Warranty Claims Prediction with a Combined Model of Market Absorption and Failure Process

Davorin Kofjač[1], Andrej Škraba[1], Aleš Brglez[2]
[1]University of Maribor, Faculty of Organizational Sciences, Kidričeva cesta 55a, Kranj, Slovenia
[2]Gorenje gospodinjski aparati d.d., Partizanska cesta 12, Velenje, Slovenia
{davorin.kofjac, andrej.skraba}@fov.uni-mb.si, ales.brglez@gorenje.si

Abstract
The paper addresses an important part of enterprise activity, which is the warranty claims control and prediction. Warranty claims have a significant impact on the financial aspect of the company, because of the funds that need to be reserved for repair of the defected products within the warranty period. The failure process is considered as a combination of market absorption and failure process. Prediction and determination of terminal call rate (TCR) is important for monitoring of the production system efficiency. Several models to predict TCR have been analyzed, such as Weibull model and the Markov Modulated Fluid model (MMFM). The models were validated against two types of data: a) from sales to failure, and b) from production to failure. The results show that MMFM model provides promising prediction results.
Keywords: warranty claims prediction, terminall call rate, weibull model, markov modulated fluid model

1 Introduction

Warranty claims control and its prediction are an important part of enterprise activity, because all products are unreliable and they eventually fail. Most products are sold with a warranty that offers protection to buyers against early failures over the warranty period. A warranty is a contract between buyer and manufacturer that becomes effective on the sale of an item. The purpose of a warranty is basically to establish liability in the event of a premature failure of an item, where failure is meant as the inability of the item to perform its intended function (Kim et al., 2004).

Offering warranty implies additional costs to the manufacturer. This is the cost of repairing item failures (through corrective maintenance) over the warranty period. Hence, the manufacturer tends to minimize this cost. Minimizing the warranty cost is not in the scope of this paper directly and the reader is referred to, for example, Blischke and Murthy (1994), Eliashberg et al. (1997), and Zuo et al. (2000) for warranty cost evaluation. However, anticipating the share of products that will eventually fail during the warranty period is crucial in determining this cost.

Taking into account financial and marketing considerations, a multitude of business decisions are being made based on the predicted number of warranty claims for the predetermined warranty period. Therefore it is important to improve the process of warranty claims prediction with models that provide an acceptable accuracy for business decision making. In addition, a parallel need for warranty claims prediction in industry also arises when the first few months of warranty claims are being analyzed for the purpose of forward extrapolation and development of appropriate corrective actions (Kleyner and Sandborn, 2005).

Many quality and reliability engineers who are involved in the warranty claims predictions use empirical models based on past data of products with similar design and complexity adjusted by certain, experience-based correction factors. Therefore, we stress that an accurate, scientific based, and user-friendly model could help to accomplish warranty prediction tasks with better precision this enhancing the process of business decision making (Kljajić et al., 2000, Škraba et al., 2011).

This paper is dealing with anticipation of the number of products failing within the warranty period with the proposed MMFM method (Rogers, 1994). Further, the anticipation of parameters for these methods is presented as well as a framework for the implementation of a decision support system to help with warranty claims predictions. To produce a quality estimate a large number of data is needed. To process such a large amount of data, however, requires substantial processing power. Hence, grid computing is also addressed as a possibility to solve this issue.

The rest of the paper is organized as follows. In Section 2 the problem formulation is given. Section 3 provides the results of model validation. In Section 4 the framework for the implementation of a decision support system for warranty claims prediction is discussed. Finally, conclusions are drawn in Section 5.

2 Problem Formulation

Figure 1 shows the temporal aspect of the failure process of the products. Production is followed by a failure of products after a certain period of time. An item failure can occur early in its life due to manufacturing defects or late in its life due to degradation of the item. The degradation is dependent on age and usage (Kim et al., 2004).

In this paper we are dealing with a forecasting activity for current products, where the warranty claims are known for the first few months of service and the objective is to anticipate the final numbers of warranty returns at the end of the warranty cycle.

The products are produced in a series which in our case is a monthly aggregate of a production of a certain product or a product family. If the failed products can be traced back to a specific production series, this data can be converted into a more comprehensive format usually referred as 'layer cake', which combines all produced and failed units on a monthly basis, as presented in the Table 1. The acquired data are aggregated in the form of an upper triangular matrix with diagonal. The table shows that, for example, after the first four months of production we can collect 10 data points about the failures that can be used as an input for the prediction.

The data presented in a layer cake format allows more sophisticated data processing, because we are able to obtain exact failure time intervals and the number of failed items. This allows for the implementation of distribution best-fit approaches and provides the confidence intervals on the results of the best-fit approximation. (Kleyner and Sandborn, 2005).

Table 1: Example of a layer cake data.

Month	Number of produced products	Number of failures by month			
		Month 1	Month 2	Month 3	Month 4
1	4553	133	56	34	20
2	4224		125	76	21
3	8521			186	110
4	9661				224

2.1 Terminal Call Rate

Terminal call rate (TCR) is an expected maximum value of the failed products within the warranty period. Fig. 1 shows an example of quarterly gathered failures data. There are three curves within each quarter, each curve representing a cumulative failure rate for a specific monthly production series. Such a representation might indicate difficulties in the production process or perhaps a problematic semi-product that is integrated in the finished product, if a certain series' failure rate is above the others.

Hence, determination of TCR is important for monitoring of the production system. On the other hand, determination of TCR also covers the financial aspect. Its determination influences the evaluation of funds that are to be reserved for maintenance during the warranty period.

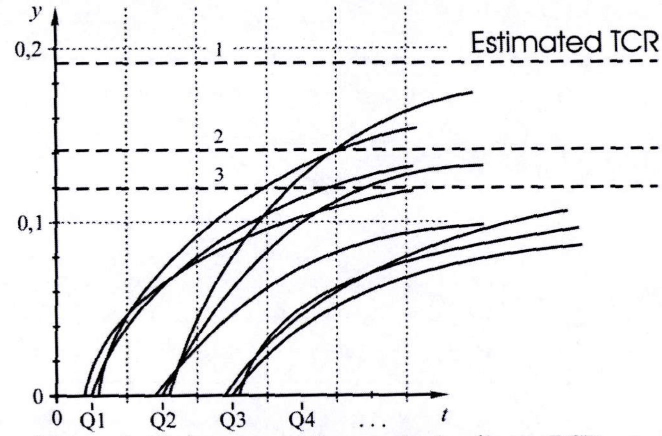


Figure 1: Estimation of the terminal call rate TCR.

2.2 Failure process

Mostly, realiability is expressed in terms of commonly used distributions: Weibull, Exponential, Normal, and Lognormal (Xie and Lai, 1995, Hall and Strut, 2003, Kleyner and Sandborn, 2005). In our case there are two types of data available for the warranty claims prediction, depending on the type of products: a:) only the time of sale is known while the production time is unknown *I(sales)*, and b) only the time of production is known while the time of sale is unknown *I(production)*. Hence, two types of data collections result in two different probability density functions (see Fig. 2) and its resulting cumulative density functions (see Fig. 3).

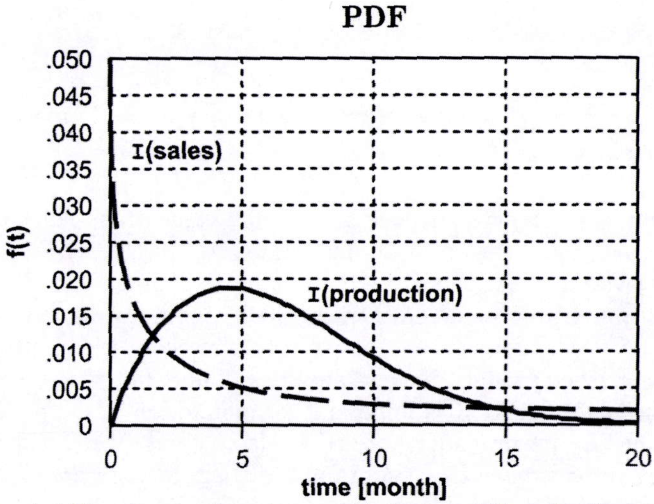


Figure 2: Probability density functions of I(sales) and I(production) type of data.

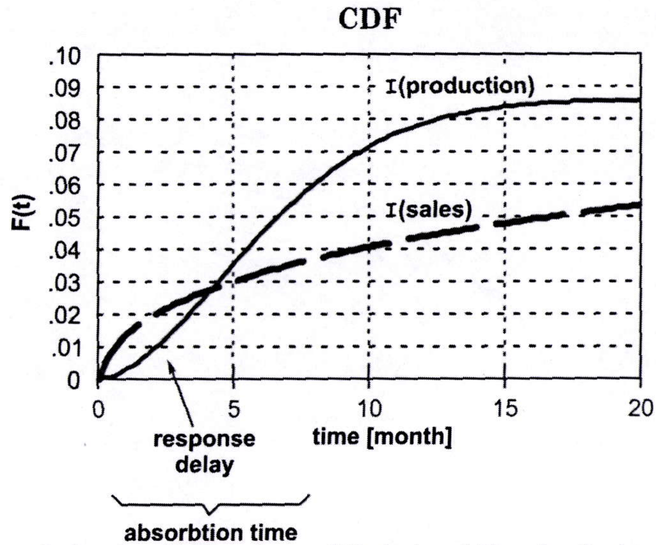


Figure 3: Cumulative density functions of I(sales) and I(production) type of data.

Obviously, the *I(sales)* is following the exponential distribution while the *I(production)* is following the Webull distribution. From the figures one can observe that *I(production)* function has an additional part of the curve that is characterized by the market absorbtion time, i.e. the time from production to the point of sale. Hence, we can characterize the *I(sales)* function as the failure process without delay and the *I(production)* function as the failure process with delay. Since the point of sale for products, characterized by the *I(production)* function, is not known, the parameters have to be anticipated from the currently available data.

To anticipate the CDF curves for both data types, we propose the Markov Modulated Fluid Model (see, e.g. Lenin and Parthasarathy, 2000) given by the following equation:

$$F(t) = \frac{at^b}{c + t^b} \tag{1}$$

where a represents the limit value of the process, i. e. $a = \lim_{t\to\infty} F(t)$, while b and c represent the shape of the distribution.

Figure 4 shows an example of F(t) according to the equation (1) and variation of parameter a.

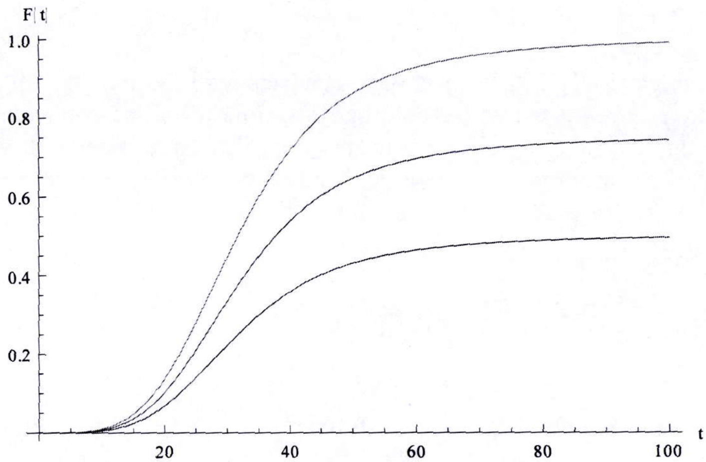


Figure 4: F(t) with the variation of parameter a.

First, we need to determine a way to calculate the parameters a, b and c.
Let

$$z = 1 - \frac{F(x)}{a} \tag{2}$$

Apply natural logarithm *ln* and rewrite:

$$c = e^{b \ln x - \ln\left(\frac{1}{z} - 1\right)} \tag{3}$$

Now the following terms determine the parameters of the model:

$$a = Fx^{-b}(c + x^b)$$

$$b = -\frac{\ln\left(-\frac{a\left(-1+\frac{F}{a}\right)}{cF}\right)}{\ln x} \tag{4}$$

$$c = \frac{(a-F)x^b}{F}$$

Now we have obtained the formulae needed for the parameters' estimation. When analyzing data, we need to select a method for estimating the parameters for the chosen distribution. There are several methods available, such as rank regression on X (RRX) or Y (RRY) or maximum likelihood estimation (MLE) (Tomblin and Seneviratne, 2011).

Regression generally works best for data sets with smaller sample sizes. Rank regression on Y is best used with data other than time-to-failure data, such as free-form data. An example of this would be warranty data that have unreliability estimates for each month of a warranty period (Reliability Hotwire, 2007). Hence, we have chosen RRY method for parameter determination that can be generally described by the following equation:

$$\sum_{i=1}^{N}(\hat{a} + \hat{b}x_i - y_i)^2 = \min(a,b)\sum_{i=1}^{N}(a + bx_i - y_i)^2 \tag{5}$$

Where $\hat{a}$ and $\hat{b}$ are the least square estimates of a and b. This equation is minimized with the following values of $\hat{a}$ and $\hat{b}$:

$$\hat{a} = \frac{\sum_{i=1}^{N} y_i}{N} - \hat{b}\frac{\sum_{i=1}^{N} x_i}{N} = \bar{y} - \hat{b}\bar{x}$$

$$\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i - \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N}}{\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}} \tag{6}$$

where N is the number of (x_i, y_i) data coordinates. One of the advantages of the rank regression method is that it can provide a good measure for the fit of the curve to the data points with the he correlation coefficient ρ, generally expressed as:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{7}$$

where σ_{xy} is the covariance of x and y, σ_x is the standard deviation of x, and σ_y is the standard deviation of y. The estimate of the correlation coefficient is then given by:

$$\hat{\rho} = \frac{\sum_{i=1}^{N} x_i y_i - \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N}}{\sqrt{\left(\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}\right)\left(\sum_{i=1}^{N} y_i^2 - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N}\right)}} \tag{8}$$

3 Model Validation

It is vital to perform model validation for evaluating if the model was developed for the purpose it was intended. To obtain a valid model, we need to measure the inputs and outputs of the real system and compare them against the variables of the model get as accurate model response as possible (Kleijnen, 1995). Validation of the model has a particular significance in the quality control,where the following issues are important:

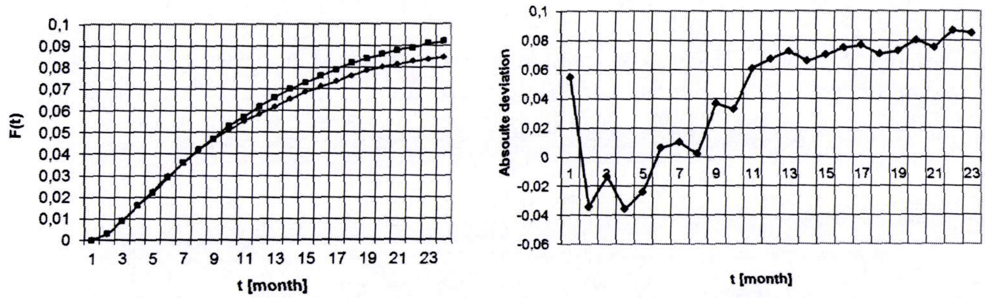
- Availability of the failures data,
- Quality of data,
- Numerus of a particular production series,
- Large datasets makes this process time consuming.

Validation is performed on the input matrix of dimensions 18x18 (upper triangle matrix with diagonal, 171 data points). Thus, the prediction of the cumulative proportion of each product is performed on historical data for the past 18 months at the time or 24 months that represents the end of the warranty period. We present two validation cases: a) the one with all historical data included in prediction, and b) the one with production series that have had 6 or more failures.

Table 2 and Fig. 5 present model validation results on a product with all historical data included in prediction. The historical TCR curve is marked with a diamond shape while the predicted is marked with a square. The historical TCR at the end of 24 months is 0,085 and the predicted TCR is 0,092. The prediction accuracy in this case is 8,2%.

Table 2: Comparison of historical and predicted TCR with all historical data included.

Historical value	0,085
Predicted value at 24m	0,092
Absolute deviation	0,007
Deviation [%]	0,082

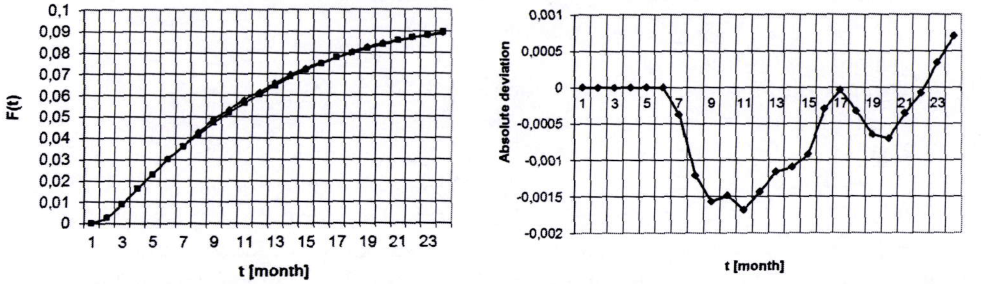


a) b)

Figure 5: Comparison of: a) historical (diamond) and predicted TCR (square) with all historical data included, and b) absolute deviation between historical and predicted data.

Table 3: Comparison of historical and predicted TCR with product series with at least 6 failures.

Historical value	0,0887
Predicted value at 24m	0,0894
Absolute deviation	0,0007
Deviation [%]	0,0079



a) b)

Figure 6: Comparison of: a) historical (diamond) and predicted TCR (square) with product series with at least 6 failures, and b) absolute deviation between historical and predicted data.

Table 3 and Fig. 6 present model validation results on a product with product series with 6 or more failures used for prediction. The historical TCR curve is marked with a diamond shape while the predicted is marked with a square. The historical TCR at the end of 24 months is 0,0887 and the predicted TCR is 0,0894. The prediction accuracy in this case is 0,79%. Obviously, the prediction accuracy is much higher than in previous case due to the quality data achieved by excluding outliers that have negatively influenced the prediction.

4 Application Development and Implementation

So far, we have presented the metholodogical framework for warranty claims prediction. Since the model has given valid results, we were able to integrate it in a decision support system presented in Fig. 7. The requirements for such a system were:

- Application of both data types (sales~failure and production~failure),
- Web application with possibility of concurrent clients,
- Application of existing production database,
- Tabelaric report with absolute and relative failure frequencies,
- Application of color codes of historic/combined/predicted data,
- Aggregation of the results on a quarterly and yearly basis,
- Graphs of failure functions,
- Visual validation of results.

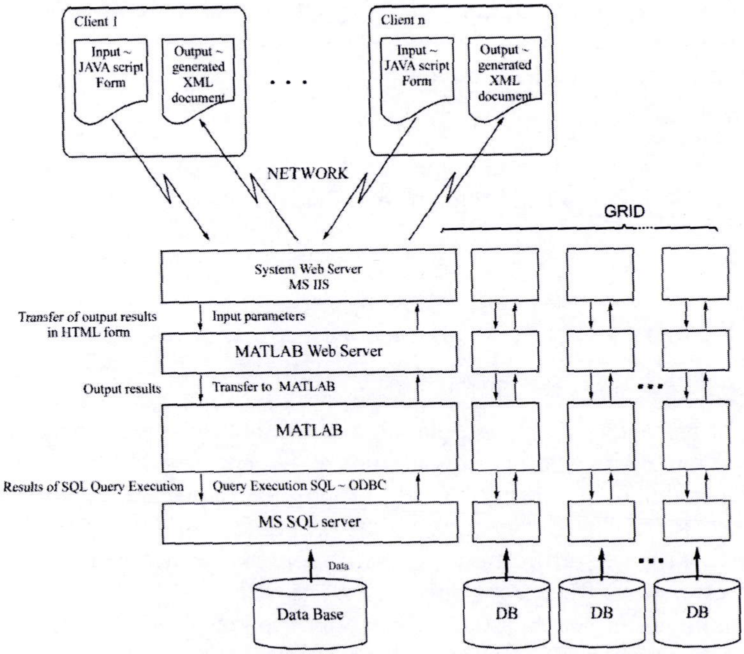


Figure 7: The framework for warranty claims prediction decision support system.

The technological platform used to implement the information system was: Microsoft SQL Server to store the data, MATLAB as a computational core, Microsoft IIS Server and MATLAB Web Server for website hosting and user interface.

Because the prediction process involves a large amount of data it is computationaly costly. Hence, grid computing is a preferable option to solve this problem, because the MATLAB platform itself supports such computation through distributing jobs to a network of computers.

5 Conclusion

This paper is dealing with the warranty claims prediction for two types of data: a) from sales to failure, and b) from production to failure, where the market absorbtion time is considered. We have proposed the Markov Modulated Fluid Model for warranty claims prediction. The model was verified and validated against the real world data, yielding quality prediction results, especially in a case, where the data outliers have been removed. In this case the prediction accuracy is less than 1% for a selected product.

Further, we have presented the framework for a warranty claims prediction decision support system that integrates a database with a large amount of data, computational core and a web user interface with a possibilty of concurrent clients. We also propose an application of grid computing because of computationaly costly prediction.

The presented decision support system has a clear industrial application, since the prediction of failures is an important part of the production process. Not only it might indicate the possible problems in the production process, it also gives a foundation for estimation of preventive maintenance costs.

Acknowledgements

The research was funded by Gorenje d.d. and Slovenian Research Agency ARRS, (Contract No.: MK-2306-03 and Program No. UNI-MB-0586-P5-0018).

References

Blischke W.R., Murthy D.N.P. (1994). Warranty cost analysis. Marcel Dekker: New York.

Hall P., Strutt J. (2003). Probabilistic physics-of-failure models for component reliability using Monte Carlo simulation and Weibull analysis: aparametric study. Reliability Engineering and System Safety, Vol. 80, pp. 233–242.

Kim C.S., Djamaludin I., Murthy D.N.P., (2004). Warranty and discrete preventive maintenance, Reliability Engineering and System Safety 84 (2004) 301–309.

Kleijnen J.P.C. (1995). Verification and validation of simulation models. European Journal of Operational Research, Vol. 82, pp. 145-162.

Kljajić M., Bernik I., Škraba A. (2000). Simulation Approach to Decision Assesment in Enterprises. Simulation, 75 (4), Simulation Councils Inc., pg:199–210.

Kleyner A., Sandborn P. (2005). A warranty forecasting model based on piecewise statistical distributions and stochastic simulation. Reliability Engineering and System Safety, Vol. 88, pp. 207–214.

Lenin R.B., Parthasarathy P.R. (2000). Fluid Queues Driven by an M/M/1/N Queue. Mathematical Problems in Engineering, Vol. 6, pp. 439-460.

Rogers L.C.G. (1994). Fluid models in queueing theory and Wiener-Hopf factorization of Markov chains. Annals of Applied Probability, Vol. 4, No. 2, pp. 390-413.

Reliability Hotwire (2007). When should rank regression or maximum likelihood estimation (MLE) be used when conducting life data analysis in Weibull++? Reliability Hotwire – The Magazine for the Reliability Professional, No. 80, http://www.weibull.com/hotwire/issue80/tooltips80.htm, accessed 15.1.2013.

Škraba A., Kljajić M., Papler P., Kofjač D., Obed M. (2011). Determination of recruitment and transition strategies. Kybernetes, vol. 40, iss. 9/10, str. 1503-1522, doi: 10.1108/03684921111169512.

Tomblin J., Seneviratne W. (2011). Determining the Fatigue Life of Composite Aircraft Structures Using Life and Load-Enhancement Factors. Final report, Air Traffic Organization, Washington DC, USA.

Xie M., Lai C.D. (1995). Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. Reliability Engineering and System Safety, Vol. 52, pp. 87–93.

Zuo M.J., Liu B., Murthy D.N.P. (2000). Replacement-repair policy for multistate deteriorating products under warranty. European Journal of Operational Research, Vol. 123, No. 3, pp. 519–530.