufacturers' component lifetime distributions are often uncertain and complicated by actual use and operating environments. Two important issues on the estimation of parameters in reliability assessment are the statistical model employed and the available data sample sizes (Yang et al., 2018). An effective statistical model provides a better fit to the valid lifetime data and an efficient parameter estimation method can reduce the estimation error. In general, reliability assessments are utilized on large sample sized, i.e. a size of larger than 30, to attain statistical inferences (Yang et al., 2018). In order to determine the reliability models of railway components, long-term failure data records are needed. However, due to money, time and similar constraints, the railway system may need to be improved with the short-term data. Due to the complex structures of components in high-speed trains and long lifetimes owing to their high costs, the reliability assessment has to be performed by using proper methods for small sample sizes (Yang et al., 2018).

In this study, we used real failure data records of high-speed train in Turkey to estimate the lifetimes of some electronic components used in railway signaling. The components were chosen to be the ones that have the failure records more than others and cause significant delays in passenger transport when they break down at unexpected times. Due to having long time life periods of the components, failure records generate small sample sizes. We used 14 different lifetime distributions to model the lifetimes of the selected components. Among them, we observed the following 3 distributions are suitable to use for the examined failure data. These are three parameter Weibull, the largest extreme value and the smallest extreme value distributions. After determining the distribution fit of the relevant component, we used 5 different estimation methods, namely, the maximum likelihood, the least squares, method of moments, method of *L*-moments and robust estimators for the parameter estimates of the distribution. These estimation methods are superior to other estimation methods in certain situations such as the small size nature of the data and also existence of some outliers. The validity and reliability of the result to be obtained from a method depends on the usage assumptions of that method. We also performed a Monte Carlo simulation to compare the estimator performances and finally determine the best estimations for lifetimes of the components. Simulation results demonstrated that since the failure data is a small sized data and it contains outliers, the classical lifetime estimation models may not be suitable. Instead, the robust estimators can model the lifetimes of the components better. The lifetime estimation framework mentioned in this study can be applied to some other components in similar ways.

The organization of the paper is as follows. In the second section, we give the mentioned lifetime distributions, their parameter estimation methodologies and a simulation study to determine the most effective estimation method for different situations. After that, we present the experimental work by first introducing the components used in the study and then giving the real data experimental results. At the end, we finish the paper with the conclusion.

2. METHODOLOGY

The reliability analysis includes to characterize how failures are distributed over the equipment's life. The scope of this study is to interpret failures over time for certain electronics equipment. Failures over time for certain components can be modeled by various probability distribution functions. In this study, we used the most widely used and the most commonly applicable distributions for lifetime analysis. The distributions used in this study are Weibull, Smallest Extreme Value (SEV), Largest Extreme Value (LEV), Rayleigh, Maxwell, Loglogistic, Birnbaum-Saunders, Exponential, Exponential Power, Gamma, Folded Normal, Half Normal and Inverse Gaussian distributions. One can refer to (Cox et al., 2018) for the details of these distributions. Distributions include parameters that are necessary to be determined for a proper fit to the data. The lifetime distributions used for reliability usually use at most three types of parameters which can be called as the shape, scale and location parameters. The scale parameter defines how spread out the distribution is or where the bulk of the distribution lies. The shape parameter defines the general shape of a distribution. The location parameter defines the location of the origin of a distribution. In terms of lifetime distributions, the location parameter represents a time shift.

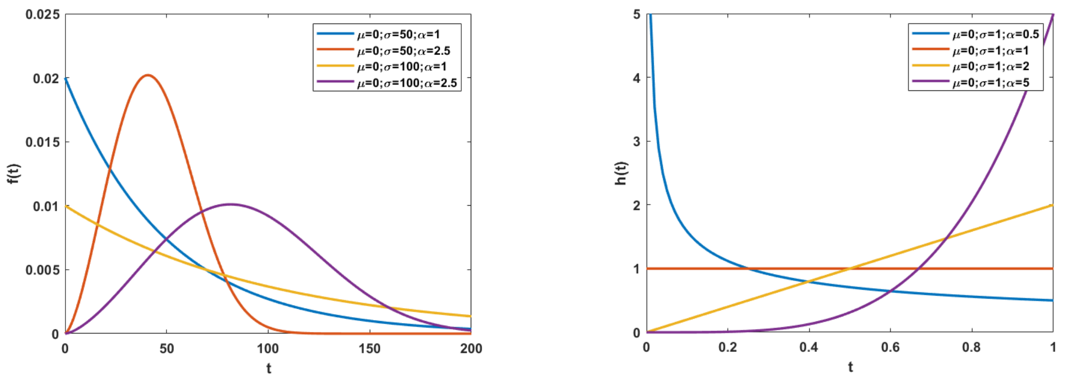
The probability distributions are denoted in terms of the probability density function (pdf), denoted as $f(\cdot)$, or cumulative distribution function (cdf), denoted as $F(\cdot)$ in general. In lifetime analysis, the hazard function, $h(\cdot)$, or the survival function, $S(\cdot)$, can be used to model the chance of death or failure of an element as a function of their age or time. The survival function indicates the probability that the event of interest has not yet occurred by time t and the hazard function represents a conditional density given that the event has not yet occurred prior to time t. The lifetime analysis deals with calculating the expected duration of time until a component failure occurs in mechanical systems. This is usually referred as the survival analysis in statistics or reliability analysis in engineering fields.

Presentation of mathematical formulations of all distributions examined in the article will unnecessarily lengthen the article. Therefore, we will describe the methodology followed in the article by using the three parameter Weibull distribution which is one of the most widely used distribution type in reliability analysis. The procedure operated on this distribution which is chosen as an example, can also be operated on other distributions with different characteristics and different numbers of parameters.

The three parameter Weibull distribution is a flexible distribution for modeling many different data sets (McCool, 2012). Zhu and Liu (2013) used the Weibull distribution in high-speed train bearing reliability estimation and confirmed that they obtained robust results. When the shape parameter of Weibull distribution is 1, it becomes the exponential distribution, if the shape parameter is 2, it becomes Rayleigh distribution and if it is 3.4, it becomes the normal distribution. The probability density function of the three-parameter Weibull distribution is defined as follows:

$$f(t;\alpha,\sigma,\mu) = \frac{\alpha(t-\mu)^{\alpha-1}}{\sigma^{\alpha}} e^{-(\frac{t-\mu}{\sigma})^{\alpha}}, t \leq \mu \leq 0 \tag{1}$$

where σ is the scale, α is the shape and μ is the location parameter. These parameters are



(a) the pdf (b) the hazard function

Figure 1. The pdf and hazard functions of the Weibull distribution for different shape parameters

always positive (Bain and Engelhardt, 1991). The pdfs of Weibull distribution for the different values of the parameters are shown in Figure 1a.

The reliability, $R(\cdot)$, and the hazard, $h(\cdot)$, functions of the Weibull distribution are given in Equations 2 and 3.

$$R(t;\alpha,\sigma,\mu) = e^{-\left(\frac{t-\mu}{\sigma}\right)^{\alpha}}, t > \mu > 0, \alpha > 0, \sigma > 0 \tag{2}$$

$$h(t;\alpha,\sigma,\mu) = \frac{\alpha(t-\mu)^{\alpha-1}}{\sigma^{\alpha}}, t > \mu > 0, \alpha > 0, \sigma > 0 \tag{3}$$

The hazard function is shown in Figure 1b for different values of the shape parameter. If the shape parameter is equal to 1 (exponential distribution), the hazard function is constant; if it is less than 1, the hazard function is exponentially decreasing; if it equals to 2 (Raiyleigh distribution), the hazard function is linearly increasing and if it is greater than 2, the hazard function is exponentially increasing.

Many estimation methods have been proposed to estimate the parameters of the Weibull distribution. Some of the most popular ones are the maximum likelihood estimation (MLE) (Sirvanci and Yang, 1984), ordinary least square estimation (OLS) (Wang et al., 2008), method of moments (MM) (Hall, 2005), Bayesian estimation method (Tsionas, 2003), logarithmic moment estimator (Wang et al., 2010), probability weighted moment estimation (LMOM) (Bartolucci et al., 1999), and robust estimation (Huber, 2009). However, there is no certain proof showing practically which of these methods is the most suitable for use in databases with few samples.

2.1. The Maximum Likelihood Estimation

The maximum likelihood estimation method is a widely used method in statistical estimation theory. It is based on the maximization of the likelihood function (Casella and Berger, 2001). Maximum likelihood estimators of three parameter Weibull distribution are found by maximizing the log-likelihood function with respect to the parameters of interest. The likelihood function, $L(\cdot)$, of three parameter-Weibull distribution is given in Equation 4 (Moeini et al., 2013).

$$L(t_1,\dots,t_n;\alpha,\sigma,\mu) = \prod_{i=1}^{n} \frac{\alpha}{\sigma}\left(\frac{t_i-\mu}{\sigma}\right)^{\alpha-1} e^{-\left(\frac{t_i-\mu}{\sigma}\right)^{\alpha}}; \quad \alpha,\sigma > 0, t_i \geq \mu. \tag{4}$$

The MLE estimators of the related parameters are obtained by taking partial derivatives with respect to the parameter and equalizing the resulting differential equation to zero. Equation 5, 6, 7 present the equations to be solved when taking the partial derivatives with respect to σ, α and μ respectively. Although there is no closed form solution to these equations, the solution can be found by using iterative methods.

$$\frac{n}{\alpha}+\sum_{i=1}^{n}\log\left(\frac{t_i-\mu}{\sigma}\right)-\sum_{i=1}^{n}\left(\frac{t_i-\mu}{\sigma}\right)^{\alpha}\log\left(\frac{t_i-\mu}{\sigma}\right)=0 \tag{5}$$

$$-\frac{n\alpha}{\sigma}+\frac{\alpha}{\sigma}\sum_{i=1}^{n}\log\left(\frac{t_i-\mu}{\sigma}\right)^{\alpha}=0 \tag{6}$$

$$-(\alpha-1)\sum_{i=1}^{n}\frac{1}{t_i-\mu}+\frac{\alpha}{\sigma}\sum_{i=1}^{n}\left(\frac{t_i-\mu}{\sigma}\right)^{\alpha-1}=0 \tag{7}$$

2.2. The Method of Moments

The method of moments is one of the oldest estimation methods that has been used for decades (Barbosa, 2018). Moment methodology solves by equalizing the theoretical moments related to the distribution and moments obtained from the sampling to estimate the unknown parameters. Moment estimators of the three parameter Weibull distribution are found by equating the sample moments to the corresponding theoretical moments which are summarized in Equation 8.

$$\begin{aligned}\mu&=\frac{\mu_1'\mu_4'-\mu_2'^2}{\mu_1'+\mu_4'-2\mu_2'},\\ \sigma&=\frac{\mu_1'-\mu}{\Gamma\left(1+\frac{1}{\alpha}\right)},\\ \alpha&=\frac{\ln(2)}{\ln(\mu_1'-\mu_2')-\ln(\mu_2'-\mu_4')}\end{aligned} \tag{8}$$

Here, the k^{th} order moment obtained from the sample to show the observations in the random sample $t_{(1)}\le t_{(2)}\le\cdots\le t_{(n)}$ as follows:

$$m_k'=\frac{\sum_{i=1}^{n}t_i^k}{n} \tag{9}$$

The moments estimation of the parameters is obtained by solving the system of equations given in Equation 8.

2.3. The Method of L-Moments

Hosking introduced L-moments as an analogous to the conventional moments (Hosking, 1990). L-moments can be defined for any random variable whose mean exists and they can be estimated by a linear combination of order statistics (Hosking, 2007). They are more resistant to the influence of sample variation and robust to the outliers in the data (Abdul-Moniem and Selim, 2009). L-moments are often used for a more efficient parameter estimation of a parametric distribution than the maximum likelihood method, especially for small samples (Arslan et al., 2014).

Let T be a continuous random variable with the cumulative distribution function $F(t)$ and quantile function $Q(t)$, then L-moments of r^{th} order random variable is obtained as follows:

$$\lambda_r = \frac{1}{r}\sum_{j=1}^{r-1}(-1)^j \binom{r-1}{j} E\left(T_{r-j:r}\right); r = 1,2,3,\dots \tag{10}$$

where E is the expected value, given in Equation 11, and $T_{k:n}$ denotes the k^{th} order statistic in an independent sample of size n.

$$E\left(T_{r-jr}\right) = \frac{n!}{(r-1)!(n-r)!}\int_0^1 Q(F)F^{r-1}(1-F)^{n-r}dF \tag{11}$$

Let $t_1, t_2, \dots, t_n$ be samples and $t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(n)}$ be the ordered samples, then the r^{th} unbiased sample L-moments can be written as in Equation 12.

$$\ell_r = \binom{n}{r}^{-1} \sum_{1\leq i_1} \sum_{<i_2<\dots} \cdots \sum_{<i_r\leq n} \frac{1}{r}\sum_{j=1}^{r-1}(-1)^j \binom{r-1}{j} t_{ir-jn}; \tag{12}$$

$$\cdot = 1,2,\dots,n.$$

L-moment estimators of the three parameter Weibull distribution are calculated by solving the three equations-unknowns set given in Equation 13. Each line in the equation is formed by equating the sample L-moments to the corresponding theoretical moments.

$$\begin{aligned} \ell_1 &= \sigma_L \Gamma\left(1+\frac{1}{\alpha}\right) + \mu_L \\ \ell_2 &= \sigma_L \Gamma\left(1+\frac{1}{\alpha}\right)\left[1-\left(\frac{1}{2}\right)^{\frac{1}{\alpha}}\right] \\ \ell_3 &= \sigma_L \Gamma\left(1+\frac{1}{\alpha}\right)\left[1-3\left(\frac{1}{2}\right)^{\frac{1}{\alpha}}+2\left(\frac{1}{3}\right)^{\frac{1}{\alpha}}\right] \end{aligned} \tag{13}$$

2.4. The Method of Least Squares

Let $T_1, T_2, \ldots, T_n$ be a random sample of size n from the cumulative distribution function $F(\cdot)$. Least squares (LS) estimators of the unknown parameters of $F(\cdot)$ are obtained by minimizing the Equation 14 with respect to the parameters of interest. In the equation, $T_{(1)} \le T_{(2)} \le \cdots \le T_{(n)}$ are the ordered random variables.

$$\sum_{i=1}^{n}\left\{F\left(T_{(i)}\right)-\frac{i}{n+1}\right\}^2 \tag{14}$$

Then, the LS estimators of three parameter Weibull distribution are obtained by minimizing the following function with respect to the parameters α, σ and μ.

$$W(t;\alpha,\sigma,\mu) = \sum_{i=1}^{n}\left\{1-e^{-\left(\frac{t_{(i)}-\mu}{\sigma}\right)^{\alpha}}-\frac{i}{n+1}\right\}^2 \tag{15}$$

Parameter estimation for this method is obtained by using iterative methods similar to maximum likelihood estimators.

2.5. Robust Estimation

Some of the data may become quite different from the overall data in reliability analysis. This kind of data is usually referred to as outliers. The outliers have a negative effect on classical estimators. Especially, the existence of outliers causes estimators to converge to wrong parameter estimations and decreases estimator efficiency. On the other hand, robust estimators are less affected by model assumptions, deviations from assumptions and outliers. Robust estimators have a large family. In this study, *M* estimators were used because of the easiness of their computation and common use. These estimators are called *M* estimators as they are maximum likelihood type estimators. *M* estimators were proposed by Huber in 1964 (Huber, 2009). Rather than minimizing the sum of squares of errors, *M* estimators minimize the errors according to a ρ function which is differentiable and symmetrical with a single minimum at zero. Robust M estimators of the unknown parameters of $F(\cdot)$ are obtained by minimizing the following equation.

$$\sum_{i=1}^{n} \rho \left\{ \frac{\left(F\left(T_{(i)}\right) - \frac{i}{n+1} \right)}{\hat{\sigma}} \right\} \tag{16}$$

Robust estimators of three parameter Weibull distribution are obtained by minimizing the function given in Equation 17.

$$W(t; \alpha, \rho, \mu) = \sum_{i=1}^{n} \rho \left\{ \frac{\left(1 - e^{-\left(\frac{t_{(i)} - \mu}{\sigma}\right)^{\alpha}} - \frac{i}{n+1} \right)}{\hat{\sigma}} \right\} \tag{17}$$

Here, $\hat{\sigma}$ is taken as the mean absolute deviation (MAD) where $MAD = \frac{median|e_i - median(e_i)|}{0.6745}$ and the constant 0.6745 is chosen so that it is asymptotically unbiased for the normal error case. The robust *M* estimators for the parameters are obtained by taking the derivative with respect to the related parameter and solving the final equation.

3. SIMULATION STUDY

In this section, we present a comprehensive Monte Carlo simulation that we did to compare the performances of different estimators. In the simulation, the performances of the estimators were compared according to the mean squared error (MSE) criteria and bias value by using small size full samples with and without outliers. Due to the relationship between Weibull and SEV distributions, the simulation is formed based on Weibull and SEV distributions if the failure data follows a Weibull distribution and the logarithm of the failure data follows a SEV distribution (Lv et al., 2015). That's why, no extra simulation was performed for SEV distribution as the results obtained in Weibull distribution would be valid for the SEV. As the reliability analysis was performed with small data sets, we used $n = 5$, 10, 20 and 50 sample sizes. In addition, the cases whether the data set contains an outlier or not is taken into account. For this purpose, the data set with a positive outlier in the x direction was studied. We conducted the simulation by selecting the following parameter values: β =1, 2, 3.4 and 8; $\mu = 0$; $\sigma = 1$ for Weibull distribution and $\mu = 0$; $\sigma = 1$ values for LEV distribution. The estimators used were compared for both n and all parameter combinations for data sets with and without outliers. R package program has been used to perform calculations. After $\frac{100000}{n}$ Monte Carlo experiments, the MSE and (Bias) values were calculated as follows. The results obtained for LEV distribution are given in Table 1.

$$MSE(\hat{\theta}) = \frac{1}{\left(\frac{100000}{n}\right)} \sum (\hat{\theta} - \theta)^2 \tag{18}$$

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta \tag{19}$$

$$E(\hat{\theta}) = \frac{1}{\left(\frac{100000}{n}\right)} \sum \hat{\theta} \tag{20}$$

Table 1 shows the bias and MSE values of the parameter estimators of the LEV distribution for different sample sizes. The predictor with the smallest MSE value is called the efficient predictor (Casella and Berger, 2001). Observations in Table 1 demonstrate that for the outlier data, if the sample size for the M location parameter is between 5 and 20, the predictor with the smallest MSE value is the LMOM estimator. But when n=50, the predictor with the smallest MSE value is the M estimator. Next comes the MLE estimator. This result also coincides with the theory. Similar observations are valid for the σ parameter of the LEV distribution. When the sample size increases, robust *M* estimator gives more effective results. Under the existence of outliers, we found that the efficiency of all estimators decreased and the MSE values increased. In this case, the robust *M* estimator is the predictor with the smallest MSE value for both parameter estimations in all sample sizes.

In case of a small number of data sets, such as the failure data experimented in this study that contains outlier, LS estimators, which are often used for parameter estimation, are the least effective estimator. In other words, the estimates obtained by this method do not reflect the truth. Robust *M* estimator, on the other hand, is the most efficient estimator for the data sets that have outliers.

Simulation results for Weibull distribution is presented in Tables 2 and 3. In Table 2, LMOM estimation gives the most effective results to estimate the parameters of Weibull distribution in the data set without outlier for the shape parameter that is equal to 1 (which corresponds to the exponential distribution). In Table 2, for cases of β=1, the data without outlier and the sample size is between 5 and 20, it is seen that the most effective estimator for all parameters is LMOM. When the sample size is 50, robust *M* and MLE estimators are effective. This is because the LMOM estimator is based on order statistics. Estimators based on order statistics give more effective results in small samples. As the sample size increases, the efficiency of the predictor decreases as the importance of the ranking decreases. The most

Table 1. Simulation results for LEV

Results without Outlier								
Estimators	n=5				n=10			
	μ=0		σ=1		μ=0		σ=1	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
ML	0.0962	0.2454	0.0811	0.1961	0.0436	0.1220	0.0625	0.099
LS	0.0203	0.2686	0.3211	0.7564	0.0092	0.1313	0.1120	0.1317
MOM	0.0658	0.2339	0.0717	0.1897	0.0261	0.1189	0.4790	0.0912
LMOM	0.0038	**0.2098**	0.0445	**0.1807**	0.0035	**0.1141**	0.0121	**0.0849**
M Estimator	0.0943	0.2439	0.0803	0.1975	0.0437	0.1207	0.0609	0.0951
	n=20				n=50			
ML	0.0132	0.0543	0.0619	0.0743	0.0089	0.0204	0.0077	0.0132
LS	0.0081	0.0603	0.0841	0.0830	0.0045	0.0234	0.0196	0.0205
MOM	0.0103	0.0557	0.3810	0.0649	0.0078	0.0215	0.0024	0.0196
LMOM	0.0025	**0.0518**	0.0108	**0.0647**	0.0015	**0.0209**	0.0085	**0.0146**
M Estimator	0.0129	0.0537	0.0306	0.0735	0.0087	0.0197	0.0055	0.0123
Results with Outlier								
	n=5				n=10			
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
ML	0.1043	0.2960	0.1920	0.3611	0.0643	0.2164	0.0977	0.1088
LS	0.0351	0.3085	0.6789	0.9250	0.0129	0.2313	0.3120	0.3024
MOM	0.0736	0.3016	0.1827	0.3582	0.0612	0.2218	0.0880	0.1464
LMOM	0.0125	0.2983	0.1820	0.3458	0.0113	0.2165	0.0989	0.1411
M Estimator	0.1041	**0.2825**	0.1581	**0.3393**	0.0601	**0.2124**	0.0957	**0.1041**
	n=20				n=50			
ML	0.0245	0.0727	0.0671	0.0892	0.0106	0.0402	0.0562	0.0271
LS	0.0105	0.0842	0.1345	0.1053	0.0073	0.0542	0.0924	0.0472
MOM	0.0216	0.0728	0.0679	0.1025	0.0128	0.0421	0.0582	0.0343
LMOM	0.0108	0.0736	0.0741	0.1011	0.0057	0.0408	0.0681	0.0292
M Estimator	0.0106	**0.0703**	0.0518	**0.0780**	0.0097	**0.0399**	0.0472	**0.0203**

effective predictor is the robust *M* estimator in all parameters and in all sample sizes for results in the data with outlier case.

If there is only one outlier in the data set, Robust *M* estimator is the most effective in all sample sizes and all parameter estimates. The results of the simulation obtained for Weibull distribution when the shape parameter is 2 (which corresponds to the Raiyleigh distribution) demonstrates that when the sample size is small (i.e., $n < 30$), the most efficient estimator for all parameters is LMOM in the data set without outliers. When the sample size is increased (i.e., $n = 50$), the most effective estimators are listed as LMOM, Robust *M*, ML, MOM and LS. Robust *M* estimator gives the most effective result in the data with outlier for all cases.

In Table 3, the most effective estimator for all parameters is seen as LMOM in the data set without outlier ($n < 20$). However, when the sample size larger than 20, the efficiencies of all estimators approach to each other. This is because when the shape

Table 2. Simulation results for 3 parameter Weibull for β =1 and 2.

Estimators	Data without outlier						Data with outlier					
	n=5											
	β=1		μ=0		σ=1		β=1		μ=0		σ=1	
	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
ML	-0.1341	0.3919	0.1974	0.0784	-0.0529	0.7303	0.2348	0.4386	-0.2735	0.1718	-0.1295	0.8730
LS	0.1695	0.3957	-0.2272	0.1093	-0.0623	0.7524	0.2695	0.4405	-0.3213	0.1923	-0.1632	0.8842
MOM	0.1423	0.3942	0.2017	0.0851	0.0684	0.7459	0.2423	0.4371	0.2852	0.1524	0.1368	0.8746
LMOM	0.1235	0.3891	0.1981	0.0716	0.068	0.7045	0.2385	0.4382	0.2451	0.1712	0.1268	0.8745
M Estimator	0.1393	0.3916	0.2041	0.0791	-0.0531	0.7314	0.2315	0.4316	-0.2711	0.1342	-0.1283	0.8720
	n=10											
ML	0.0853	0.2801	0.0891	0.0208	-0.0387	0.4116	-0.1835	0.3233	-0.1829	0.1218	-0.0874	0.5241
LS	0.0873	0.3286	-0.1983	0.0773	-0.0388	0.4568	-0.1973	0.3313	-0.2798	0.1837	-0.0939	0.5426
MOM	0.0816	0.2951	0.1016	0.0268	0.0465	0.4303	-0.1836	0.3237	0.1642	0.1257	0.0927	0.5346
LMOM	0.0647	0.2769	0.0861	0.0199	0.0467	0.4018	-0.1837	0.3232	0.1639	0.1199	0.0865	0.5329
M Estimator	0.0850	0.2818	0.0903	0.0209	-0.0391	0.4165	-0.0921	0.3149	-0.1109	0.0902	-0.0489	0.4718
	n=20											
ML	-0.0641	0.0376	0.0472	0.0109	-0.0321	0.1627	-0.1465	0.0936	-0.1342	0.0918	-0.0665	0.3934
LS	0.0763	0.0463	0.2547	0.0701	-0.0365	0.1827	-0.1492	0.0946	-0.2098	0.1374	-0.0708	0.4066
MOM	0.0751	0.0452	0.0504	0.0126	0.0418	0.1721	-0.1469	0.0945	0.1234	0.0946	0.0693	0.4018
LMOM	0.0512	0.0365	0.0451	0.0105	0.0431	0.1618	-0.1472	0.0936	0.123	0.0898	0.0695	0.3995
M Estimator	0.0638	0.0377	0.0469	0.0110	-0.0314	0.1626	-0.0691	0.0736	-0.0835	0.0679	-0.0362	0.3088
	n=50											
ML	0.0374	0.0298	0.0191	0.0076	-0.0372	0.0336	-0.0738	0.0455	-0.0639	0.0438	-0.0324	0.1911
LS	0.0483	0.0342	0.2634	0.0364	-0.0409	0.0381	-0.0786	0.0461	-0.1025	0.0662	-0.0348	0.1966
MOM	0.0401	0.0347	0.0216	0.0084	0.0521	0.0377	-0.0743	0.0460	0.0669	0.046	0.0336	0.1945
LMOM	0.0349	0.0327	0.0195	0.0082	0.0643	0.0366	-0.0745	0.0458	0.0682	0.0441	0.0337	0.1932
M Estimator	0.0310	0.0281	0.0189	0.0076	-0.0298	0.0335	-0.0339	0.0335	-0.0413	0.0319	-0.0169	0.1451
	n=5											
	β=2		μ=0		σ=1		β=2		μ=0		σ=1	
	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
ML	-0.8090	3.0161	0.3111	0.2515	-0.3968	0.4268	-1.2318	3.3825	-0.2388	0.2627	-0.1032	0.7686
LS	-0.8191	3.0435	-0.3890	0.2810	0.4661	0.5198	-1.2438	3.4259	-0.2481	0.2742	-0.1173	0.7898
MOM	0.8201	3.0398	0.3754	0.2615	0.4208	0.4395	-1.2368	3.3938	0.2394	0.2646	0.1154	0.7787
LMOM	0.8087	3.0115	0.3109	0.2498	0.3908	0.4226	-1.2335	3.3785	0.2391	0.2613	0.1023	0.7725
M Estimator	-0.8088	3.0159	0.3110	0.2518	-0.3971	0.4373	-1.2305	3.2324	-0.2282	0.2579	-0.1033	0.7630
	n=10											
ML	-0.7456	1.1377	0.2253	0.0914	-0.3019	0.2197	-0.9123	1.8608	-0.1813	0.1303	-0.0764	0.4231
LS	-0.8209	1.5176	-0.2855	0.1699	0.3258	0.2812	-0.9207	1.8797	-0.1891	0.1764	-0.0871	0.4356
MOM	0.7890	1.4916	0.2641	0.1126	0.3168	0.2616	-0.9153	1.8672	0.1792	0.1343	0.0856	0.4295
LMOM	0.7010	1.1218	0.2218	0.0903	0.3016	0.2130	-0.9127	1.8597	0.1772	0.1299	0.0761	0.4248
M Estimator	-0.7449	1.1372	0.2256	0.0921	-0.3020	0.2200	-0.5045	1.5192	-0.0943	0.1129	-0.0442	0.4067
	n=20											
ML	-0.4125	0.3906	0.1413	0.0387	-0.1915	0.0960	-0.5931	0.7451	-0.1151	0.0531	-0.0501	0.1684
LS	-0.4514	0.4078	-0.1803	0.0413	0.2195	0.1616	-0.5984	0.7545	-0.1197	0.0597	-0.0572	0.1739
MOM	0.4418	0.3928	0.1642	0.0395	0.2019	0.1151	-0.5946	0.7468	0.1164	0.0568	0.0564	0.1712
LMOM	0.4029	0.3876	0.1397	0.0381	0.1911	0.0912	-0.5934	0.7438	0.1153	0.0508	0.0498	0.1658
M Estimator	-0.4119	0.3901	-0.1419	0.0391	0.1914	0.0958	-0.2594	0.5321	-0.0862	0.0467	-0.0228	0.1209
	n=50											
ML	-0.2073	0.1121	0.0750	0.0175	0.0231	0.0093	-0.3852	0.3294	-0.0747	0.0213	-0.0323	0.0867
LS	-0.3241	0.1345	0.0821	0.0236	0.0372	0.0103	-0.3899	0.3845	-0.0775	0.0226	-0.0368	0.0887
MOM	0.3029	0.1182	0.0783	0.0198	0.0329	0.0094	-0.3867	0.3816	0.0749	0.0215	0.0361	0.0874
LMOM	0.1918	0.1135	0.0741	0.0183	0.0280	0.0093	-0.3859	0.3791	0.0748	0.0214	0.0389	0.0845
M Estimator	-0.1984	0.1009	0.0749	0.0175	0.0235	0.0093	-0.1386	0.2391	-0.0238	0.0196	-0.0113	0.0554

parameter is 3.4, Weibull distribution is converted to the Normal distribution. The most effective estimator for the data set with outliers is the robust M estimator. For the shape parameter of 8, LMOM is the most efficient estimator for the data set ($n < 50$) without outliers. When the sample size is 50, the efficiencies of all estimators approach to each other. In the data set with outliers, robust M estimator gives the most effective result with the distortion of the distribution. Since SEV distribution is derived from a transformation of Weibull distribution, the same results would be obtained.

When all of the simulation results are evaluated together, we obtain the following interpretation. Even though LS or ML estimators are widely used in the literature, they are not appropriate to use with small sized data sets like the failure data used in this study. In our work, Monte Carlo simulation results verify that it is appropriate to choose LMOM estimator for the data set without outliers and Robust M estimator for the data set with outliers.

Table 3. Simulation results for 3 parameter Weibull for β =3.4 and 8.

Estimators	Data without outlier						Data with outlier					
	n=5											
	β=3.4		μ=0		σ=1		β=3.4		μ=0		σ=1	
	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
ML	0.3102	4.6103	0.0572	0.3666	-0.6521	0.4728	-2.3523	5.7978	0.0824	0.4322	-0.0910	0.6767
LS	0.3982	4.6280	-0.0681	0.4219	0.7288	0.5124	-2.4244	5.9392	-0.0972	0.4460	0.1024	0.7108
MOM	0.3827	4.8134	0.0601	0.4028	0.6785	0.4955	-2.3824	5.8996	0.0846	0.4333	0.0945	0.6773
LMOM	0.3789	**4.6013**	0.0559	**0.3514**	0.5841	**0.4547**	-2.3459	5.7806	0.0819	0.4321	0.0921	0.6735
M Estimator	0.3114	4.6059	0.0569	0.3670	-0.6601	0.4735	-2.3008	**5.7003**	0.0817	**0.4221**	-0.0907	**0.6266**
	n=10											
ML	0.1452	1.2519	0.0352	0.1939	-0.4034	0.2519	-1.9802	4.0585	0.0671	0.2596	-0.0765	0.5282
LS	0.2160	1.3640	-0.0436	0.2811	0.5312	0.3542	-2.0319	4.1573	-0.0819	0.3122	0.0860	0.5529
MOM	0.2084	1.4103	0.0391	0.2261	0.4522	0.2814	-2.0012	4.1299	0.0724	0.2603	0.0794	0.5467
LMOM	0.1981	**1.2495**	0.0349	**0.1721**	0.3865	**0.2457**	-1.9711	4.0468	0.0691	0.2595	0.0774	0.5271
M Estimator	0.1461	1.2521	0.0360	0.1941	-0.4051	0.2529	-0.7831	**3.0564**	0.0284	**0.1921**	-0.0308	**0.4175**
	n=20											
ML	0.0267	0.7942	0.0246	0.1015	-0.2774	0.1266	-1.3663	2.5413	0.0322	0.1256	-0.0643	0.2948
LS	0.0493	0.8021	-0.0281	0.1103	0.3212	0.1316	-1.7069	2.9207	-0.0691	0.1326	0.0722	0.3874
MOM	0.0341	0.8213	0.0264	0.1125	0.2915	0.1321	-1.6819	2.8909	0.0609	0.1294	0.0667	0.3867
LMOM	0.0244	**0.7912**	0.0231	**0.1010**	0.2719	**0.1254**	-1.6559	2.8323	0.0591	0.1287	0.0651	0.3691
M Estimator	0.0321	0.7977	0.0259	0.1018	-0.2769	0.1262	-0.2671	**1.4232**	0.0099	**0.1035**	-0.0125	**0.2593**
	n=50											
ML	0.0070	**0.5321**	0.0148	**0.0419**	0.0501	**0.0031**	-0.1147	1.8968	0.0054	0.0524	-0.0054	0.0786
LS	0.0165	0.5327	-0.0179	0.0421	0.0515	0.0032	-1.4349	2.0542	-0.0594	0.0794	0.0607	0.2712
MOM	0.0185	0.5343	0.0152	0.0436	0.0506	**0.0031**	-1.4134	2.0263	0.0516	0.0698	0.0563	0.2707
LMOM	0.0145	0.5338	0.0141	0.0440	0.0496	**0.0031**	-1.3908	2.0126	0.0497	0.0693	0.0547	0.2583

M Estimator	0.0052	**0.5321**	0.0148	**0.0417**	0.0499	**0.0031**	-0.1092	**1.1452**	0.0036	**0.0502**	-0.0042	**0.0516**
	n=5											
	β=8		μ=0		σ=1		β=8		μ=0		σ=1	
	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
ML	-0.5780	24.5989	0.6636	0.5857	-0.7448	0.5640	-5.0224	55.7298	0.2842	0.9216	-0.1923	1.0341
LS	-0.6258	25.4558	0.6978	0.6987	0.8259	0.8650	-5.0226	55.9139	-0.2972	0.9595	0.2043	1.8754
MOM	0.5941	24.6847	0.6821	0.5874	0.7745	0.6258	-5.0224	55.8800	0.2855	0.9326	0.1935	1.8426
LMOM	0.5517	**24.0130**	0.6614	**0.5086**	0.7045	**0.5610**	-5.0223	55.7914	0.2839	0.9211	0.1920	1.6339
M Estimator	-0.5800	24.6121	0.6641	0.5861	-0.7467	0.5642	-5.0223	**53.7209**	0.2839	**0.8214**	-0.1927	**0.9340**
	n=10											
ML	-0.4424	7.7845	0.4541	0.3394	-0.5556	0.3776	-4.5116	23.1658	0.2562	0.4793	-0.1735	0.7512
LS	-0.5013	8.2159	0.5223	0.4595	0.6324	0.4108	-4.5208	23.3225	-0.2675	0.5139	0.1842	0.7881
MOM	0.4842	7.9254	0.4859	0.3657	0.5843	0.3935	-4.5199	23.2926	0.2571	0.4894	0.1746	0.7586
LMOM	0.4315	**7.6914**	0.4520	**0.3132**	0.5507	**0.3718**	-4.4826	23.2126	0.2561	0.4792	0.1732	0.7509
M Estimator	-0.4428	7.7862	0.4538	0.3389	0.5531	0.3781	-2.2648	**18.6312**	0.0149	**0.3762**	-0.0898	**0.4506**
	n=20											
ML	-0.3921	3.1893	0.3804	0.2027	-0.3987	0.2208	-3.7065	18.9148	0.0125	0.3263	-0.1159	0.2948
LS	-0.4003	3.4518	0.4247	0.3193	0.4286	0.3025	-4.1589	20.0574	-0.2461	0.3952	0.1689	0.3574
MOM	0.3948	3.3211	0.3918	0.2549	0.4026	0.2515	-4.1586	20.0326	0.2365	0.3246	0.1601	0.3067
LMOM	0.3887	**3.1745**	0.3704	**0.2005**	0.3883	**0.2192**	-4.1582	19.9514	0.2354	0.3159	0.1592	0.3011
M Estimator	-0.3918	3.1887	0.3805	0.2018	-0.3983	0.2206	-1.1583	**8.7892**	0.0095	**0.2288**	-0.0069	**0.2759**
	n=50											
ML	-0.2708	1.0332	0.2632	0.1109	-0.2728	**0.0911**	-1.9645	10.5549	0.1247	0.1326	-0.0856	0.3426
LS	-0.3102	1.2650	0.2921	0.1516	0.2296	0.0932	-3.0359	12.6368	-0.1797	0.2546	0.1235	0.4269
MOM	-0.2928	1.1759	0.2831	0.1312	0.2418	0.0940	-3.0358	12.6059	0.1725	0.2109	0.1169	0.4102
LMOM	-0.2725	1.1762	0.2763	0.1126	0.2321	0.0940	-3.0352	12.5763	0.1716	0.2053	0.1160	0.4036
M Estimator	-0.2702	**1.0321**	0.2710	**0.1109**	-0.2649	**0.0911**	-0.0998	**4.0032**	0.0863	**0.1268**	-0.0048	**0.3319**

4. LIFETIME ANALYSIS OF SOME ETCS COMPONENTS

The first high speed train experience of Turkey has been started with Ankara-Eskisehir line in 2009. Following the first high speed line, two more lines were put into service in 2011 (Ankara-Konya) and 2014 (Ankara-Istanbul). In this section, reliability analysis of some components in the high-speed train signaling system of Turkey has been made. Components were selected as signaling components that deteriorated more frequently than most of the others, according to failure records obtained in the mentioned 3 lines above up to the year 2020.

4.1. Used Equipments

Railway signaling equipments are a system of components used to direct railway traffic and keep trains away from possible accidents. This system contains a wide variety of electronic materials or blocks. In this study, we selected 5 components of the high-speed train signaling system that more frequent failures of them has been reported and when they fail, they cause significant delays. These are the signal lamps, electronic input-output cards, signal electronic control cards, balise transmission modules and SRG213/214 and RG400 connectors.

4.1.1. Signal Lamp

The wayside signals are the elements transmitting the information to perform the train headway regulation and are located in the stations and in the routes between stations. These wayside signals give the movement authority to the train. Thus, each wayside signal protects a track section and the aspect shown to the driver indicates the authorization given to move forward though the protected track. The wayside signals for high-speed railways can be classified as the main signals, dwarf signals and alpha-numeric signals as. The main signals may have three or four aspects, having the same kind of mast, base and ladder for each of them. The main signals with four aspects are located in the station areas and are used as entry and exit signals when protecting a turnout. The main wayside signals with three aspects (Figure 2a) will be located in station areas and in the routes between stations to protect blocks. The dwarf signals are located in the station areas for secondary tracks (Figure 2b). The alpha-

(a) The main signal	(b) The dwarf signal	(c) The alpha-numeric signal

Figure 2. Some illustrations of signal lamps

numeric signals are used to indicate the speed and/or direction for the next turnout. This signal is always installed on the main signal pole with a special connection (Figure 2c).

4.1.2. Electronic Input/Output Card

Electronic input/output (E-I/O) card is located in the interlocking cabinet. E-I/O card is used to control the receiving and sending of commands of turnouts, track circuits, locks, blocks and each element has a separate input/output voltage. Each E-I/O card can send 16 commands that are sent via two independent channels (Channel-A and B). The outputs have backup. At the same time, these cards can receive 16 notifications.

4.1.3. Signal Electronic Control Card

Signal electronic control card (SEC) is located in the interlocking cabinet. SEC controls the lamps up to 8 pieces. It turns on and off the lamps. It provides a circuit for measuring current and voltage. It detects and controls the short circuit and burning situations that may occur in the lamps. Figure 3 shows E-I/O and SEC cards located on the electronic interlocking cabinet.

Figure 3. Electronic input/output and signal electronic control cards

4.1.4. Balise Transmission Module

Balise transmission module (BTM) is an important equipment that is responsible to carry safety-related information between track side equipment and the train. This equipment has been consisted of an antenna which is used for activation and the reading of the data and a tool for maintenance of eurobalises (Stanley et al., 2011). Balise is a punctual transmission

(a) Balise antenna (b) Fixed eurobalise (c) Programmable eurobalise

Figure 4. Eurobalise components

device that are installed on the track. It can be energized by balise antenna (Figure 4a), which produces radio frequency. When the train goes over it, it sends a telegram to the on-board equipment. Telegram has some information about the geographical position of train, gradient of track, radius of curvature and level transitions. Eurobalises are of either a fixed (Figure 4b) or programmable (Figure 4c) type.

4.1.5. SRG213/214 and RG400 Connector

RG 400 and RG 213/214 are connectors of antenna connection. These are the connectors of the cables that provide communication between antenna and balise reader. Coded signaling data (telegrams) received from balise are sent to the balise reader by RG 400 and RG 213/214 connectors. Balise reader converts this information to the digital data and shares with the high-speed train. Figure 5a and 5b shows RG213/214 N type male and Figure 5c shows RG400 SMA type male connectors.

(a) RG213/214 N type (male) (b) RG213/214 N type (male) (c) RG400 SMA type (male)

Figure 5. SRG213/214 and RG400 connectors

5. EXPERIMENTS

In this section, we present the reliability analysis of the signaling components mentioned in the previous subsections. The data of E-I/O and SEC cards has been obtained from the malfunctions recorded within the period between 2009-2019. These failures occurred in the technical buildings at the 456th and 498th kilometers in Ankara-Istanbul Line Project Phase-1, respectively. The data of the electrans green signal lamp consists of malfunctions within the 10-years period between 2011-2020. The malfunctions of the relevant component belong to the S03 numbered signal in Siding-2, one of the regions within the scope of Ankara-Konya Line Project. Balise antenna, failure data of RG213/214 and RG400 connectors were obtained from 65008 TCF, which operates between various provinces and districts within the scope of Ankara-Istanbul YHT Project Phase-1 and Ankara-Konya YHT Project. The data of these components consist of failures recorded over a 8-years period between 2012-2019. The suitability of distribution fit for each equipments' failure data were performed by using Kolmogorov-Smirnov test (Huber-Carol et al., 2012). The obtained results are summarized in Table 4.

The values presented in the table are the probability (p) values. These values show how the relevant equipment is compatible with each of the tested distributions. The highest probability value for each equipment is written in bold. The distribution that has the highest probability value implies that it is the most suitable distribution for modeling of failure data. The most suitable probability distribution for the E-I/O card is determined as a three parameter Weibull distribution. The Largest Extreme Value is the most suitable distribution method for the failure data of SEC card and Electrans green signal lamps. The Smallest Extreme Value is the most suitable distribution method for the failure data of Balise antenna, RG213/214 and RG400 connectors.

Once the distribution of the failure data has been determined, it should be investigated whether the data set contains an outlier. The outliers are observations that have a very different structure than the rest of the data and have negative effects on parameter estimation. Figure 6 shows the box-plots of the components considered in the study. Box-plot

Table 4. Distribution fitting of failure data

Distributions	E I/O (n=16)	Balise (n=13)	SEC (n=14)	RG213 (n=14)	GSL (n=18)	RG400 (n=14)
Birnbaum-Saunders	0.959	0.411	0.545	0.581	0.282	0.576
Exponential	0.019	0.512	0.046	0.621	0.040	0.613
Exponential Power	0.949	0.382	0.569	0.46	0.870	0.491
Folded Normal	0.927	0.670	0.261	0.601	0.859	0.604
Gamma	0.944	0.541	0.452	0.627	0.797	0.616
Half Normal	0.941	0.538	0.421	0.507	0.137	0.508
Inverse Gaussian	0.959	0.128	0.550	0.527	0.277	0.529
Largest Extreme Value	0.968	0.592	**0.725**	0.656	**0.894**	0.638
Loglogistic	0.939	0.613	0.680	0.691	0.742	0.682
Maxwell	0.921	0.581	0.293	0.665	0.858	0.643
Rayleigh	0.926	0.526	0.289	0.659	0.620	0.638
Smallest Extreme Value	0.963	**0.742**	0.151	**0.742**	0.815	**0.789**
Weibull	0.935	0.535	0.300	0.623	0.843	0.613
Weibull (3-Parameter)	**0.980**	0.664	0.113	0.605	0.806	0.628

representation is a widely used tool in data analysis which shows the basic structures of the data set such as minimum, maximum, percentage and mean value. It is useful for giving information about the variability and distribution of the data. In the box-plot, samples that are out the range of the minimum and maximum values are called outliers. There is no outlier in the failure data in box-plot drawings for EIO. It is also seen that the distribution is close to the symmetry. It is also noticed that the failure data for part RG213-214 had an asymmetrical distribution and did not contain outliers. On the other hand, there exists outliers in the failure data of SEC, Green lamp, Balise and RG400 components.

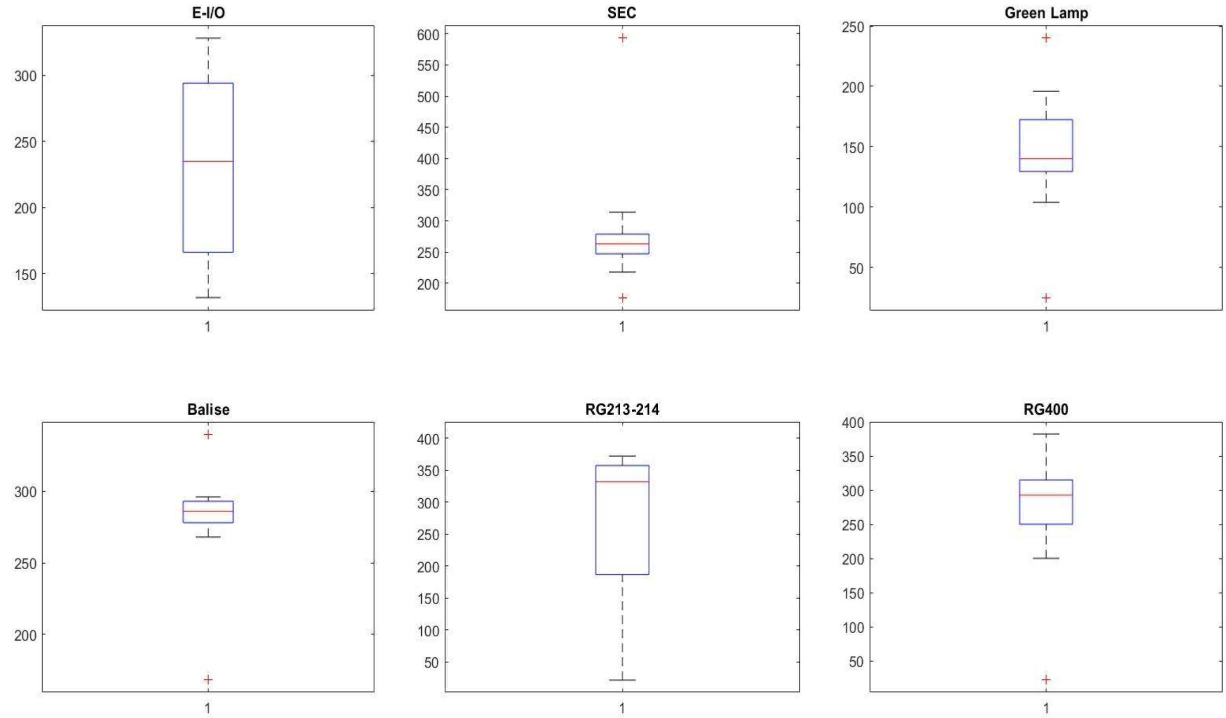


Figure 6. Box-plots of the components

Table 5. The parameter estimations of lifetime distributions of the components.

	MLE Likelihood	Moments	LMOM	LS	M
E I/O (Weibull distribution with three parameter)					
μ	123.129	70.836	39.676	98.254	129.142
α	1.534	2.564	2.856	3.141	1.612
σ	115.150	176.358	210.677	211.462	123.241
SEC (LEV)					
μ	257.328	250.241	193.722	268.415	190.316
σ	64.981	85.435	58.375	56.438	54.247
Electrans Green Signal Lamp (LEV)					
μ	135.469	139.990	136.540	173.275	138.214
σ	68.653	50.692	56.670	62.916	42.163
Balise Antenna (SEV)					
μ	312.200	162.928	157.288	356.102	175.204
σ	117.984	127.028	136.798	131.162	98.246
RG213/214 Connector (SEV)					

μ	289.771	135.639	130.450	324.675	188.138
σ	129.729	128.827	137.818	142.253	112.605
RG400 Connector (SEV)					
μ	286.317	135.262	129.833	312.421	229.035
σ	127.559	126.208	135.613	128.136	116.127

After the suitable distribution method of failure data is determined for each equipment, the parameters of these probability distributions have to be estimated. In this study, we used five different methods to estimate the parameters of each distribution. The estimated values of the corresponding components with different parameter estimation methods are given in Table 5.

In Table 5, E-I/O card shows quite different results. Since there are only 16 failure entry for this component, the parameter estimation is realized by using a very small sampling size. In addition, the shape parameter has a great importance for Weibull distribution. The estimation values obtained for the shape parameter vary between 1.5 and 3.14. This parameter will affect hazard and survival functions, so it has an influence on the characteristics of the failure model. Since, there is no outliers contained in the box-plot of E-I/O card and the simulation results for Weibull distribution without outliers, we used LMOM estimator which is the most effective method for parameter estimation of Weibull distribution for data sets without outliers. Therefore, in calculation of the hazard and survival functions, the use of LMOM estimates will give more reliable results. According to the survival function calculated by using LMOM estimates, the survival probability after 180th day is 0.73. At the 210th day of the component, the probability of lifetime decreases to 0.578. Therefore, it should be recommended to check E-I/O card between 180th and 210th days and replace it if necessary.

SEC card parameter estimation results for LEV distribution are given in Table 5. There are only 14 failure data records for this component and there are two outliers as presented in the corresponding box-plot. In the case of both small sample sized and the existence of outliers, the robust *M* estimator provides effective results for the parameter estimation of LEV distribution. The survival probability after 180th day is 0.7 according to the survival function calculated using robust *M* estimator. The survival probability decreases to 0.5 on the 210th day. Therefore, it should be recommended that the SEC card needs to be checked and replaced if necessary between 180th and 210th days. If LS estimates are used for the SEC card, the probability of SEC card to run smoothly after the 270th day is 0.62. In other words, it should not be controlled for a longer period but it may fail during this period.

Therefore, robust *M* estimates reflect more accurate results to plan the maintenance and repair.

Electrans green signal lamp failure records best fit to the LEV probability distribution. However, in the box-plot, there exists two outliers in the failure data set of this component. That's why, robust *M* estimation method were used to estimate the distribution parameters. If the robust *M* estimation parameters presented in Table 5 is used to calculate the probability of survival for the lamp, the probability of the component to survive after the 150th day is calculated as 0.53. The probability drops to 0.31 on the 180th day. Therefore, it should be recommended to maintain Electrans green signal lamp after the 150th day.

SEV distribution is obtained as the probability distribution of balise antenna failure data. The estimation results based on 13 observations which are quite different due to the existence of two outliers that are verified by the box-plot. For this reason, the robust *M* estimations must be preferred in Table 5. Then, the survival probability of balise antenna after 150th day is calculated as 0.5. Therefore, it can be recommended to maintain it on the 150th day. If the same results are evaluated by using LS estimations, the survival probability of balise antenna is dropped to 0.5 after 300th days. This period is quite different and shows the effect of the parameter estimation method on estimates.

There are 14 failure records for RG 213/214 connector and the data does not contain outliers. Since the probability distribution method is determined as SEV, as presented in Table 4, the estimation method that gives the most effective parameter value is the LMOM. The survival probability that is calculated by corresponding estimations indicates that after the 90th day, the connector will survive with a probability of 0.658 and on the 120th day with a probability of 0.579. When the survival probability is close to 0.50, it would be appropriate to make the maintenance of this equipment.

There are 14 failure records for RG 400 connector which also include only one outlier as demonstrated in the corresponding box-plot. In Table 4, the probability distribution for this component has been found to be SEV. Because of being a small sample sized data and having an outlier, robust *M* estimation must be preferred in determinations of the parameter estimations. Survival probabilities for RG 400 connector has been calculated as 0.60 on the 150th day and 0.52 on the 180th day. Therefore, it is necessary to perform maintenance and repair activities on the 180th day.

6. CONCLUSION

In this study, we emphasized the parameter estimation of the random distortion model based on the failure data of some electronic components used in high speed train signaling system. High speed train signaling systems and similar large scale systems consist of hundreds of expensive components having a small number of failures in long time intervals. This results in small sized failure data which needs to be analyzed by using proper methods. We used the most popular distributions to model the lifetimes of the experimented components and observed that Weibull, LEV and SEV distributions has been determined to be the most appropriate distributions in modeling the failure data obtained from the high speed train lines in Turkey.

Parameter estimation is very important in failure models. In practice, some components may fail in very short periods of time or some parts may last much longer than their average lifetimes. These situations highly affect the estimation of model parameters. For this reason, we discussed 5 different methods in the study for parameter estimation of the random distortion model. The effectiveness of the parameter estimators of the Weibull, LEV and SEV distributions was compared by performing a Monte Carlo simulation for the small sample case with and without outliers. According to the simulation results, if there is no outlier in the data set for parameter estimation of the 3 distributions, LMOM estimator was determined as the most effective parameter estimator. When the sample size increases ($n = 50$), it is recommended to use ML and robust M estimators. In case of outliers in the data set, the most effective estimator is the robust M method for all samples considered.

The 5 components discussed in the study are Electrans Green Signal Lamp, Electronic Input/Output Card, Signal Electronic Control Card, Balise Transmission Module, SRG213/214 and RG400 connectors. They are chosen because they cause sudden breakdowns more frequently than other components and can cause train trips to be delayed due to immediate repair needs when broken. The distribution that the failure data of each individual component was determined by the Kolmogorov-Simirnov test. Since there are no outliers in two of these components (E-I/O and RG 213/214), the LMOM estimator is used in parameter estimation. With the help of these estimates, the probability of the recommended time for maintenance and repair was determined. There are outliers in the data sets of other components. The predictive results of robust M estimator were used for the parameter estimation of the random deterioration model for these components.

As a result, for the cases that the data set has outliers, we found the robust *M* estimator performs the best, for the cases that the data set has no outliers, LMOM estimator performs the best in lifetime probability estimations of the components. Large systems such as high speed railways contain hundreds of signaling system components. The results of this study would be inspiring to other large scale systems having small sample sized failure data as well.

ETHICAL DECLARATION

In the writing process of the study titled "Robust Lifetime Estimation with Small Sample Sized High Speed Railway ETCS Component Data", there were followed the scientific, ethical and the citation rules; was not made any falsification on the collected data and this study was not sent to any other academic media for evaluation.

REFERENCES

Abdul-Moniem, I. and Selim, Y. M. (2009), Tl-moments and L-moments estimation for the generalized pareto distribution, *Applied Mathematical Sciences*, 3(1):43-52.

Antoni, M. (2009), The ageing of signalling equipment and the impact on maintenance strategy. *In 2009 International Conference on Computers & Industrial Engineering*. IEEE.

Arslan, T., Bulut, Y. M. and Yavuz, A. A. (2014), Comparative study of numerical methods for determining weibull parameters for wind energy potential, *Renewable and Sustainable Energy Reviews,* 40, 820-825.

Bain, L. J. and Engelhardt, M. (1991), *Statistical analysis of reliability and life-testing models: Theory and methods*. Number vol. 115 in Statistics, textbooks and monographs. M. Dekker, New York, 2nd ed edition.

Barbosa, J. F., Carlos Silverio Freire Ju´nior, R., Correia, J. A., De Jesus, A. M. and Cal¸cada, R. (2018), Analysis of the fatigue life estimators of the materials using small samples, *The Journal of Strain Analysis for Engineering Design*, 53(8):699-710.

Bartolucci, A. A., Singh, K. P., Bartolucci, A. D. and Bae, S. (1999), Applying medical survival data to esti- mate the three-parameter weibull distribution by the method of probability-weighted moments. *Mathematics and computers in simulation*, 48(4-6), 385-392.

Bemment, S. D., Goodall, R. M., Dixon, R. and Ward, C. P. (2017), Improving the reliability and availability of railway track switching by analysing historical failure data and introducing functionally redundant sub- systems. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 232(5), 1407-1424.

Casella, G. and Berger, R. L. (2001), *Statistical Inference,* Thomson Learning.

Cox, D. R. and Oakes, D. (2018), *Analysis of survival data.* Chapman and Hall/CRC.

G´omez, M., Corral, E., Castejon, C. and Garc´ıa-Prada, J. (2018), Effective crack detection in railway axles using vibration signals and WPT energy. *Sensors*, 18(5), 1603.

Hall, A. R. (2005), *Generalized method of moments. Advanced texts in Econometrics,* Oxford University Press, Oxford; New York. OCLC.

Hosking, J. R. (1990), L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1), 105-124.

Hosking, J. R. (2007), Distributions with maximum entropy subject to constraints on their l-moments or expected order statistics, *Journal of statistical planning and inference*, 137(9), 2870-2891.

Huber, P. (2009), *Robust statistics,* Wiley, Hoboken, N.J.

Huber-Carol, C., Balakrishnan, N., Nikulin, M. and Mesbah, M. (2012), *Goodness-of-fit tests and model validity*, Springer Science & Business Media.

Jia, L., Wang, L.,and Qin, Y. (2021), Hi*gh-speed railway transportation organization status. In High-Speed Railway Operation Under Emergent Conditions,* 1-29, Springer Berlin Heidelberg.

Lv, S., Niu, Z., He, Z. and Qu, L. (2015), *Estimation of lower percentiles under a weibull distribution*, In 2015 First International Conference on Reliability Systems Engineering (ICRSE). IEEE.

McCool, J. I. (2012), *Using the Weibull Distribution: Reliability, Modeling, and Inference,* Wiley.

Moeini, A., Jenab, K., Mohammadi, M. and Foumani, M. (2013), Fitting the three-parameter weibull distribution with cross entropy, *Applied Mathematical Modelling*, 37(9), 6354-6363.

Mokhtarian, P., Namzi-Rad, M.-R., Ho, T. K. and Suesse, T. (2013), *Bayesian nonparametric reliability analysis for a railway system at component level*. In 2013 IEEE International Conference on Intelligent Rail Transportation Proceedings. IEEE.

Shangguan, W., Zang, Y., Wang, H., and Pecht, M. G. (2020), Board-level lifetime prediction for power board of balise transmission module in high-speed railways, *IEEE Access*, 8, 135011-135024.

Sirvanci, M. and Yang, G. (1984), Estimation of the weibull parameters under type i censoring, *Journal of the American Statistical Association*, 79(385), 183-187.

Stanley, P., Hagelin, G., Heijnen, F., Lofstedt, K., Por´e, J., Suwe, K.-H. and Zoetardt, P. (2011), *ETCS for Engineers*, Eurail press.

Tsionas, E. G. (2003), Bayesian quantile inference, *Journal of statistical computation and simulation*, 73(9), 659-674.

Wang, D., Hutson, A. D. and Miecznikowski, J. C. (2010), L-moment estimation for parametric survival models given censored data, *Statistical Methodology*, 7(6):655-667.

Wang, L., Xu, Y. and Zhang, J. (2008), Research on reliability analysis model for key components and parts of railway equipment and its application, *Journal of the China Railway Society*, 30(4), 93-97.

Yang, J.-W., Wang, J.-H., Huang, Q. and Zhou, M. (2018), Reliability assessment for the solenoid valve of a high-speed train braking system under small sample size, *Chinese Journal of Mechanical Engineering*, 31(1).

Zhang, D., Long, Z., Xue, S. and Zhang, J. (2012), Optimal design of the absolute positioning sensor for a high-speed maglev train and research on its fault diagnosis, *Sensors*, 12(8), 10621-10638.

Zheng, P., Quan, S. and Chu, W. (2021), Analysis of market competitiveness of container railway transportation, *Journal of Advanced Transportation*, 1-8.

Zhu, D. and Liu, H. (2013), Reliability evaluation of high-speed train bearing with minimum sample, *Journal of Central South University (Science and Technology)*, 44(3), 963–969