An Investigation of Estimation Performance for a Multivariate Poisson-Gamma Model with Parameter Dependency

Rafael Schwarzenegger

Supervisors:

Prof. Lesley Walls & Prof. John Quigley

A thesis presented in fulfilment of the requirements for the degree of:

Doctor of Philosophy (PhD)

University of Strathclyde, Depatment of Management Science

Glasgow, Scotland, UK - 2023

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date:

Signed Statement

This statement is intended clearly to define the extent of my contribution to previously published work for which I have been jointly responsible.

Following contributions have resulted from the work included in the thesis. I developed the ideas and papers under the able guidance of my supervisors (Professor Lesley Walls and Professor John Quigley). The contribution of other academics is limited to providing feedback on the draft version of the manuscripts and facilitating empirical research.

Chapter 3:

- Schwarzenegger, R, Quigley, J & Walls, L 2020, Evaluation of empirical Bayes estimators of correlated event rates and implications for predicting new product reliability. in P Baraldi, F Di Maio & E Zio (eds), *E-proceedings of the 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference: ESREL2020 PSAM15.* Singapore, pp. 3022, The 30th European Safety and Reliability Conference and The 15th Probabilistic Safety Assessment and Management Conference, Venice, Italy, 1/11/20. doi: 10.3850/978-981-14-8593-0_4439-cd

- Schwarzenegger, R, Quigley, J & Walls, L 2021, Is eliciting dependency worth the effort? A study for the multivariate Poisson-Gamma probability model, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability.* doi: 10.1177/1748006X211059417.

Abstract

Statistical analysis can be overly reliant on naive assumptions of independence between different data generating processes. This results in having greater uncertainty when estimating underlying characteristics of processes as dependency creates an opportunity to boost sample size by incorporating more data into the analysis. However, this assumes that dependency has been appropriately specified, as mis-specified dependency can provide misleading information from the data. The main aim of this research is to investigate the impact of incorporating dependency into the data analysis.

Our motivation for this work is concerned with estimating the reliability of items and as such we have restricted our investigation to study homogeneous Poisson processes (HPP), which can be used to model the rate of occurrence of events such as failures. In an HPP, dependency between rates can occur for numerous reasons. Whether it is similarity in mechanical designs, failure occurrence due to a common management culture or comparable failure count across machines for same failure modes. Multiple types of dependencies are considered. Dependencies can take different forms, such as simple linear dependency measured through the Pearson correlation, rank dependencies which capture non-linear dependencies and tail dependencies where the strength of the dependency may be stronger in extreme events as compared to more moderate one. The estimation of the measure of dependency between correlated processes can be challenging.

We develop the research grounded in a Bayes or empirical Bayes inferential framework, where uncertainty in the actual rate of occurrence of a process is modelled with a prior probability distribution. We consider prior distributions to belong to the Gamma distribution given its flexibility and mathematical association with the Poisson process.

For dependency modelling between processes we consider copulas which are a convenient and flexible way of capturing a variety of different dependency characteristics between distributions. We use a multivariate Poisson – Gamma probability model. The Poisson process captures aleatory uncertainty, the inherent variability in the data. Whereas the Gamma prior describes the epistemic uncertainty. By pooling processes with correlated underlying mean rate we are able to incorporate data from these processes into the inferential process and reduce the estimation error.

There are three key research themes investigated in this thesis.

First, to investigate the value in reducing estimation error by incorporating dependency within the analysis via theoretical analysis and simulation experiments. We show that correctly accounting for dependency can significantly reduce the estimation error. The findings should inform analysts a priori as to whether it is worth pursuing a more complex analysis for which the dependency parameter needs to be elicited.

Second, to examine the consequences of mis-specifying the degree and form of dependency through controlled simulation experiments. We show the relative robustness of different ways of modelling the dependency using copula and Bayesian methods. The findings should inform analysts about the sensitivity of modelling choices.

Third, to show how we can operationalise different methods for representing dependency through an industry case study. We show the consequences for a simple decision problem associated with the provision of spare parts to maintain operation of the industry process when dependency between event rates of the machines is appropriately modelled rather than being treated as independent processes.

To my father, Friedrich Schwarzenegger,

and

grandfather, Prof. Dr. Ivan Šantavý, CSc.!

Acknowledgements

I would like to acknowledge and thank my supervisors Professor Lesley Walls and Professor John Quigley. Their guidance and support have sustained me on this journey. My progress would not be possible without their guidance and supervision. Thank you for your patience, guidance, trust, input and feedback. I am honoured to have been on this very special journey.

I would like to thank my external international advisor plk. gšt. Prof. David Vališ for his valuable support, experience, help and useful discussions.

A special acknowledgement goes to my mother, whose relentless support, words of encouragement, sacrifice and love made my completion possible.

I would like to express my gratitude for receiving funding from the University scholarship with additional contribution from Irvin-GQ. For my last chapter, I would like to thank Siemens Analyse for providing me with industrial data accompanied with support.

List of Figures

List of Tables

Contents

List of Symbols

α	Shape parameter of gamma prior distribution.
β	Scale parameter of gamma prior distribution.
$[\widehat{\alpha}, \widehat{\beta}]$	Estimates of parameters of the gamma prior distribution by the method of moments, likelihood estimation or expert judgment.
t_1	Exposure time of first variable.
K	Total number of failure cause.
k	Index for a specific failure cause.
E	Expectation function.
$C_\theta(u_1, v_1)$	Copula function with dependency parameter θ of variables u_1 and v_1.
c	Copula density function.
MSE	Mean squared error.
$MSPE$	Mean squared percentage error.
$bias$	The bias as the difference of the estimated parameter value $\hat{\theta}$ from the true parameter value θ as $\hat{\theta} - \theta$.
$\%bias$	Percentage bias increase.
λ_1	First gamma random variable.
n_i	The i-th element of random vector of Poisson count data.
$\underset{\sim}{N}$	Vector of realizations of Poisson count data.

N_1	Number of faults for first variable originating from a Poisson distribution.
N_{ik}	Number of faults of machine i within failure cause k.
$\Gamma(\alpha, \beta; \rho)$	Gamma bivariate probability density function of specified by α, β and correlation parameter ρ.
$G(\lambda_1; \alpha, \beta; \rho)$	Marginal density of gamma bivariate cumulative density function of specified by λ_1, α, β and correlation parameter ρ.
$\pi(\lambda)$	Probability density function of the variable λ.
$\mathbb{1}_{k\times 1}$	Unit vector of size $(k \times 1)$.
ρ_A	Assumed linear correlation.
ρ_T	True underlying linear correlation.
ρ_S	Spearman's rank correlation.
q	Percentile
p	Percentile value.
cop_A, *cop_T*	Assumed Gaussian copula *cop_A* and true underlying copula *cop_T* notation. We appoint on the place *cop* one of the four used copulas, the Gaussian, Frank, Gumbel and Clayton copula.

Chapter 1

Introduction

The research is motivated by the need to address risk and reliability decision analysis. The management of arising challenges with the use of the Poisson-Gamma probability models is in central focus. The Poisson-Gamma probability model is a popular choice for such analysis due to its flexibility and ability to handle count data with excess zeros. Research by Walls and Quigley (1999; 2011) aims to address the challenges in risk and reliability analysis using Poisson-Gamma probability models, specifically in the context of new product development and risk mitigation decision support. The model encompassed interdependencies within a one-shot system in the stage of developers advancing to the target engineering design. The new product is an updated version of an existing design line, and the development team relied on data from previous design versions, along with new test data, to inform certain aspects of the updated system.

A setting with the features of high-consequence, low-frequency events is undertaken by Quigley et al. (2007). They aimed at developing empirical Bayes framework suitable for the investigated and distinctive features modelled by subjecting the inputs to a Poisson-Gamma model. The method suited the setting of estimating the frequency for train derailments within the UK. The authors are targeting in their objectives the reduction of risk to a reasonable possible level.

A new inference framework found in the Bayes linear Bayes model is expressed by Quigley et al. (2013). The underpinning presumptions count with a homogeneous Poisson process. The analysis is positioned in comparison to a full Bayesian model with a multivariate gamma prior. The paper used an application on correlated occurrence rates in supply chain for various users displaying how neglecting dependencies results in biased estimates. It is shown that the new subjective inference method was more accurate than the traditional approach that ignored the correlation between event rates.

Over the last 20 years (i.e. 2002–2022) a semi-structured literature review shows that the Poisson-Gamma probability model has been adopted in different aspects of reliability and risk analysis. For example, this model has been used in various applications, new estimation strategies have been developed and, for the multivariate case consideration has been given to the way in which dependency between model parameters has been approached. A summary follows.

In the area of reliability, the papers often extend the applications to other fields. Bourezaz et al. (2020) examine the metric "time to interrupt" in reliability to represent the monthly failure number in supercomputer components. The authors note that the analysis can be used likewise for survival analysis with taking the metric "time to death". Chen et al. (2019) use the Poisson-Gamma model within a shock model for reliability engineering with a complementary application on insurance claims. It is examined under which conditions the number of shocks will cause the system to fail, exceeding a threshold.

Risk related papers are often modelling zero-inflated problems, addressing risks of missing out on payments, security breaches and claims in insurance. Which means that there are more zeros observed than we would expect. Such excess zeros are causing biased results in statistical models. A way of adjusting the modelling strategy and obtain more accurate estimates to this condition is to use a zero-inflated model. There are risks of bank failure, risk for agricultural lenders, risk connected to bond pricing and traffic risks of car crashes. While building models, there can be focus on specific modelling parts like tails and extreme events. See (Chen and Leong, 2022; Diers et al., 2013; Gerhold et al., 2010; Hemrit and Ben Arab, 2012; Mouatassim, 2012; Wang et al., 2015).

Homogeneous Poisson-Gamma models are used furthermore in other areas. Within spare parts management authors predict future spare parts repairs, give risk-based ranking to important spares and make use of zero-inflated unrepresented failure alarms (Percy, 2004; Hassan et al., 2012) and including ageing (Allella et al., 2002). In the area of food and nutrition are HPP used to describe microbial counts and affecting food sampling strategies. Making comparisons between the use of the Poisson-gamma and Poisson-lognormal in the microbial food counts. Authors research on within-batch and between-batch variability characterized by comparably low concentrations of the investigated microbial counts. These concentrations suit to be modelled using Poisson-Gamma regression with random effects assuming that certain model parameters have a random effect on the model. Meaning they are then being treated as random variables. A deeper explanation gives (Green et al., 1960, p. 131): “When a sample exhausts the population, the corresponding variable is fixed; when the sample is a small (i.e. negligible) part of the population the corresponding variable is random.” That is especially convenient for larger amounts of zero counts in the data. Some papers are combining other methods like deep learning with the Poisson-Gamma model (Cabras, 2021; Gonzales-Barron and Butler, 2011; Gonzales-Barron et al., 2013). Tweedie distributions are a family of distributions that for the Tweedie power parameter $1 < p < 2$, where $p \in \mathbb{R}$ are occasionally named Poisson-Gamma distributions. Studies can be conducted as an extension of the univariate dispersion Tweedie family. The Poisson-Gamma model is applied within risk management in risk capital allocations. See (Abid et al., 2020; Furman and Landsman, 2010; Hasan and Dunn, 2011).

Poisson-Gamma is an accepted modelling family in risk and reliability analysis for certain types of problems, and although this thesis considers one model we are aware that there exist other models, including in the Poisson family even though we have chosen not to make them the subject of your research.

1.1 Modelling Dependency in Homogeneous Poisson Processes

Earlier studies (Quigley et al., 2007; 2013) that motivated this research were concerned with modelling the dependency between data generating processes; particularly for models that assumed underlying homogeneous Poisson processes (HPP) or Poisson-Gamma probability models when full Bayesian or empirical Bayes inference was to be accommodated. More generally, there has been wider consideration of modelling dependency in the context of stochastic processes in risk and reliability analysis. We first consider some of the wider issues associated with modelling dependency using probability models.

With the growing need of correctly representing underlying dependencies in models, parameters and rates, authors are developing advanced modelling frameworks withing the reliability context. Having the data independently generated, but not observing dependency in the underlying structure in their failure rates. In the context of rare events, many critical data aspects may be missing. Enriching the information pool with likewise data can show vital in supporting the modelling of the reliability of a new design during development by extending the amount of data used for estimation to obtain more accurate results. Described by (Quigley et al., 2013; Olatubosun and Zhang, 2018; Khosravi et al., 2018; Wang et al., 2020).

A way of introducing a dependency structure offers the usage of copulas. Copulas are useful in inducing structures into dependency. Work has been accomplished on topic of copula dependency to model multivariate prior distributions and apply them on count data (Caliendo et al., 2021; Park et al., 2021; Dimitriou, 2022), where copulas were present mainly as part of a larger formula modelling the failure rate.

Using copulas relies on selecting a suitable copula structure with an implication of risking a possible mis-specification, as the current state of art shows (Durrleman et al., 2000; Trivedi and Zimmer, 2007; Acar et al., 2019; Park et al., 2021).

A wide class of applications (Pörn, 1990; Bunea et al., 2005) in the risk and reliability context show that failures can be well approximated with a Poisson distribution. Two-stage Bayesian models are often introduced by giving failure rates a prior distribution in order to enhance the complexity of the Poisson element. This work is bounded to a homogeneous Poisson process (HPP). We consider the situation where we have a collection of HPP, each with its own underlying rate, so denoting the rate of the jth process with λ_j. We are considering in our case two parallel running processes λ_1 and λ_2. These rates observe correlation. λ_1 represents the examined rate and λ_2 a rate that is similar (correlated) to λ_1. So we have 2 rates between whom we observe dependency. Moreover, a structure encompassed with a copula function. The HPP events N_1, N_2 themselves are conditionally independent after taking account of underlying failure rates, as illustrated on Figure 1.1. The copula aspect is changing the degree of dependency for different percentiles of the marginal distribution of the prior distribution. Engineering applications in the area have been published by (Dutfoy and Lebrun, 2009; Achibi et al., 2012; Pan et al., 2018; Wang et al., 2020 Zhou et al., 2019). Applications of road crash data are seen in (Savolainen et al., 2011; Wali et al., 2018; Yasmin et al., 2018; Caliendo et al., 2021; Park et al., 2021).

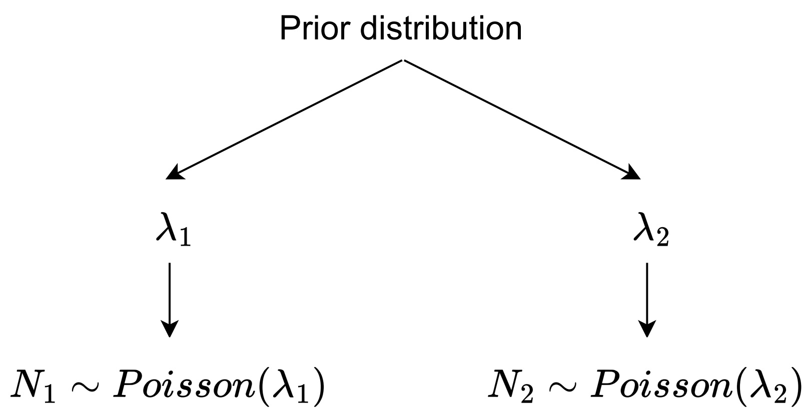


Figure 1.1: Diagram illustrating the dependency occurring between the rate parameters of the Poisson distribution in a hierarchical Bayesian model.

Dependency between parameters in the area of risk and reliability may arise due to various reasons. We may observe correlation within a model or moreover the probabilistic parameters. This can be due to the fact that systems may be subjected to the same environmental factors or show a similar inherent behaviour. Uncertainties can be affected by the same correlation source. A possible way of modelling these underlying dependencies in failure rates is with the use of multivariate prior distributions or by using copulas.

1.2 Value of Capturing Dependency in Probability Modelling

Accounting for dependency will possibly prove valuable as will be evidenced during this thesis. The value of dependence comes to light when compared with treating the data generating processes as independent. This contributes to properly understand the overall risk. In cases where complex relations and dependencies between variables, rates or data are present, assuming independence does not have full justification. Practically, it can be argued for example a diminishing effect of drugs (medicamentations) on bugs (illnesses) due to dependency in the work on rates of bug-drug resistance, pointing out systemic risk (Werner et al., 2017; 2021).

Savings can be quantified with considering dependency. 29% safety stock can be saved in (Song, 2002) who researches inventory multi-item systems backorders within the area of management science. The authors examine the influence of considering common components within service and inventory. This paper further scrutinizes the if constructing a common component will enhance inventory service. Nevertheless the achieved savings, it is acknowledged that an all-including reduction in cost can be hard to identify. Nevertheless, authors detail savings on safety stocks. A 27% in bias in the estimation of Value at Risk in multivariate portfolios states Fantazzini (2009) in finding that when mis-specifying a Student's t-copula with a Normal copula, the bias increases. If negative dependencies are present, the bias is much larger.

Notably, 6%, most commonly 3% is the underestimation in expected revenue from the actual revenue a retailer observes in a multi-period clearance pricing problem by Yahav and Shmueli (2012) with considering dependency for multivariate Poisson counts with correlation seemingly between occurrence rates. Authors are considering the multivariate Normal, then the univariate Poisson and lastly the multivariate Poisson distributions.

Eliciting dependency is challenging. Whether it is eliciting linear correlation coefficient or rank correlation coefficients between multivariate quantities or tail probabilities, conditional quantities and concordance measures. When experts face decision problems under uncertainty, they are interested in reducing uncertainty as much as possible. To incorporate as much related data as possible in an appropriate manner will help to achieve better estimates. Some processes as similar to each other, so the more similar two processes are then the more representative their data are of each other. The degree of similarity is in this sense the degree of dependency. Assuming independence is not sensible. Therefore we need to take into account dependence. Historical data to access dependencies can be unavailable or too expensive to acquire. Structured expert judgment therefore sometimes offers the only sensible solution for eliciting uncertainties of interest, including dependencies.

Dependency might be feature of the problem and there are models capable of representing this dependency. Earlier studies (Park et al., 2021; Quigley et al., 2013; Song, 2002; Werner et al., 2021; Yahav and Shmueli, 2012) show estimation accuracy gains in modelling dependency when it exists. However, when building models for real problems, the demands of data collection increase when dependency might exist. There remains issues to explore about the value of actually getting dependency data.

1.3 Aims and Objectives

The overall aim is to investigate the performance of an analytical approach based on a multivariate Poisson-Gamma probability model with parameter dependency to learn about the effects of choices on the way dependency is represented, the type of statistical inference conducted, and the implications of mis-specifying the form or strength of the dependency. The research is bounded to consider to the multivariate Poisson-Gamma probability model to allow investigation of issues not yet reported in the literature.

The research is motivated by industry problem of accurately estimating the reliability of a new variant engineering system design. Data about an existing system is present and available to be drawn upon. Can we make effective use of all existing available data? The question is how? Additionally, we include some knowledge about the assumed failure rate and its uncertainty around it in form of a prior distribution with parameters, which can be elicited for example via expert judgement facilitating and enhancing the analysis. To address the posed question, a suggested approach would involve a mechanism that includes the integration of all relevant usable data with the utilization of a proposed prior distribution.

The described situation is not only observed by manufacturers, but also arises in various circumstances. Whether it is measuring counts of accident types in road safety data or looking at similar counts of failure occurrences of failure modes between two machines. The similarity capturing the strength of the statistical dependency between the target failure rate and the failure rate of a similar product is measured via the linear correlation parameter. For a copula framework, we use rather the rank correlation.

Our research objectives are:

- *R.O. 1: To investigate the estimation accuracy using empirical Bayes inference when there is correlation between the mean rates in a multivariate Poisson-Gamma probability model.*

Estimating failure rates using data implies we are without the knowledge of its true underlying rate. Using empirical Bayes inference, we evaluate at which accuracy are the estimates obtained.

- *R.O. 2: To assess the changes in estimation accuracy when the correlation between the mean rates in a multivariate Poisson-Gamma probability model is mis-specified.*

Since eliciting dependencies can be challenging, we like to question ourselves about what would happen if the expert assessments are subject to systematic bias. Would it be of significant importance? We will investigate the magnitude of such an induced error for various scenarios.

- *R.O. 3: To examine the effect on estimation accuracy and precision when the dependency in a multivariate Poisson-Gamma probability model is represented using a copula, including when the copula family is mis-specified.*

We may experience mis-specification of the dependency structure represented via copula function.We might assume a strong correlation value on one tail dependence and weak correlation on the other tail, so we choose the suitable copula. But what if we misjudge these dependencies and choose a certain copula, while the actual true underlying copula is another?

- *R.O. 4: To illustrate the application of methods to account for dependency in the parameters of a Poisson-Gamma model using both empirical Bayes inference and copulas through an industry case. Further:*
  1. *To compare the relative performance of the different methods for this case.*
  2. *To explore how the results of modelling might be translated to a decision-making context.*

For the considered case study, we utilize data provided by Siemens Analyse about machine failure. Can we make use of similar counts of failure occurrences of failure modes between two machines? Is it possible to apply the developed framework. Confirm the underlying statistical assumption by performing relevant tests? The results would enable us to make predictions about future events.

1.4 Thesis Structure

Chapter 2 is a modelling chapter aimed at establishing the foundation for the modelling framework for our research. Poisson processes and homogeneous Poisson processes (HPP) are introduced establishing groundwork in modelling assumptions for subsequent analyses. The chapter introduces the reader with Bayesian inference. Alongside, derives from it the relatively more novel and compared to the Bayesian analysis lesser-known case of the empirical Bayesian inference. Notably, the copula theory is introduced as a tool to capture dependencies within the prior distribution, thereby enhancing the overall modelling framework flexibility. Lastly, the chapter culminates by discussing estimation approaches with particular attention to the bias and mean squared error (MSE) as metrics to gauge the efficacy and accuracy.

Within this thesis in Chapter 3 (peer-reviewed journal article), we examine whether it is worthwhile eliciting subjective judgments to account for dependency in a multivariate Poisson-Gamma probability model. We have conscientiously selected the Poisson-Gamma model to account for on one side the inherent randomness in the form of aleatory uncertainty and on the other side for the knowledge-based uncertainty described as epistemic uncertainty. The Poisson element captures aleatory uncertainty and resembles a homogeneous Poisson process (HPP). A counting process characterized by a mean rate. The gamma distribution brings epistemic uncertainty to the mean rate of the Poisson distribution. It brings ambiguity around the parameter. Research suggests it to be a suitable model for modelling under-dispersed and over-dispersed data. The challenge of estimating reliability during product design motivated the choice of model class. The work is motivated by modelling supports the reliability program of a new product in development for a one-shot system. For this new product design family were identified some similar elements from earlier designs. With the help of expert knowledge, engineers are able to enrich the limited pool of available empirical data with relevant rich historical information. For the multivariate Poisson-Gamma model, we adopt an empirical Bayes methodology to present an estimator with improved accuracy of failure rate estimation. Engineers are facing modern-day reliability problems of different types. When working with similar failure-generating processes, they encounter small data samples, they need to be flexible and adjust to changing parameters of assumed marginal distribution. Aditionally, there is a substantial need to capture dependency precisely. As Dias et al. (2017, p. 124) points out, ignoring dependence in data estimation overconfidence and inappropriate assumptions. To investigate the estimation error, we take the route of a simulation study. Different parameters of the marginal distribution, pool sizes and dependencies are being modelled. A simulation study investigates the estimation error of this estimator for different degrees of dependency and examines the impact of dependency being mis-specified when assessed by subjective judgment. Our theoretical and simulation findings give analysts insights into the value of eliciting dependency.

Chapter 4 investigates the impact of an incorrect choice of a copula on the error estimate. We develop a Poisson – copula with Gamma marginals probability model. We wish to consider situations with a strong or weak association in one of the tails. In applications can be observed that in the joint cumulative distribution do the marginal distribution have an increased/decreased association on a distribution tail. Applied, failure rates in the lower tail may be more (or less) correlated than failure rates in the upper tail. To specify the structure of the dependence, we work with 4 widely used copulas. For these we choose an assumed and a true copula. From $4 \times 4 = 16$ possible combinations, we identify 4 main combinations that matter. Looking at two authoritative measures, namely the bias and MSE. For these we see, how changes occur over percentiles (10th, 50th and 90th). Guidance is provided on suggested situations when such a fallacy matters.

Chapter 5 shows how it is possible to improve the failure rate estimation of a certain machine using additional data of a similar machine on data kindly provided by Siemens Analyse. It is rationalized that similar counts are observed for same failure causes. Aim to reduce the variability around the investigated estimates and provide more certainty to decision makers. The underlying assumptions of working with a homogeneous Poisson process are tested. The results and research is finally validated by a K-fold cross-validation.

Chapter 6 summarizes the thesis and points out the highlights of each chapter. It addresses and evaluates, how are the research objectives are met and is shedding light on the contributions to knowledge of the thesis. The chapter points out limitations and future work.

Chapter 2

Modelling Background

The purpose of this chapter is to provide material to set up the modelling family in which we are conducting our research.

We shall describe homogeneous Poisson processes and pave the path to them by earlier defining stochastic processes and counting processes. This gives a pretext to introducing the Poisson model. The model is compared in the context of standard frequentist estimation, Bayesian estimation. As a next step we introduce the empirical Bayesian estimation and an interesting aspect of including partial dependency.

Variables that fluctuate randomly over time are called stochastic processes. They are a collection of random variables indexed by time. They are indicating randomly occurring events. Researchers encounter them in form of such as Markov processes, Brownian motion, and Poisson processes. Stochastic processes can be used to describe complex models such as fault detection models and fault removal processes shown in (Walls and Quigley, 1999). Poisson processes have proved themselves well to model failures or repairs in a given time frame. A **stochastic process** is a family of random variables $\{X_\theta\}$, indexed by a parameter θ, where θ belongs to some index Θ. A **counting process** $X(t)$ is a process in discrete or continuous time for which the possible values of $X(t)$ are the natural numbers $(0, 1, 2, \ldots)$ with the property that $X(t)$ is a non-decreasing function of t (Ross, 1995).

2.1 Poisson Processes and HPP

One of the most used counting processes is the Poisson process (Baxora, 2021). Let us imagine a starry sky. Stargazers enjoying the beauty of a clear view ponder, how to describe the shining dots statistically. There is no clear pattern and points are scattered without a visible regularity. These phenomena are well captured within Poisson processes. From the starry sky moving towards a reliability context, we shall remove one dimension and experience dots lined up on one axes, now representing time. This would lead us to a one dimensional Poisson process with modelling instead of stars failure occurrence. The Poisson process describes a counting process $N(t)$, specifying the total number of occurrences up to time t. The process is governed by the intensity function $\lambda(t)$, detailing the instantaneous rate of occurrences. The process is well suitable for statistically describing changes through time in the number of buses arriving at a station, number of mutations in a DNA sequence and most importantly, the number of failures in a machine. The number of occurrences is governed by a failure rate. A counting process $\{N(t), t \in [0, \infty)\}$ is called a Poisson process with rate $\lambda > 0$ (fixed) under defined circumstances.

The counting process is a **homogeneous Poisson process (HPP)** by Ross (2006, p. 313), if having three features

1. $N(0) = 0$.
2. The process has independent increments.
3. The number of events in any interval of length t is Poisson distributed with mean λt. That is, for all s, $t \geq 0$:

$$P\{N(t+s) - N(s) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n = 0, 1, \ldots \tag{2.1}$$

2.1.1 Stationarity

Stationarity is an important assumption of stationarity made in formulating the homogeneous Poisson process (Gabbiani and Cox, 2010, p. 256).

Definition 2.1.1 (strict-stationarity)**.** *A stochastic process* $X(t)$ *by (Zhang, 2022, p. 4) is strictly stationary if the sets of random variables* $X(t_1+\tau),\ldots,X(t_k+\tau)$ *and* $X(t_1),\ldots,X(t_k)$ *have identical joint probability distribution for any* $\tau > 0$ *and* $t_1, t_2, \ldots, t_k$ *where* $k = 1, 2, \ldots$

Thus, a stochastic process is strict-sense stationary iff:

$$f_{X(t_1+\tau),X(t_2+\tau),\ldots,X(t_k+\tau)}(x_1, x_2, \ldots, x_k) = f_{X(t_1),X(t_2),\ldots,X(t_k)}(x_1, x_2, \ldots, x_k) \quad (2.2)$$

for all k, all choices of $n_1, n_2, \ldots, n_k$ and all time offsets τ.

Definition 2.1.2 (wide-stationarity)**.** *A stochastic process X(t) by Zhang (2022) is wide-sense stationary, if*

1. *its mean is constant, i.e.* $E\{X(t)\} = \mu_X$*;*
2. *its second order moment is bounded, i.e.,* $E\{X(t)X^T(t)\} = E\{|X(t)|^2\} < \infty$
3. *its correlation function is time independent, i.e.,*

$$C_{XX} = E\{[X(t) - \mu_X][X(t-\tau) - \mu_X]^T\}$$

Speaking about stationarity, we shall consider wide-sense stationarity. If a stochastic process is strict-sense stationary, then in addition is the process wide-sense stationary.

When modelling event data using a stochastic process then the tests for stationarity are the ADF (Augmented Dickey Fuller) and KPSS (Kwiatkowski-Phillips-Schmidt-Shin) test. The KPSS test is in the null hypothesis not rejecting stationarity, whereas the ADF test can not accept the hypothesis of non-stationarity (Kwiatkowski et al., 1992).

2.2 Bayesian Inference

For the Bayesian method, we are assuming that the data originate from a distribution, whose parameters are known to us. The method allows us to combine information contained in the data with that of an a priori guess represented with prior distribution Bayesian methods offer an opportunity to make use of expert judgement on the value of the parameter. The need arises to specify in first place the prior distribution. The association of the prior with the data follows via application. As a result, we obtain the posterior distribution. The updated assigned probabilities for distribution of the underlying parameters. So that conclusions about the parameter distribution may be drawn, by the Bayes Theorem. Let us at first define the conditional probability. It described the probability of even A given event B has already occurred. It equals to the probability of events A and B jointly occurring ($A \wedge B$) divided by the probability of B.

Definition 2.2.1 (Conditional Probability)**.**

$$P(A|B) = \frac{P(B \wedge A)}{P(B)} \tag{2.3}$$

The Bayes theorem takes the form:

Definition 2.2.2 (Bayes Theorem)**.**

$$\begin{aligned} P(A_i|B) &= \frac{P(B \wedge A_i)}{P(B)} \\ &= \frac{P(B|A_i)P(A_i)}{\sum\limits_{i=1}^{n} P(B|A_i)P(A_i)} \end{aligned} \tag{2.4}$$

In Bayesian statistics, the Bayes Theorem provides a tool to update the probability about an investigated parameter based on obtained new information. $A_1, \ldots A_n$ represent exclusive and exhaustive events. Let us replace A_i with the parameter θ and B with data x. The density of the posterior distribution is proportional to

$$\underbrace{f(\theta|x)}_{\text{Posterior}} \propto \underbrace{f(x|\theta)}_{\text{Likelihood}} \times \underbrace{f(\theta)}_{\text{Prior}} \tag{2.5}$$

where $f(x)$ is the normalising constant. With x denoting the observed data and θ the investigated parameter. Using the Bayes Theorem, we can update the prior distribution with data. The likelihood distribution results in the posterior distribution describing the investigated parameter θ after observing data x. The denominator of the Bayes Theorem, which is the normalising constant $f(x)$, is a numerical constant that can numerically estimated and does not have any implications on the inference. It can be omitted since not depending on θ. If new data becomes available, the posterior is used as the new prior that will be updated.

The data, more precisely the *likelihood distribution* stands in the Bayes Theorem as a term that updates the prior distribution.

$$f(x|\theta) = L(\theta) \propto P(\text{Data}|\theta) \tag{2.6}$$

We suppose that data are observed. These data are assumed to originate from a probabilistic distribution with a vector of parameters θ (Lawless, 2011, p. 49). The likelihood function, as a function of θ, describes the probability of the data we have observed. Often under assumption of total independence of observations (i.i.d), the likelihood function is constructed as the product of individual distributions of data probability distributions, a function of random samples $X_1, \ldots, X_n$.

$$P(\text{Data}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \tag{2.7}$$

The *prior distribution* is represented by $f(\theta)$. The distribution represents the knowledge about the parameter before observing any data. Priors have different levels of informativeness. This is the information contained in the prior distribution ranging from considerable level of certainty to complete uncertainty, an uninformative prior. An uninformative prior is often called a flat prior. The parameters that control the prior are named hyperparameters. The assigning of hyperparameter values may be based on a scientifically educated guess, expert judgement. (van de Schoot et al., 2021, p. 2) urge to conduct a prior sensitivity analysis to examine what impact will different values of the prior have on the posterior. Since the data are assumed to originate from the prior data population, one might expect that the posterior will not diverge largely from the prior. The likelihood will conform to the prior. However, a contrary outcome is not necessarily faulty as likelihood might be misjudged or we worked with biased data (van de Schoot et al., 2021, p. 5). The outcome of the Bayes analysis is not just a single value, but a full distribution, the posterior distribution.

We will assume the Gamma prior distribution $\pi(\lambda)$, given in Equation 3.4. The Gamma distribution is chosen especially for analytical convenience between the prior and likelihood function. This means that the posterior is of the same distribution as the prior distribution, but with different parameters. Another requirement is to choose a non-negative distribution. Specific forms of the bivariate distributions (Equations 2.27 and 3.14) are chosen to enable capturing dependency in the prior distribution.

For the prediction of new values is used the *predictive distribution* (Gelman et al., 1995). For the case, before any data is observed and only the prior is available, statisticians work with the *prior predictive distribution* translated as:

$$p(y) = \int p(y, \theta) d\theta = \int p(\theta) p(y|\theta) d\theta \tag{2.8}$$

It is essentially a weighted average of the distributions that the data follow, where the weights are provided by the prior distribution. After observing data (size k), the *posterior predictive* distribution becomes available with the predicted variable y^*:

$$\begin{aligned} p(y^*|y) &= \int p(y^*, \theta|y) d\theta \\ &= \int p(y^*|\theta, y) p(\theta|y) d\theta \\ &= \int p(y^*|\theta) p(\theta|y) d\theta \end{aligned} \tag{2.9}$$

2.3 Empirical Bayesian Inference

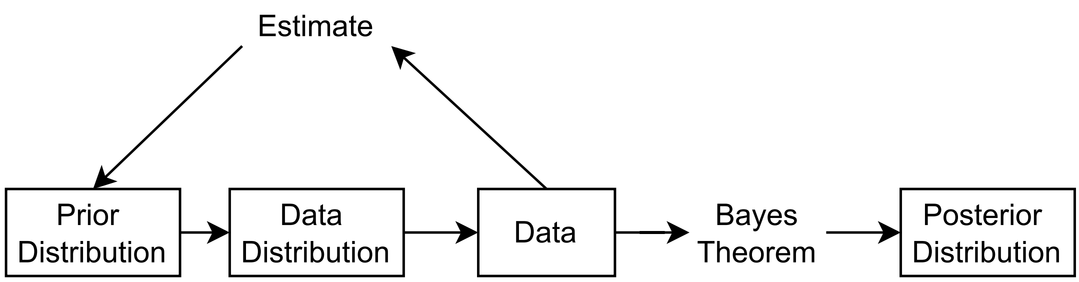


Figure 2.1: Empirical Bayes Process for obtaining the posterior distribution.

Between the classical frequentist and Bayes method lies the empirical Bayesian approach (EB). The empirical Bayesian method combines Bayesian and frequentist method. The difference to the Bayesian method is in the data being the source of information for determining the prior parameter, whereas in the Bayesian method the prior distribution is determined before the analysis possibly with the use of expert judgement. The development of the empirical Bayes methodology dates back to Robinns in the 1950s as Efron (2019) mentions. For a comparison of the Bayes and empirical Bayes methods Carlin and Louis (2010) details how these methods compare in theory and practice. The empirical Bayes method is being used in many applications and further developed (Zhang, 2003; Scott and Berger, 2010; Louis, 1991; Carlin and Louis, 2010; Leng and Kendziorski, 2015; Corotis, 2015; Quigley and Walls, 2011; Mukhopadhyay and Fletcher, 2018).

We are using the data to estimate the distribution parameters. After estimating the prior parameters, the analysis considers prior as known and proceeds towards Bayesian updating using the Bayes theorem. From the obtained posterior distribution of the parameter space, we may wish to take a single quantity as a representative such as the posterior mean. As in the way of the Bays method.

In Figure 2.1 is explained the process of obtaining the EB estimate. We assume data are following a probability distribution, whose parameter is governed by a prior distribution. We seek to estimate the prior parameters by pooling the data, i.e. the empirical prior distribution. After estimating the prior distribution by the Bayes Theorem, we can conduct Bayesian updating to obtain the posterior distribution.

There are two main pathways for estimation. One the non-parameteric EB estimation and the parametric EB estimation. In the thesis is applied only parametric EB. The parametric EB requires to choose a (data-based) value of the prior parameter and to specify the posterior distribution. The value of the prior parameter is typically specified using MLE or MME estimation procedure. One of the main advantages is that analysts are able to make inference about parameters that would be difficult set by expert judgement. The non-parameteric approach assumes an unknown form of the prior parameter.

Empirical Bayesian developments are well described by (Zhang, 2003 as cited in Efron, 2019). As Efron and Hastie (2016) points out, empirical Bayes has various usage. One of them is empirical Bayes hypothesis testing. Next to model selection (Frost and Savarino, 1986), parameter estimation and prediction. Or another usage is Robin's formula based on Bayes formula as a non-parametric point estimate. The estimation methods in the same way as in Carlin and Louis (2010) is divided into point estimates and interval estimates, with the possibility to be applied in the Poisson-Gamma model for reliability (Martz and Waller, 1982). As Efron (2019) points out, it is possible to evaluate the accuracy of the empirical Bayes inference from an omnibus perspective or an individual perspective, as Efron (2019) uses in this case finite Bayes inference. Efron (2019) also praises the impact on statistical practice, but points out the impediment in form of finding suitable data sets. Associated techniques to the empirical Bayes method, besides the frequentist approach and Bayes approach, can be seen as the Bayes linear methodology in Coolen et al. (2001) and Quigley et al. (2013), or Oracle Bayes in Jiang and Zhang (2009). The developed methodology in Quigley et al. (2007) aims to estimate rare events. The empirical Bayes methodology in this case is discussed to improve the accuracy of the estimates in light of having little data available. The methodology is useful especially when few data is available per hazard, but a pool of similar hazards can be constructed to assist in the inference Quigley et al. (2007).

Empirical Bayes has applications in risk, finance, economics, genetics, sport analytics and more. The EB method is used for economic problems like GDP growth as in Scott and Berger (2010), where the GDP growth regression is explained using various geographical, social and political predictors. For accident data EB can model insurance claims as in (Carlin and Louis, 2010). An application in pharmacy shows Louis (1991), in genetics Leng and Kendziorski (2015). Different concepts of uncertainty in mechanical and civil engineering can be found in Corotis (2015). Efron et al. (2001) uses EB in genetics in genes identification. Quigley et al. (2007) shows applications in technical risk analysis and Quigley et al. (2018) conducts an evaluation of suppliers within a supply chain. Gelman and Stern (2006) predicts with EB the game outcomes in sports. Frost and Savarino (1986) estimate risk and assemble portfolios. To conclude, the empirical Bayes methodology is a valuable methodology for improving accuracy in estimation generally and is particularly useful for risk reliability estimation.

2.4 Inference for the Multivariate Poisson - Dependency Context - Copula

By Trivedi and Zimmer (2007, p. 3), "copulas are functions that connect multivariate distributions to their one-dimensional margins". Nelsen et al. (1999, p. 8) defines a copula as follows in Definitions 2.4.1. Other type of definitions include a sub-copula (Nelsen et al., 1999, p. 19), from Sklar's theorem (Schweizer, 1991), or as a multivariate distribution (Joe, 1997, p. 12). Copulas are offering great flexibility as they can couple with no limits distributions together with complex dependence structures. The marginal choice is indeed arbitrary. We may see copulas as parametrically specified joint distributions that originate from given marginals (Trivedi and Zimmer, 2007, p. 7). Depicting the researched copulas, Figure 2.2 shows symmetry in the dependence structure for the Gaussian and Frank copula. The Gumbel copula shows increased dependence in the right tail, whereas the Clayton copula performs an increased dependence in the left tail.

Definition 2.4.1. *A copula is a function C: $[0,1]^2 \to [0,1]$ with the following proper-*

ties:

1. *For every* u, v *in* $\mathbf{I}$,

$$C(u, 0) = 0 = C(0, v) \tag{2.10}$$

and

$$C(u, 1) = u \;\; \text{and} \; C(1, v) = v \tag{2.11}$$

2. *For every* u_1, u_2, v_1, v_2 *in* $\mathbf{I}$ *such that* $u_1 \leq u_2$ *and* $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0. \tag{2.12}$$

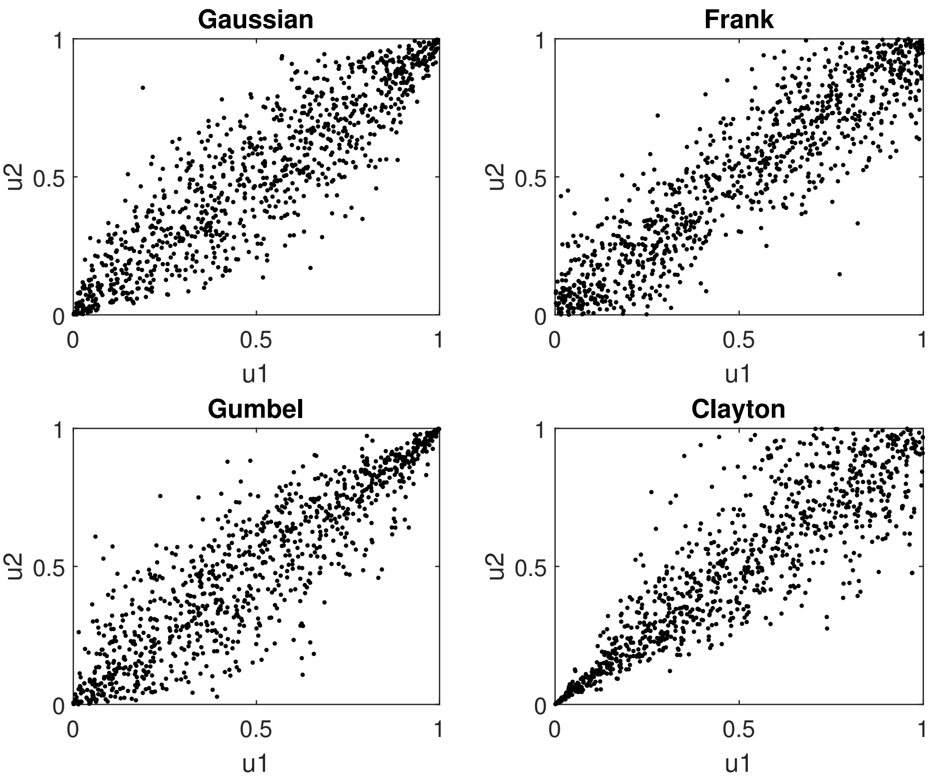


Figure 2.2: Copula samples of the Gaussian, Frank, Gumbel and Clayton copula when the dependence parameter is $\rho_S = 0.9$ with 1,000 generated samples.

Sklar's theorem (Sklar, 1959) portrays and elucidates, how the marginal and a copula are describing any multivariate distribution. By this, we justify the usage of copulas for multivariate distributions.

$$F(y_1, \dots y_d) = C(F_1(y_1), \dots, F_d(y_d)) \tag{2.13}$$

Implications from Sklar's theorem are by Nelsen (2007b) in (Ghosh, 2010, p. 7) from Equation 2.13 as follows:

1. We can define new families of multivariate distributions with requested specifications
2. We can study the copula and the marginal separately.

2.4.1 Archimedean Copulas

A widely popular class of copulas are the Archimedean copulas. Instead of a construction by Sklar's theorem, they are a Laplace transform (inverse) to a so-called generator. The most wide used Archimedean copulas are the Frank, Gumbel and Clayton copulas. Whereas the Gaussian would be an elliptical copula, a copula generated by an elliptical distribution (Frahm et al., 2003, p. 277). "Archimedean copulas are popular because they are easily derived and are capable of capturing wide ranges of dependence." (Trivedi and Zimmer, 2007, p. 41).

A copula C is called Archimedean if it can be written as

$$C(u_1, \dots, u_d|\theta) = \psi^{[-1]}(\psi(u_1; \theta) + \dots + \psi(u_d; \theta)|\theta) \tag{2.14}$$

, ψ is the so-called generator and $\psi^{[-1]}$ the pseudo-inverse. For more classes are books Joe (1997) and Nelsen (2007a) giving an overview and review.

2.4.2 Common Bivariate Copulas

We name some copulas that are present in the research as (Trivedi and Zimmer, 2007) details.

Gaussian copula

$$\begin{aligned} C(u_1, u_2|\theta) &= \Phi_G(\Phi^{-1}(u_1), \Phi^{-1}(u_2)|\theta) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \times \left(\frac{-(s^2 - 2\theta st + t^2)}{2(1-\theta^2)} \right) dsdt \end{aligned} \tag{2.15}$$

dependency θ within the interval $(-1, 1)$. Allowing for both positive and negative dependency. The Gaussian copula shows flexibility by permitting both positive and negative dependency. It is a symmetric copula. Φ stands for the cdf of the standard normal distribution and $\Phi_G(u_1, u_2; \theta)$ the bivariate standard normal distribution with dependency parameter θ.

Frank copula

$$C(u_1, u_2|\theta) - \theta^{-1} \log \left\{ 1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right\} \tag{2.16}$$

Also negative dependency and symmetric alike to the Gaussian copula. Weak tail dependencies characterise both tails. The strongest dependency is around the centre. All values are permissible for the dependency parameter, $\theta \in (-\infty, \infty)$.

Gumbel copula

$$C(u_1, u_2|\theta) = \exp\left(-(\bar{u}_1 + \bar{u}_2)^{1/\theta} \right) \tag{2.17}$$

with $\bar{u}_j = -\log(u_j)$, with the dependence parameter $\theta \in [1, \infty)$. Suggests a strong right tail and comparable weak left tail. Only positive dependency is permitted.

Clayton

$$C(u_1, u_2|\theta) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta} \tag{2.18}$$

With the parameter $\theta \in (0, \infty)$, the copula is used to study cases, where strong left tail dependencies are observed. Whereas the right tail dependencies are weak. The generator of the copula is

$$\phi(s; \theta) = (1 + \theta s)^{-1/\theta} \tag{2.19}$$

The copula allows only positive dependency.

Special copulas (in the bivariate case $d = 2$) for dependency extremes are the *independency*, *co-monotonicity* and *counter-monotonicity* copulas. The independence copula is defined as

$$C^{\text{ind}}(u_1, \dots, u_d) = u_1 \dots u_d \tag{2.20}$$

The co-monotonicity with perfect positive dependence as

$$\begin{aligned} C^M(u_1, \dots, u_d) =& P(U \leq u_1, \dots, U \leq u_d) = P(Y \leq \min(u_1, \dots, u_d)) \\ =& \min(u_1, \dots, u_d) \end{aligned} \tag{2.21}$$

counter-monotonicity C^{CM} with perfect negative dependence as

$$\begin{aligned} C^{CM}(u_1, u_2) &= P(U \leq u_1 \,\&\, 1 - U \leq u_2) \\ &= P(1 - u_2 \leq U \leq u_1) = \max(u_1 + u_2 - 1, 0) \end{aligned} \tag{2.22}$$

All the above copula converge in the limiting cases for dependency to the independence copula, for full-dependency to the co-monotonicity copula and where is negative dependency feasible to the counter-monotonicity copula (Ruppert, 2011, p. 178).

2.5 Multivariate Poisson with Copula in Prior Distribution

To model varying dependency structures in the prior distribution, we create a Poisson-Gamma model with copula dependency in the prior rates (Chapters 5 and 6).

$$P\left(N_i = n_i | \Lambda_i = \lambda_i\right) \sim Poisson\left(\lambda_i\right) \tag{2.23}$$

We assume that the data follow a Poisson distribution.

$$\Lambda_i \sim Gamma\left(\alpha, 1\right) \tag{2.24}$$

The rates originate from a gamma distribution.

$$G(\lambda;\alpha,1)=\int_0^{\lambda}\frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}dx \tag{2.25}$$

With the according cdf.

$$F(\lambda_1,\lambda_2)\sim C_\theta\left(G(\lambda_1;\alpha,1),G(\lambda_2;\alpha,1)\right) \tag{2.26}$$

The distribution functions are now coupled with a copula, given the dependency parameter θ.

$$\pi(\lambda_1,\lambda_2)=\frac{d^2C_\theta(G,G)}{d\lambda_1 d\lambda_2} \tag{2.27}$$

The joint density function of λ_1 and λ_2 is derived from the copula via differentiation.

$$\pi(\lambda_1,\lambda_2|\,n_1,n_2)\propto Poisson(n_1;\lambda_1)\,Poisson(n_2;\lambda_2)\,\pi(\lambda_1,\lambda_2) \tag{2.28}$$

It is possible to formulate the density as the product of the data's probability functions with the copula density.

$$\begin{aligned}\pi(\lambda_1,\lambda_2|\,n_1,n_2)&=\frac{\frac{\lambda_1^{n_1}\lambda_2^{n_2}e^{-(\lambda_1+\lambda_2)}}{n_1!n_2!}\pi(\lambda_1,\lambda_2)}{\int_0^\infty\int_0^\infty\frac{\lambda_1^{n_1}\lambda_2^{n_2}e^{-(\lambda_1+\lambda_2)}}{n_1!n_2!}\pi(\lambda_1,\lambda_2)\,d\lambda_1 d\lambda_2}\\&=\frac{\lambda_1^{n_1}\lambda_2^{n_2}e^{-(\lambda_1+\lambda_2)}\pi(\lambda_1,\lambda_2)}{\int_0^\infty\int_0^\infty\lambda_1^{n_1}\lambda_2^{n_2}e^{-(\lambda_1+\lambda_2)}\pi(\lambda_1,\lambda_2)\,d\lambda_1 d\lambda_2}\end{aligned} \tag{2.29}$$

In detail.

$$E[\Lambda_i]=\frac{\int_0^\infty\int_0^\infty\lambda_i\lambda_1^{n_1}\lambda_2^{n_2}e^{-(\lambda_1+\lambda_2)}\pi(\lambda_1,\lambda_2)}{\int_0^\infty\int_0^\infty\lambda_1^{n_1}\lambda_2^{n_2}e^{-(\lambda_1+\lambda_2)}\pi(\lambda_1,\lambda_2)\,d\lambda_1 d\lambda_2} \tag{2.30}$$

The marginal distribution is derived from the joint distribution. The formulas can be evaluated using Monte Carlo integration described in Section B.1.

To highlight:

- The joint density $f(y, x)$:

$$f(y, x) = c(F_Y(y), F_X(x)) f_Y(y) f_X(x)$$

- The conditional density $f_{Y|X}(y|X = x)$:

$$f_{Y|X}(y|X = x) = f_Y(y) c(F_Y(y), F_Y(x))$$

- The k-th moment $E(Y^k|X = x)$:

$$E(Y^k|X = x) = \int_{-\infty}^{\infty} y^k c(F_Y(y), F_X(x)) dF_Y(y)$$

The theoretical MSE under mis-specification can be formally written by (Acar et al., 2019, p. 11) as

$$\begin{aligned}
&MSE(\lambda_1|c_A, c_T, N_1, N_2, \rho, \alpha, \beta) = \\
&= E_0[\{m_W(\lambda_1) - \Lambda_1\}|\Lambda_2 = \lambda_2] \\
&= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \lambda_1 \; c_W(F_{\Lambda_1}(\lambda_1), F_{\Lambda_1}(\lambda_2)) f(\lambda_1) d\lambda_1 - \lambda_1 \right\}^2 \\
&\times c_T \left(F_{\Lambda_1}(\lambda_1), F_{\Lambda_2}(\lambda_2)\right) \; f(\lambda_1) d\lambda_1
\end{aligned} \tag{2.31}$$

where $m_W(\lambda_1)$ is the expectation under the assumed (working) copula. Examining similar methods in mis-specification. Comparing Acar et al. (2019) to our results, we look on the case, when the true copula equals the assumed to assure ourselves, that the method provides the correct choice. Whereas Dupuis (2007, p. 385) uses quantile differences, which are as well adopted in this research.

2.6 Classical Estimation Approaches

We define the *bias* as the expected difference between the estimate $\hat{\theta}$ minus the true value θ.

$$bias = E\left[\hat{\theta} - \theta\right] \tag{2.32}$$

This order is important for the bias contrary to the MSE, where the order of the true values and estimate is invariant. An estimator is said to be biased if its expected value is not equal to the true value of the parameter. A biased estimator can lead to incorrect conclusions about the estimated parameter.

The MSE (Mean Squared Error) is another standard measure of accuracy. It provides insights on how far on average are the estimated values from the true values.

$$MSE = E\left[\left(\hat{\theta} - \theta\right)\right]^2 \tag{2.33}$$

As a drawback is that the MSE is sensitive to outliers.

The *MME (Methods of Moment Estimation)* is estimating underlying model parameters. Setting sample moments equal to theoretical moments. A set of equation is then solved for the desired parameter variables. But there are known disadvantages like being sensitive to outliers (Arnold and Strauss, 1991). The k moment of a random variable is

$$\mu_k = E\left(X^k\right) \tag{2.34}$$

The matching k sample moment is

$$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k \tag{2.35}$$

The estimator is the solution of the equation $\mu_k = m_k$.

In the *MLE (Maximum Likelihood Estimation)*, the data are assumed to follow a distribution with a defined density. The joint density of the samples is forming the likelihood function. For the model parameters, the maximum of the function is found. The method has advantages like being assymptotically efficient, i.e. lowest variance among all unbiased estimators. Among drawbacks we mention not always being computationally easy and convenient for optimization Pratt (1976).

The common procedure is to take the logarithm of the likelihood, which reduces the computational complexity:

$$\hat{\theta} = \arg\max_{\theta} L(\theta|x) = \arg\max_{\theta} \log(L(\theta|x)) = \arg\max_{\theta} l(\theta|x) \tag{2.36}$$

2.7 Summary

In this chapter we have described the theoretical properties of the homogeneous Poisson model as a simple stochastic process model. We have also described the three approaches to inference which we adopt in later studies. Namely frequentist (classical estimation), full Bayes and empirical Bayes.

Chapter 3

Value of Expert Judgement Elicitation of Dependency for Multivariate Poisson-Gamma Model

3.1 Introduction

We examine whether it is worthwhile accounting for the dependency in a multivariate Poisson-Gamma probability model. More specifically, examine what is the benefit of explicitly accounting for correlation in the model towards error reduction, and hence towards estimation accuracy? Next, we examine what are the consequences of mis-specifying correlation acquired via subjective expert judgement elicitation? We do not seek to provide universal answers to this questions. Rather, we investigate this issue for a particular multivariate probability model, a Poisson-Gamma model.

This model has underpinned analysis for real industry problems. For example, recent modelling (involving two of the authors) to support decisions about the reliability of a one-shot system during new product development. This was a new generation of a product design family for which data from earlier design generations was deemed relevant for some elements of the new system together with test data generated for the new design throughout its development project. A correlation parameter represents the dependency in the multivariate Poisson-Gamma probability model used for this reliability estimation problem for the new design. The dependencies were elicited from suitably qualified engineers using a structured process based on the method described in Quigley and Walls (2018). This elicitation methodology maps the model parameters to the engineering expertise then uses relevant data (say from related past design elements and/or test) to quantify the dependency in view of the elicited judgements. The elicitation approach, grounded in a specifically designed defensible protocol, was resource intensive. It was also cognitively demanding for those expressing their subjective assessments despite having adopted an approach which asked engineers to express their engineering, rather than probabilistic, expertise.

The chapter is structured as follows. First we present an estimator for the multivariate Poisson-Gamma model that pools data from correlated processes and should result in reduced model estimation error. We develop this estimator through a comparative argument based on alternative inference approaches. We describe a simulation study to investigate the accuracy of the proposed inference approach given the degree of dependency (controlled through the correlation parameter) and the amount of data (controlled by the number of processes in the pool). This study is extended to further examine the impact of subjective mis-specification of the correlation parameter. We conclude by reflecting on the limitations of our study, the implications of our findings and provide suggestions for further work.

3.2 Model and Inference Framework

Our first objective is to develop an understanding of how pooling data from similar processes can reduce estimation error. Specifically, we consider homogeneous Poisson Processes (HPP) and we adopt an empirical Bayes framework to support the inference under the assumption of Gamma marginal prior distributions. These are conjugate to the Poisson and so are mathematically convenient as well as flexible. To motivate the value of the proposed inferential framework for our multivariate Poisson-Gamma model, we develop our reasoning through comparisons with standard methodological approaches. After describing a classical inference approach, which provides a benchmark for later assessing estimation error, we present the Bayesian approach and show the theoretical error reduction resulting by incorporating prior information. Since expressing subjective judgement in the form of a prior distribution can be challenging and resource intensive to elicit Quigley and Walls (2021), we are motivated to present an empirical Bayes approach where the Bayes mechanism is used but the prior distribution is estimated by pooling data on similar processes. Finally in this section we consider the pooling of data from multiple processes which are measurably correlated in their underlying mean rates, such that data from other processes can be explicitly incorporated into the inferential updating to reduce estimation error.

3.2.1 Classical Inference

Under classical inference we assume a probability model that measures the variation in the data as a function of a parameter. We consider a Poisson distribution with parameter λ which corresponds to the mean value of the distribution:

$$P(N = n; \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}, n = 0, 1, 2, \ldots \tag{3.1}$$

A typical classical approach to estimation would be either to estimate λ through a moment matching approach or a maximum likelihood approach. Assuming we have t observations from the same stochastic process, where the observations are denoted by n_j, then the estimator is given by:

$$\widehat{\lambda} = \frac{\sum_{j=1}^{t} n_j}{t} \tag{3.2}$$

To assess the accuracy of such an estimator we treat the data as random variables from the probability distribution and evaluate the Mean Square Error (MSE) which is given by:

$$MSE_C = E\left[\left(\lambda - \frac{\sum_{j=1}^{t} N_j}{t}\right)^2\right] = \frac{\lambda}{t} \tag{3.3}$$

The MSE_C for the classical estimator approaches 0 as the sample size increases.

3.2.2 Bayesian Inference

Under Bayesian inference we again assume the Poisson distribution but now consider it as a conditional probability distribution assuming the true mean, denoted by λ, is known. This mean is then modelled as a random variable where the uncertainty is described by a prior distribution. Here we assume the prior distribution $\pi(\lambda)$ belongs to the Gamma distribution family. This Poisson-Gamma model is given by:

$$\begin{aligned} P(N = n|\Lambda = \lambda) &= \frac{\lambda^n e^{-\lambda}}{n!}, n = 0, 1, 2, \ldots \\ \pi(\lambda) &= \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}, \alpha > 0, \beta > 0, \lambda > 0 \end{aligned} \tag{3.4}$$

Combining the Poisson and Gamma models, we obtain the predictive distribution which is essentially a weighted average of Poisson distributions where the weights are provided by the prior distribution. For this combination of Poisson data and Gamma prior, the prior predictive distribution is in the form of a Negative Binomial distribution:

$$\begin{aligned} P(N = n) &= \int_0^\infty \frac{\lambda^n e^{-\lambda}}{n!} \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} d\lambda \\ &= \frac{\Gamma(\alpha + n)}{\Gamma(\alpha) n!} \left(\frac{\beta}{\beta + 1}\right)^\alpha \left(\frac{1}{\beta + 1}\right)^n \end{aligned} \tag{3.5}$$

Similarly to the posterior predictive distribution after observing data taking the form of:

$$\begin{aligned} P(N^* = n^*|\underset{\sim}{n}) &= \frac{(k+\beta)^{\sum n_i + \alpha}}{\Gamma(\alpha + \sum n_i)\Gamma(n^*+1)} \int_0^\infty \lambda^{n^* + \sum n_i + \alpha - 1} e^{-(t+\beta+1)\lambda} d\lambda \\ &= \frac{\Gamma(n^* + \alpha + \sum n_i)}{\Gamma(\alpha + \sum n_i) n^*!} \left(\frac{\beta + t}{\beta + t + 1}\right)^{\alpha + \sum n_i} \left(\frac{1}{\beta + t + 1}\right)^{n^*} \end{aligned} \tag{3.6}$$

Since Bayesian inference incorporates prior information on the process then the mean prior, denoted by $E[\Lambda]$, should be specified before observing any data, where:

$$E[\Lambda] = \frac{\alpha}{\beta}. \tag{3.7}$$

Once data have been observed on the process (such as the aforementioned t observations), the prior distribution can be updated using Bayes' Theorem to give the following posterior distribution:

$$\pi(\lambda|\underset{\sim}{n}) = \frac{(\beta+t)^{\alpha + \sum_{j=1}^{t} n_j} \lambda^{\alpha + \sum_{j=1}^{t} n_j - 1} e^{-(\beta+t)\lambda}}{\Gamma\left(\alpha + \sum_{j=1}^{t} n_j\right)} \tag{3.8}$$

The associated posterior mean is:

$$E\left[\Lambda|\underset{\sim}{n}\right] = \frac{\alpha + \sum_{j=1}^{t} n_j}{\beta + t} \tag{3.9}$$

To facilitate comparison with the classical inference for the same model, we calculate the MSE assuming a Bayesian framework by first averaging over the mean and subsequently over the data that will be realised to obtain:

$$\begin{aligned} MSE_B =& E_N E_\Lambda \left[\left(\Lambda - E[\Lambda | \underset{\sim}{N}] \right)^2 \middle| \underset{\sim}{N} \right] \\ =& E_{\underset{\sim}{N}} \left[\frac{\alpha + \sum_{j=1}^{t} N_j}{(\beta + t)^2} \right] = \frac{\alpha}{\beta(\beta + t)} \end{aligned} \tag{3.10}$$

The MSE_B for the Bayesian estimator also approaches 0 as the sample size t increases. Further, inspection of MSE_B shows that it is less than $E[\Lambda]$, which is to say that prior to observing any data we anticipate that the MSE_B will be smaller than the expected value of the mean. Moreover, we anticipate that $MSE_B < MSE_C$ since the denominator of the former is $\beta + t$ rather than just t as for the latter. This insight is not surprising given more data are being introduced to the analysis in the form of prior information. The data are reflected in the prior distribution since the parameters of the prior distribution are estimated using moment method estimation (MME) as shows Algorithm 15. As β increases, the smaller the variance of the prior distribution (consistent with more precise judgement) and hence a smaller MSE_B.

3.2.3 Empirical Bayes Inference

An empirical Bayes inference approach presumes we have a pool of Poisson processes, each with their own λ all of which have been realised from the same probability distribution, namely the prior distribution. Thus, by pooling the data associated with the rates allows estimation of the prior distribution. Then Bayes' Theorem can be applied (as in a traditional Bayes approach) to provide a tailored posterior estimator for the process of interest.

The empirical Bayes estimator for the Poisson-Gamma model is given by:

$$E[\Lambda_i | N_i = n_i] = \frac{\widehat{\alpha} + n_i}{\widehat{\beta} + t} \tag{3.11}$$

where the estimators of the prior distribution are denoted by $\left(\widehat{\alpha}, \widehat{\beta}\right)$ and we index the mean values with subscript i to correspond to the i^{th} process in a pool of m Poisson processes. The corresponding mean square error is given by:

$$\begin{aligned} MSE_{EB} &= E_{N_i}\left[E_{\Lambda_i}\left[\left(\Lambda_i - \frac{\alpha + \sum_{j=1}^{t} N_j}{\beta + t}\right)^2 \Bigg| \underset{\sim}{N_i}\right]\right] \\ &+ E_{N}\left[E\left[\left(\frac{\alpha + \sum_{j=1}^{t} N_j}{\beta + t} - \frac{\widehat{\alpha} + \sum_{j=1}^{t} N_j}{\widehat{\beta} + t}\right)^2 \Bigg| \underset{\sim}{N}\right]\right] \\ &= MSE_B + MSE_{PE} \end{aligned} \tag{3.12}$$

The MSE_{EB} can be decomposed into two terms. The first term is the MSE_B (Equation 3.10) and the second term is the MSE_{PE}, which is the pool parameter estimation error. This implies that MSE_{EB} is affected by both the number of processes in the pool and the number of observations in each process since MSE_B only decreases as more data are observed for process i and MSE_{PE} decreases as the number of process in the pool increases. Thus, an empirical Bayes estimator provides a means of reducing estimation error since it allows the error to become closer to that of a Bayes estimator without the need for a prior distribution. However, we note that the role of increased pool size is to reduce MSE_{PE} only and not MSE_B.

3.2.4 Dependency between Processes

We now consider the situation of primary interest where we wish to include data that is correlated with a process of interest with the aim of reducing MSE_B. To accommodate this we require a multivariate prior distribution to model the correlation between the data generating processes. This multivariate prior model can be used within a Bayesian approach if obtained by subjective judgement or within an empirical Bayes approach if the parameters have been estimated from observations across a pool of processes.

As a motivating example, we could consider estimating the rate of occurrence of major accidents at a specified location. By pooling data on major accidents across multiple locations then an empirical Bayes estimator could be derived to improve the accuracy of the estimators. However, we can also include data on minor accidents at each location where the rates are likely to be correlated with each other but not necessarily perfectly. As such, the data from the minor accidents for the same location can have a direct impact on reducing the MSE_B due to its correlation with the major accidents at that location.

Here we propose a framework that could be operationalised with a Bayesian or an empirical Bayes approach to inference depending upon how the prior parameters have been obtained. Let

$$\rho = corr(\Lambda_1, \Lambda_2) \tag{3.13}$$

For $\rho < 1$ we assume the following bivariate Gamma distribution developed by Minhajuddin et al. (2004) for which the marginal distributions for each process have a Gamma prior:

$$\begin{aligned} \pi(\lambda_1, \lambda_2) = & \frac{\beta^\alpha \lambda_1^{\alpha-1} e^{-\beta\lambda_1}}{\Gamma(\alpha)} \frac{\beta^\alpha \lambda_2^{\alpha-1} e^{-\beta\lambda_2}}{\Gamma(\alpha)} \\ & \times (1-\rho)^{\alpha-2} {}_0F_1\left[[\,], [\alpha]; \left(\frac{\beta}{1-\rho}\right)^2 \rho\lambda_1\lambda_2\right] \end{aligned} \tag{3.14}$$

where

$$\begin{aligned} & {}_0F_1\left[[\,], [r]; \left(\frac{\beta}{1-\rho}\right)^2 \rho\lambda_1\lambda_2\right] \\ = & \sum_{k=0}^{\infty} \frac{\left(\left(\frac{\beta}{1-\rho}\right)^2 \rho\lambda_1\lambda_2\right)^k}{\Gamma(\alpha+k)k!} \end{aligned} \tag{3.15}$$

For $\rho = 1$, we assume the Gamma prior distribution $\pi(\lambda)$, given in Equation 3.4.

The bivariate Gamma distribution in Equation 3.14 was first proposed as a multivariate prior for an HPP by Quigley et al. (2013) where many of the results we require are derived. Here we state only those results which are key for our research. The posterior mean for this model which is given by:

$$E[\Lambda_i|n_1, n_2] = \frac{n_i + \alpha + E[K]}{t + \frac{\beta}{1-\rho}} \tag{3.16}$$

where K is a discrete random variable whose probability distribution belongs to generalised hypergeometric family of distributions (Patil and Joshi, 1968). This distribution is expressed as:

$$P(K = k) = \frac{\frac{\Gamma(\alpha)\Gamma(n_1 + \alpha + k)\Gamma(n_2 + \alpha + k)}{\Gamma(\alpha + k)\Gamma(n_1 + \alpha)\Gamma(n_2 + \alpha)} \frac{\left(\rho\left(\frac{\beta}{t(1-\rho)+\beta}\right)^2\right)^k}{k!}}{{}_2F_1\left[[n_i + \alpha], [\alpha]; \rho\left(\frac{\beta}{t(1-\rho)+\beta}\right)^2\right]}, k = 0, 1, 2, \ldots \tag{3.17}$$

Further, (Quigley et al., 2013) show that:

$$\lim_{\rho \to 0} E[\Lambda_i|n_1, n_2] = \frac{n_i + \alpha}{t + \beta} \tag{3.18}$$

$$\lim_{\rho \to 1} E[\Lambda_i|n_1, n_2] = \frac{n_1 + n_2 + \alpha}{2t + \beta} \tag{3.19}$$

These results indicate that as the correlation approaches 0, we obtain the Bayes estimate for the multivariate Poisson-Gamma model. Also, as correlation approaches 1, we obtain the same estimate as we would derive if all $2t$ observations were observed from the same Poisson process. While we can reason the effect of dependency under perfect and no correlation, we are interested to understand the effects for varying degrees of dependency.

Hence now that we have presented an estimator that, by pooling data from correlated processes, should reduce estimation error, we investigate the accuracy of this method for changes in the degree of correlation and the size of the pool of processes.

3.3 Simulation Study for Estimation Error

We conduct a simulation study to investigate the MSE of the estimator obtained from pool dependent data, where the parameters of the marginal distribution are estimated from observations using an empirical Bayes approach but the correlation parameter is specified through subjective judgement. This reflects the general modelling situation where engineering experts identify relevant data sets and provide a measure of their similarity between these data sets. Specifically this is the case for our motivating industry problem when we estimated the reliability of a new variant engineering system design.

After describing the simulation study design, we discuss the conditions that lead to under dispersed data being generated in our simulations – under dispersion occurs when the variance of the Negative Binomial distribution used to model the distribution of the data in the pool is less that the corresponding mean. Then we present the results from the simulation study and provide an expression that relates the MSE to correlation and pool size.

3.3.1 Study Design

We assume the correlation between two processes has been specified by subjective judgement but that the marginal Gamma distributions have been estimated with observed data, hence an empirical Bayes inference approach is adopted, (e.g. following Quigley and Walls (2018). We assume a pool of HPPs each with a pair of correlated observations. The purpose of the study is to assess the impact of the correlation (ρ) and pool size (m) upon the MSE of the estimator. As in the previous section, the rates are assumed realised from a Gamma distribution. But, without loss of generalisation, we set the scale parameter to be $\beta = 1$. Moreover, we assume the data are realised from the HPP given their rates. To estimate the parameters of the marginal distribution, (α, β), following (Quigley et al., 2007), we use a moment based approach to obtain the following estimators:

$$\widehat{\alpha} = \frac{U^2}{W - U^2}$$
$$\widehat{\beta} = \frac{U}{W - U^2} \tag{3.20}$$

where

$$U = \frac{\sum_{i=1}^{m} n_i}{m} \text{ and } W = \frac{\sum_{i=1}^{m} n_i^2}{m} - U \tag{3.21}$$

A range of values is specified for the three parameters we wish to control in the simulation study – correlation, pool size and shape parameter of the marginal distribution – as shown in Table 3.1.

Table 3.1: Parameter values controlled in the simulation study

Input parameter	Specified Values
m	$5, 10, 20, 30, 40, 50, 60$
α	$0.5, 1, 5, 10, 20, 30, 40, 50$
ρ	$0, 0.1, 0.2, 0.3, \ldots, 0.7, 0.8, 0.9, 1$

The algorithm for the study is as follows.

Algorithm 1 Simplified simulation steps in code inner loop:

1: Set: α, m and ρ
2: **for** $j = 1 : m$ **do**
3: Simulate: $\lambda_{1j},\ \lambda_{2j} \sim \Gamma(\alpha, 1; \rho)$
4: Simulate: $[N_{1j},\ N_{2j}] \sim Poisson(\lambda_{1j}t, \lambda_{2j}t)$
5: **end for**
6: Calculate: $[\hat{\alpha}, \hat{\beta}] \xleftarrow{\text{moment estimator}} [\underset{\sim}{N_{1\cdot}}, \underset{\sim}{N_{2\cdot}}]$
7: **if** $\hat{\alpha} \leq 0$ or $\hat{\beta} \leq 0$ **then** ▷ underdispersion
8: $E[\Lambda_1|\cdot] \leftarrow \frac{\sum_{i=1}^{2}\sum_{j=1}^{m} N_{ij}}{2m}$
9: **else**
10: **for** $j = 1 : m$ **do**
11: Calculate: $E[\Lambda_{1j}|N_{1j}, N_{2j}, \hat{\alpha}, \hat{\beta}, \rho]$
12: **end for**
13: **end if**
14: **for** $j = 1 : m$ **do**
15: Calculate: $E[\Lambda_{1j}|N_{1j}, N_{2j}, \hat{\alpha}, \hat{\beta}, \rho] - \lambda_1 = \mathsf{e}_{1j}$
16: **end for**
17: Calculate: $MSE = \left(\frac{\sum_{j=1}^{m} \mathsf{e}_{1j}^2}{m} \right)$

We create an empirical Bayes simulation, where the rate parameter $\beta = 1$ is same, but shape parameter $\alpha = 0.5, 1, 5, 10, 20, 30, 40, 50$, pool size $m = 5, 10, 20, \ldots, 60$, correlation $\rho = 0, 0.1, \ldots, 0.9, 1$ differs, as shows algorithm 1. We generate a pool of m correlated pairs of samples from a multivariate gamma distribution (Minhajuddin et al., 2004), Equation 3.14 in Step 3. From these, Poisson samples $[N_{1i}, N_{2i}]$ are drawn. Using the moment estimator (Quigley et al., 2007) in formula 3.20, we estimate the gamma prior parameters α and $\beta = 1$, Code line 6. If the samples have underdispersion $(\widehat{\alpha}, \widehat{\beta} < 0)$, we will use the mean of Poisson samples $\mathbf{N_{1\bullet}}$ for estimating the posterior mean $E[\Lambda_{1i}|\rho_A, \mathbf{v}]$ in Equation 3.16. This means the mean of the pair $[N_1, N_2]$ in Code line 7 is the best estimate of the posterior mean. In the next Section 3.3.2 we expand the discussion when we are likely to experience underdispersion.

3.3.2 Treatment of Underdispersion

Since we use a Negative Binomial distribution for the data in the pool when sampling from Poisson-Gamma model we risk encountering the problem of underdispersed data, as discussed in Kokonendji et al. (2008). Underdispersion occurs if the variance of the Negative Binomial distribution is smaller than its mean due to sampling and compromises the moment estimator proposed by Quigley et al. (2007).

Figure 3.1 shows the probability of underdispersion given the choice of pool size and shape parameter across all simulation combinations in our study. The plot shows that the highest chance of underdispersion occurring is for our smallest combination of shape parameter and pool size ($\alpha = 0.5$, $m = 5$). We find an area of low underdispersion for $\alpha \geq 5$, pool size ≥ 30 and our plot suggests that a smaller pool size matters more than a small value of α. If underdispersion occurs then we have two options to address it; either to discard the samples or to take the mean and use it as our best estimate. We chose the latter option. If we would not discard the samples and resample instead, until no underdispersion occurs, we would crop from the simulation a part of the DGP. The full data generating process encompasses alongside these samples that pose challenges.

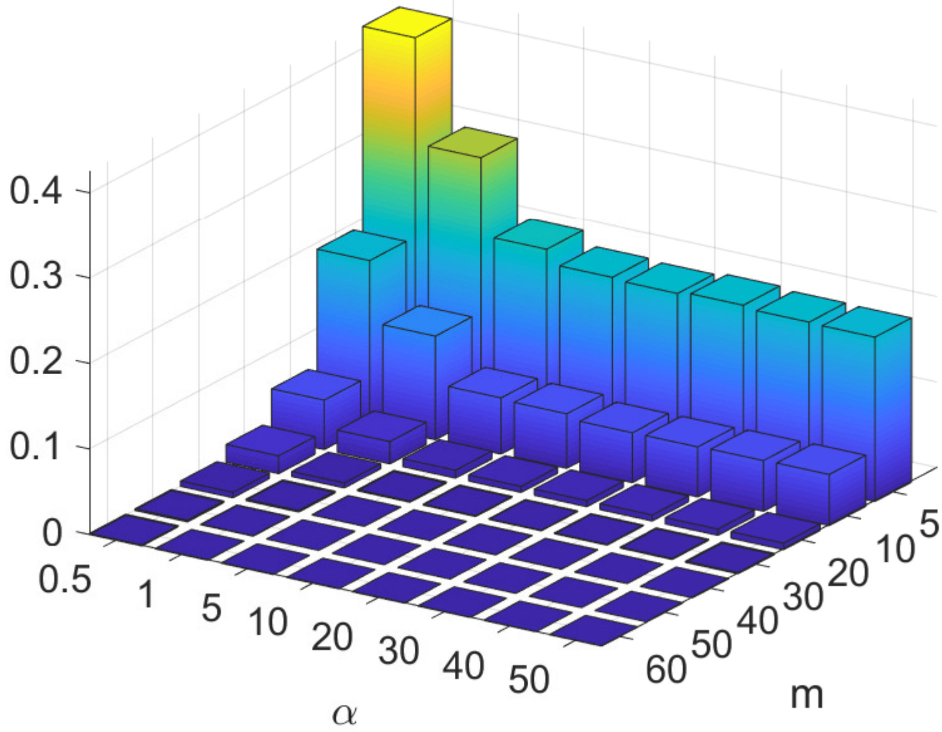


Figure 3.1: Probability of underdispersion for simulation combinations of shape parameter and pool size.

3.4 The Linear Relationship Between the MSE and the Prior Rate, Given Correlation and Pool Size

$$MSE = \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij}\alpha \cdot \mathbb{1}_{\rho(i)}\mathbb{1}_{m(j)} + \epsilon \tag{3.22}$$

$$\log\left(\frac{MSE}{MSE_{\infty}} - 1\right) = \sum_{i=1}^{I}\sum_{j=1}^{J}(a_{ij}^{0} + a_{ij}^{1}m) \cdot \mathbb{1}_{\rho(i)}\mathbb{1}_{\alpha(j)} + \epsilon \tag{3.23}$$

Our focus turns on the results of the simulation is the MSE for each correlation and pool size. We observe in Figure 3.3 that the MSE is proportional to the prior rate α. So how could we capture for each correlation value and each pool size the relation to the MSE? This model is denoted in Formula 3.22. This relationship is well evaluated in regression ($r^2 = 99.98\%, r^2_{adj} = 99.98\%$). If we model the coefficients against (Figure 3.2) the log pool size $\ln(m)$ and second power of the correlation ρ^2, the coefficients become visually linear towards these parameters. Figures 3.7 show the MSE of the simulation against the prior rate α for different pool sizes.

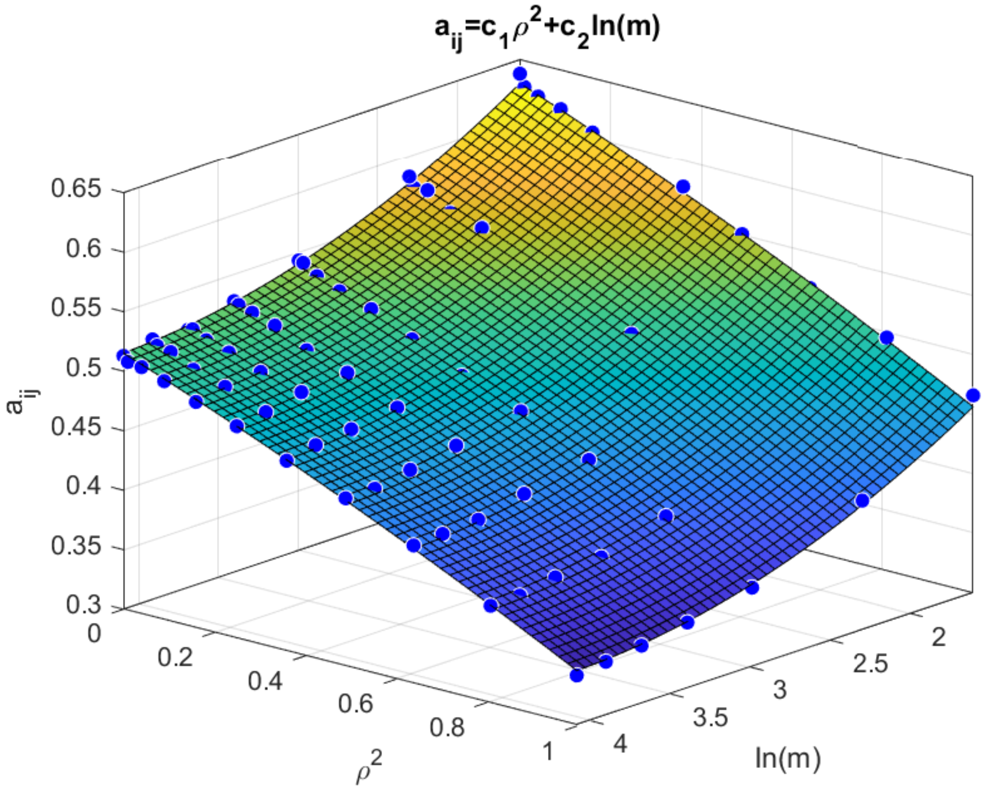


Figure 3.2: Modelling regression coefficients as a visually flat surface to justify choice of quadrature and logarithm. Z-axes contains the estimated coefficient from Equation 3.22 .

For the second set of Figures 3.4, we have the same MSE data but group them accordingly to the pool size. The MSE shows to be linear (Figure 3.5) to the pool size after the transformation in Equation 3.23 with a high degree of precision from regressing this model ($r^2 = 94.46\%, r^2_{adj} = 91.71\%$). We see that the curves in Figures 3.4 are approaching towards an asymptote. We calculated this asymptote for the case

$\rho_T = 0, 1$ in Table 3.2. On this figures we apply the transformation

$$\log\left(\frac{MSE}{MSE_\infty} - 1\right) \tag{3.24}$$

, where $\lim_{m\to\infty} MSE = MSE_\infty$. Since we don't have MSE_∞ for $\rho_T \in \{0, 0.5, 1\}$, we use $MSE_\infty = MSE_{m=60}$ for all cases. Please note that the pool size reaches 50 instead of 60 due to this implementation. After transformation, Figure 3.5 shows, what the results look like.

We can observe for Figures 3.3, 3.4 and 3.5

1. With rising pool size, the MSE is decreasing.
2. As the true correlation increases, the MSE decreases.
3. The MSE seems proportional to the prior rate α (only in Figure 3.7).

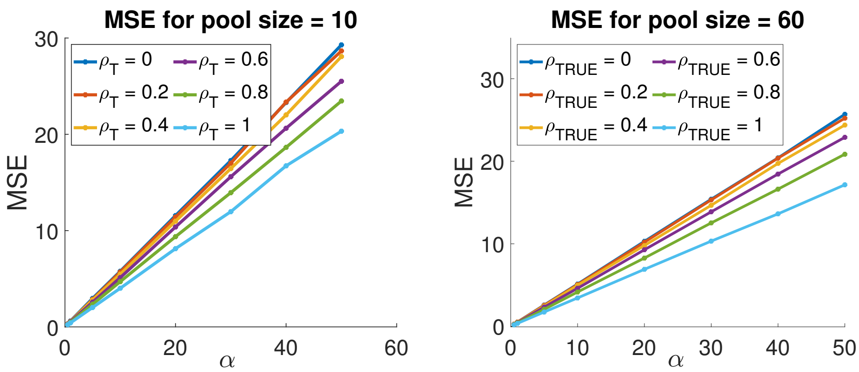


Figure 3.3: The mean squared error against the prior rate for fixed pool sizes of 10 and 60.

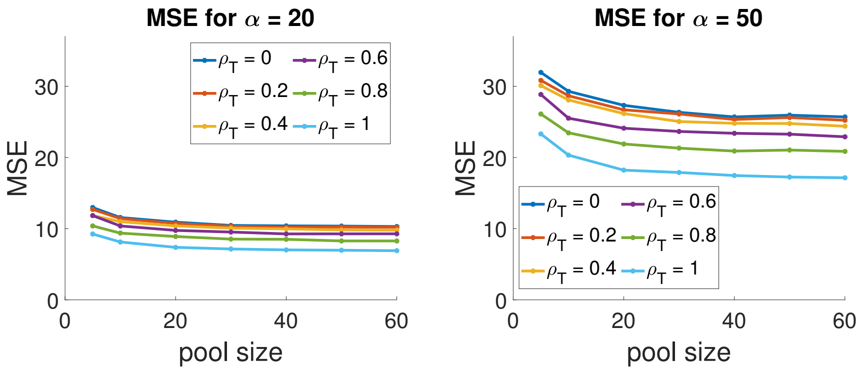


Figure 3.4: The mean squared error against the pool size for fixed prior rate sizes of 20 and 50.

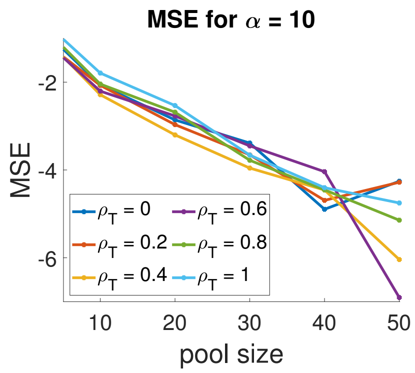


Figure 3.5: The transformed mean squared error against the pool size by Equation 3.24.

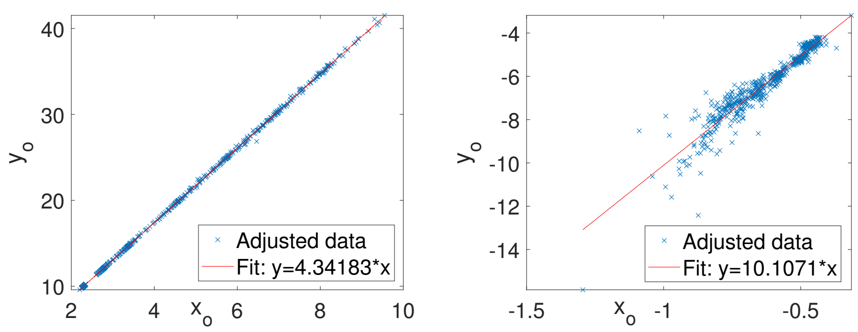


Figure 3.6: Left: the added variable plot from for regression Formula 3.22. Right: the added variable plot for the logarithmically transformed regression formula, Equation 3.23. The variables x_0 and y_0 are specified in Section B.2.1.

α	$MSE_{\infty,\rho=0}$	$MSE_{\infty,\rho=1}$
0.5	0.25	0.166667
1	1/2	1/3
5	5/2	5/3
10	5	10/3
20	10	20/3
30	15	30/3
40	20	40/3
50	25	50/3

Table 3.2: Limits for MSE in the case of no dependency and full dependency. The pool size tends to infinity $m \to \infty$.

3.4.1 Overall MSE Results

We now consider the findings of the simulation study to examine the overall impact of the shape parameter (α), pool size (m) and correlation (ρ) upon the MSE of the estimator. First, we investigate the relationship between the MSE and the shape parameter (α). We find this relationship to be linear. For example, Figure 3.7 shows the MSE as a function of α when the pool size is $m = 20$ and for six settings of the true correlation (ρ) between zero and one. Although not shown here, similar patterns are found for other input combinations. Evidence of a linear relationship between the MSE and α is not surprising given the analytical results shown in Formulas 3.26 – 3.47 which allow comparison to the simulation study.

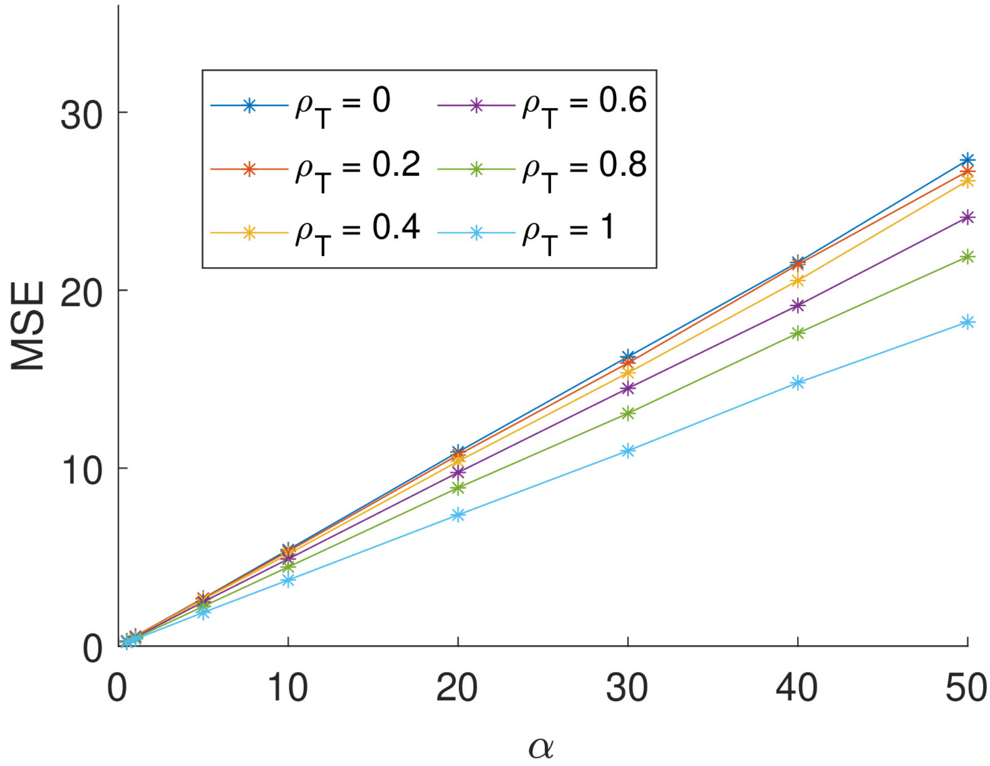


Figure 3.7: Relationship between MSE and the prior shape parameter for pool size = 20 and selected values of the true correlation.

Next, we examine the relationship between the ratio of MSE/α with the correlation and the pool size. Figure 3.8 shows the MSE values computed for simulation combinations together with a model fitted to this relationship. After investigating a variety of transformations to this relationship, we obtain the following expression for the best-fitting model through a regression analysis:

$$MSE = \alpha \cdot (0.705451 - 0.047799 \ln(m) - 0.169848\rho^2) \tag{3.25}$$

Interestingly we find that the interaction terms between ρ and m do not contribute to the model. Moreover, the impact of m and ρ are proportional to the value of the shape parameter α.

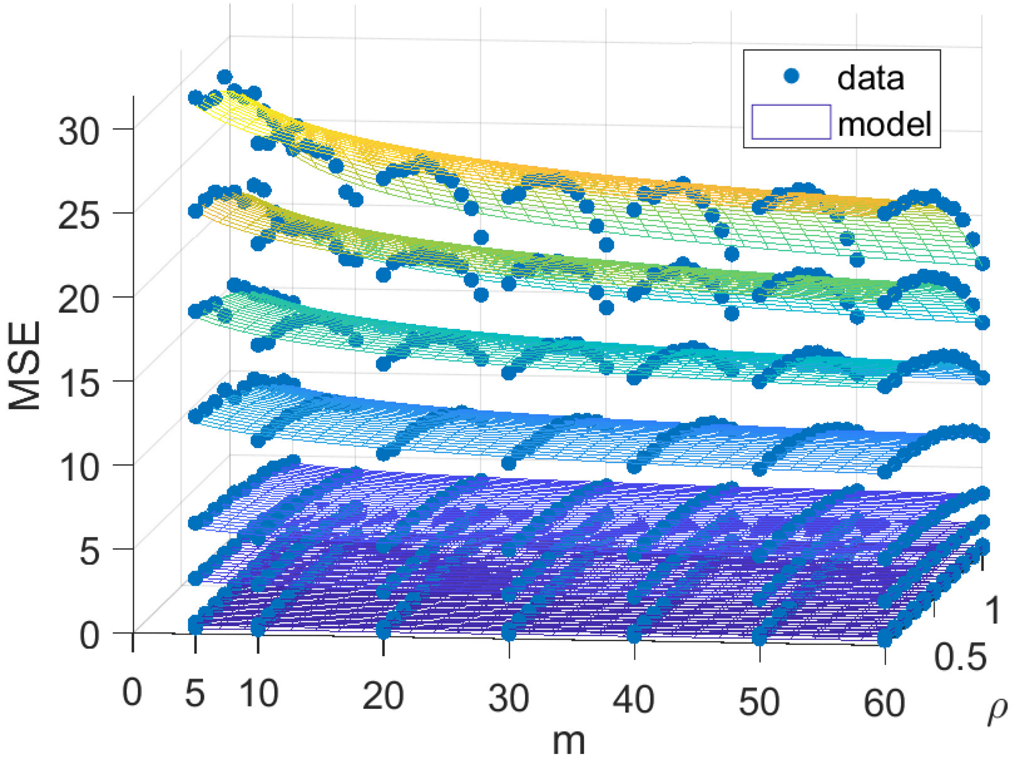


Figure 3.8: MSE at different values of shape parameter α for settings of pool size and true correlation parameter with fitted model of the form $MSE = \alpha(c_0\, ln(m) + c_1\rho^2 + c_2)$

While this regression model is defensible only for the range of parameter values used in the simulation study, we can build upon our earlier consideration of the inference approaches to develop analytical results to provide the limit of the MSE from the general formula grasping all listed cases.

$$\begin{aligned} MSE &= E\left[(E_{\rho_A}[\Lambda|N_1,N_2] - \Lambda)^2\right] \\ &= E_{N_1,N_2}\left[E_{\rho_T}\left[E_{\rho_A}[\Lambda|N_1,N_2]^2 \right.\right. \\ &\quad \left.\left. - 2\Lambda E_{\rho_A}[\Lambda|N_1,N_2] + \Lambda^2 \middle| N_1,N_2\right]\right] \end{aligned} \tag{3.26}$$

leading to an overall general formula for the MSE

$$\boxed{\begin{aligned}=&E_{N_1,N_2}\left[E_{\rho_A}[\Lambda|N_1,N_2]^2\right]\\&-2E_{N_1,N_2}\left[E_{\rho_A}[\Lambda|N_1,N_2]E_{\rho_T}[\Lambda|N_1,N_2]\right]\\&+E_{N_1,N_2}\left[E_{\rho_T}[\Lambda^2|N_1,N_2]\right]\end{aligned}} \tag{3.27}$$

To obtain specific formulas under the different assumption, we substitute into the Formula 3.27 the following terms branching the expression into all listed cases.

$$E_{\rho_A=0}[\Lambda|N_1,N_2]=\frac{\alpha+N_1}{\beta+1} \tag{3.28}$$

$$E_{\rho_A=1}[\Lambda|N_1,N_2]=\frac{\alpha+N_1+N_2}{\beta+2} \tag{3.29}$$

$$E_{\rho_A=0}[\Lambda^2|N_1,N_2]=\frac{\alpha+N_1+(\alpha+N_1)}{(\beta+1)^2} \tag{3.30}$$

$$\begin{aligned}&E_{\rho_A=1}[\Lambda^2|N_1,N_2]\\&=\frac{\alpha+N_1+N_2+(\alpha+N_1+N_2)^2}{(\beta+2)^2}\end{aligned} \tag{3.31}$$

with extending by its' expectations

$$E_{\rho_T=0}[N_1]=E_{\rho_T=0}[N_2]=\frac{\alpha}{\beta} \tag{3.32}$$

$$E_{\rho_T=1}[N_1]=E_{\rho_T=1}[N_2]=\frac{\alpha}{\beta} \tag{3.33}$$

$$E_{\rho_T=0}[N_1^2]=E_{\rho_T=0}[N_2^2]=\frac{\alpha}{\beta}+\frac{\alpha+\alpha^2}{\beta^2} \tag{3.34}$$

$$E_{\rho_T=1}[N_1^2]=E_{\rho_T=1}[N_2^2]=\frac{\alpha}{\beta}+\frac{\alpha+\alpha^2}{\beta^2} \tag{3.35}$$

$$E_{\rho_T=0}[N_1N_2]=\frac{\alpha^2}{\beta^2} \tag{3.36}$$

$$E_{\rho_T=1}[N_1N_2]=\frac{\alpha+\alpha^2}{\beta^2} \tag{3.37}$$

The initiating case states that m tends to infinity and as ρ tends to one. Since an infinite pool size corresponds to the Bayesian estimator, and if the observations are perfectly correlated, then this implies the sample size is doubled in relation to the case of no correlation when processes are statistically independent. Thus we find:

$$\lim_{\rho_A \to 1,\, m \to \infty,\, \rho_T = 1} MSE = \frac{\alpha}{3} \tag{3.38}$$

$$\lim_{\rho_A \to 0,\, m \to \infty,\, \rho_T = 1} MSE = \frac{\alpha}{2} \tag{3.39}$$

The difference between these two limits, 0.17α, is consistent with the coefficient for the correlation in the regression model (Equation 3.25). From Equation 3.25 can be anticipated the value around

$$\lim_{\rho \to 0.5, m \to \infty} MSE = 0.456\,\alpha \tag{3.40}$$

Further analysis reveals the following:

$$MSE_{\rho_A = 1,\, \rho_T = 0} = \frac{7\alpha}{9} \tag{3.41}$$

Substituted by the moments into the expression for the MSE results from having set the value $\beta = 1$ in the following:

$$\begin{aligned} MSE_{\rho_A=0, \rho_T=1} =& \frac{\alpha}{\beta(\beta+1)} \\ =& MSE_{\rho_A=0, \rho_T=0} \\ =& MSE_B \end{aligned} \tag{3.42}$$

This form is as we would expect because we are ignoring the information from N_2 as we are assuming independence. Therefore, incorporating data under the assumption it is realised from the same HPP (i.e. $\rho_A = 1$) when in fact there is no correlation ($\rho_T = 0$) can introduce considerable estimation error depending on the variability of the pool of processes.

A lower value than for no dependency and higher than full dependency is obtained from assuming $\rho_A = 1$ and having in fact neither full nor no dependency, but partial dependency.

$$\lim_{\rho_A=1,\,\rho_T=\rho} MSE = \frac{\alpha}{9}(7-4\rho) \tag{3.43}$$

Assuming dependence when the rates are independent results in the following:

$$MSE_{\rho_A=1,\rho_T=0} = \frac{(\beta^2+4\beta+2)\alpha}{\beta^2(\beta+2)^2} \tag{3.44}$$

Assuming dependence when the rates are dependent results in the following:

$$\begin{aligned} MSE_{\rho_A=1,\rho_T=1} =& \frac{\alpha}{\beta(\beta+2)} \\ =& MSE_B \end{aligned} \tag{3.45}$$

The MSE_B from Equation 3.10. Again this is as we would expect because we are correctly using a Bayesian approach with 2 observations.

However, the regression model does not have the above limits. Consider the situation where pool size is $m = 1$, which is outside the range investigated in the simulation study where the minimum is $m = 5$. When $m = 1$ an empirical Bayes approach is not appropriate because there is no pool of processes from which to estimate the pool variability. Under such circumstances where we have prior information, then we would apply a classical approach as described earlier, which would have the following limits:

$$\lim_{\rho\to 1,\, m\to 1} MSE = \frac{\alpha}{2} \tag{3.46}$$

$$\lim_{\rho\to 0,\, m\to 1} MSE = \alpha \tag{3.47}$$

Therefore, extrapolating our regression model would underestimate the MSE when processes are independent and overestimate the MSE when processes are perfectly dependent.

For addressing the percentiles of the standardized errors in Table 3.3 can be compared to the percentiles of the normal ($z_{10} = 1.28, z_{05} = 1.645, z_{025} = 1.96, z_{01} = 2.33$). The distribution of the residues we can describe as right-screwed data as shown in the histogram and the normal probability plot, Figure 3.9. We can see that the model coincides with Figures 3.7.

Table 3.3: The distribution of the standardized residuals of the regression model (Equation 3.25).

percentile	count	value
0.10%	1	-2.323150
1%	7	-1.786462
2.50%	16	-1.578797
5%	31	-1.429651
10%	62	-1.213216
90%	554	1.145075
95%	585	1.584439
97.5%	600	2.054588
99%	609	3.036317
99.9%	615	4.745185

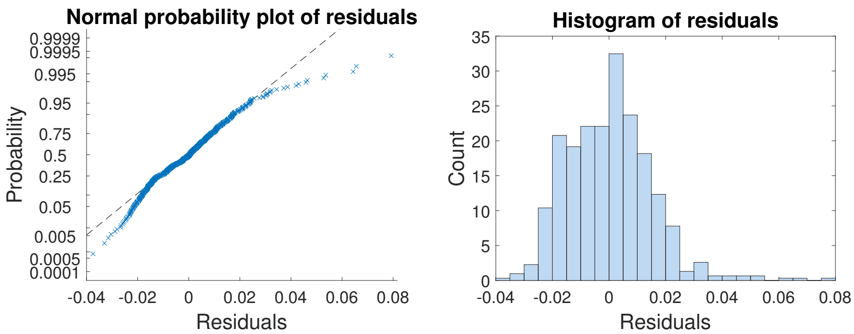


Figure 3.9: Residues of the regression model (Equation 3.25).

3.4.2 Dependency Mis-specification

Let us now investigate the implications of mis-specifying the correlation parameter. We extend the simulation study to explore situations where we assume a correlation $\rho = \rho_A$ has been specified, say by an engineering expert's subjective assessment when the true correlation is actually $\rho = \rho_T$. In the study, we simulate data under the case $\rho = \rho_T$ then analyse it as if it was specified as $\rho = \rho_A$ to mimic the parameter mis-specification. We share a selection of results to illustrate key findings.

Figure 3.10 shows the relationship between the MSE and ρ_A for situations where $\alpha = 50$, $m = 60$ and $\rho_T = 0$, 0.5 and 1. Regardless of the true correlation, we find the same MSE when no dependency is specified ($\rho_A = 0$). However when $\rho_T = 0$ and ρ_A

increase towards 1 then there is a nonlinear growth in the MSE. Whereas for $\rho_T = 1$ there is an almost linear decrease in the MSE and for $\rho_T = 0.5$ there is relatively little change in the MSE as ρ_A increases to 1. Similar patterns are found for other combinations of pool size and shape parameter.

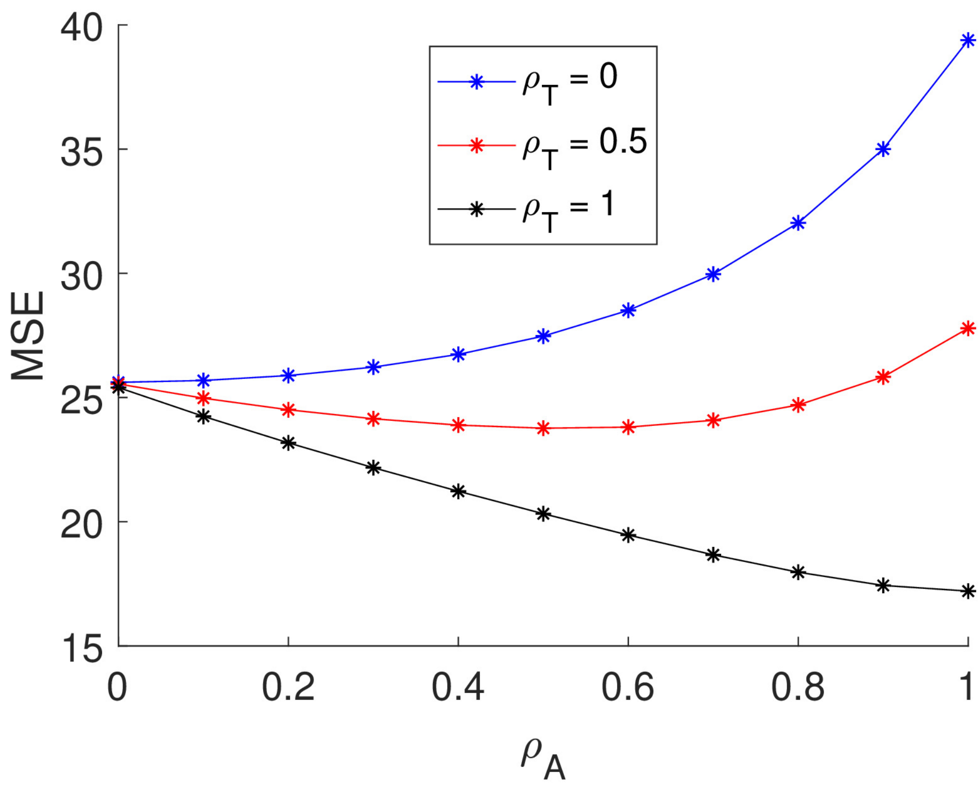


Figure 3.10: MSE when correlation is mis-specified, where ρ_A is assessed correlation, under selected true correlations ρ_T and for case of α=50, m=60.

By assuming a true α value, we can examine the effects of mis-specifying the correlation under either Bayesian inference or for the empirical Bayes asymptotic case when the pool size m tends to infinity. Figure 3.11 shows the relationship between the ratio MSE/α and ρ_A for $\rho_T = 0$, 0.5 and 1 for three settings of the prior shape parameter $\alpha = 0.5$, 10 and 50. We find that the limits agree with the findings discussed earlier (Equations 3.38 and 3.39). That is, for $\rho_A = 0$ then $MSE/\alpha = 0.5$ and for $\rho_A = \rho_T = 1$ then $MSE/\alpha = 1/3$. Figure 3.11 also shows the effects of varying α on MSE/α for $\rho_T = 0$.

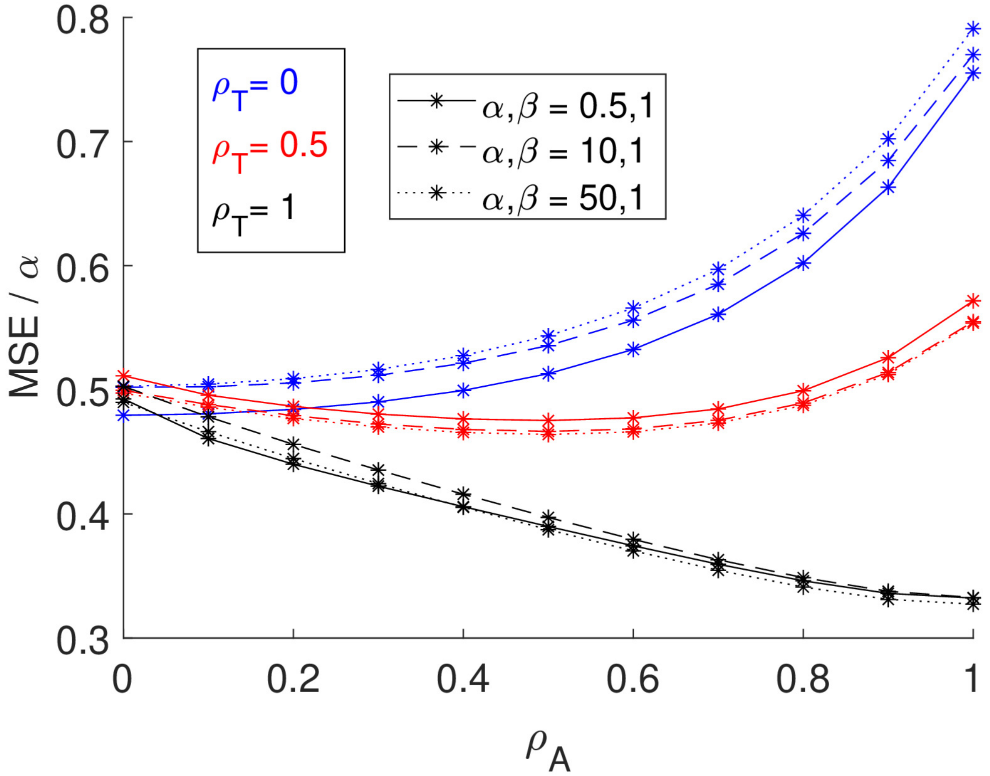


Figure 3.11: MSE/α for assessed correlations, ρ_A under selected values for the true correlations, ρ_T, and prior shape and scale parameters.

To experience, how modelling how maximum percentage changes increases vary dependent on pool size and prior shape, we use the graphical representation in Figure 3.12. Colour coded by four cases of assumed and true dependency scenarios. Modelled as such, we analyse that with increasing pool size irrespective of prior rate the percentage error increases. The more there is available data, the more certainty is present about benefit of correct dependency estimation.

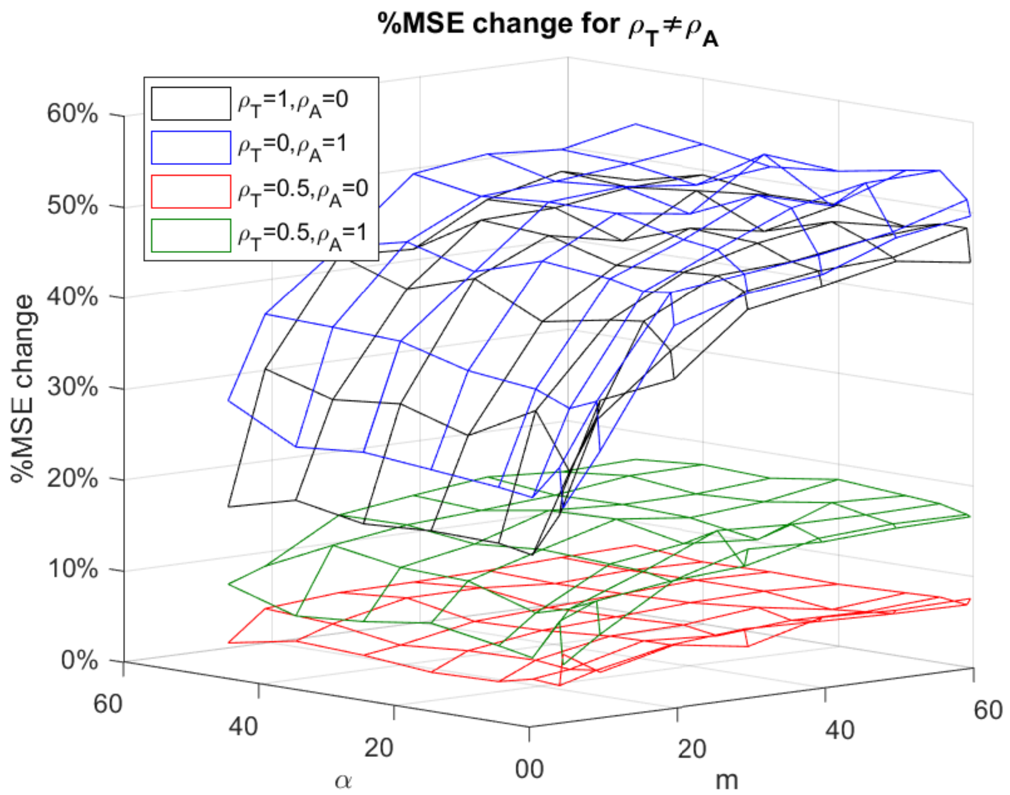


Figure 3.12: The percentage MSE change for selected points of interest for four selected cases of varying assumed and true underlying dependency.

3.4.3 Minimax Estimate

In case, we are reluctant to make any assumptions about the correlation parameter, we arrive at the question: Which correlation value would be the most sensible, conservative option to choose, for a given pool size and prior rate and any additional knowledge?

We suggest using a minimax estimate on both, the absolute (Figure 3.10) and the relative measure (Figure 3.13). The minimax estimator is more suitable for the relative measure.

$$\rho_{sensible} = \underset{\rho \in [0,1]}{\arg} \min \max \% \Delta MSE(m, \alpha) \tag{3.48}$$

, where $\%\Delta MSE(m, \alpha)$ is denoting the curves in Figure 3.13. In plain words, we are taking the lowest point of the roof the Figure 3.13, that is taking the lowers point of a curve, having no curve above. Looking e.g. at Figure 3.13, the question is, whether we should be taking 0.6 or a slightly higher value, where the blue and black curves are meeting? A proper interpolation technique would improve the quality of the curves. We suggest looking at Figure 3.14 to use in most cases the value $\rho = 0.6$. In other words, neither to assume dependency not independence.

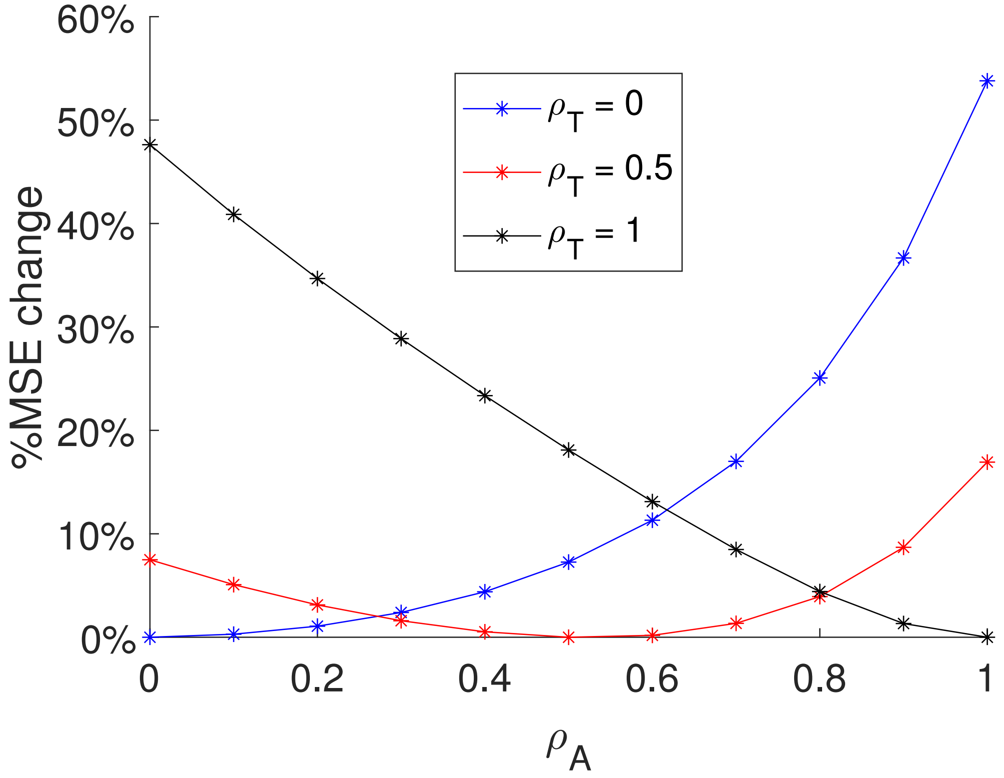


Figure 3.13: Relative measure of the case $\alpha = 50, m = 60$, in close ration to Figures 3.10 and 3.11.

3.5 Summary

Our study has investigated the effect of incorporating data that is correlated with the event process of interest to reduce estimation error. This requires the correlation between processes to be assessed by subjective judgement, which under many circumstances can require a resource-intensive elicitation exercise. Our analysis has shown that that the maximal possible reduction of the MSE is about 53.8% (Figure 3.13).

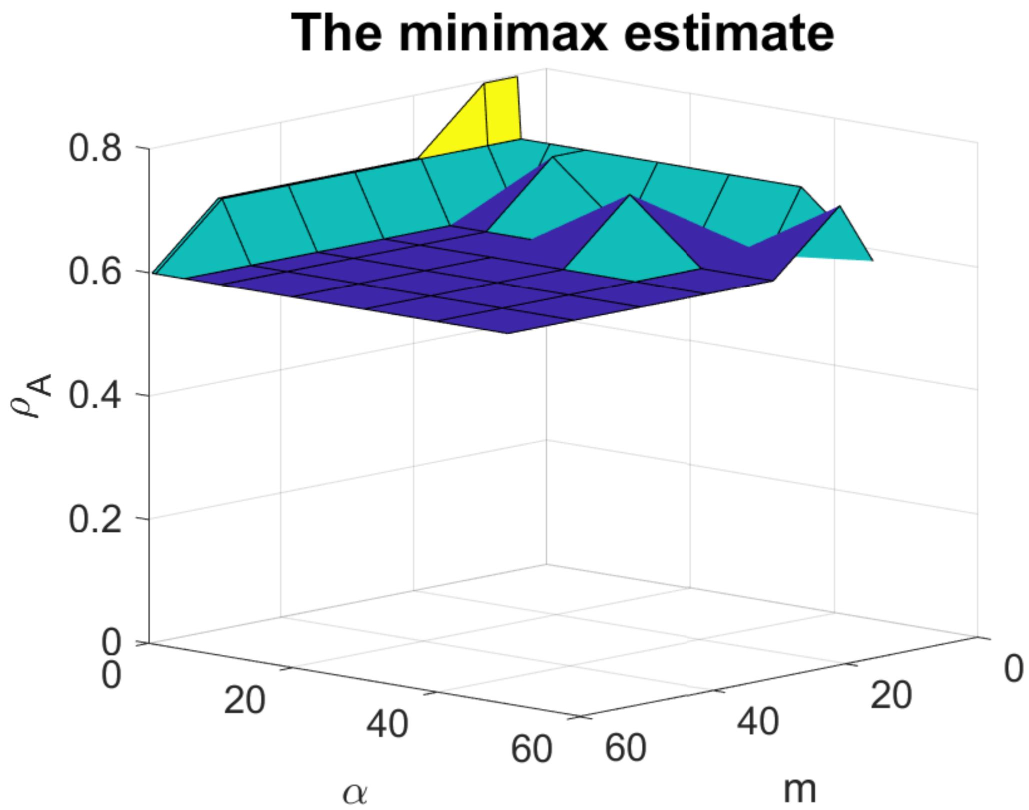


Figure 3.14: The minimax estimate. Advising the value of ρ_A for given values of α, m if the expert possesses no prior knowledge.

This means that if the analyst using the model invests in structured expert judgement to obtain perfect assessments of the dependency between rates, then it is possible to achieve an MSE that is more than half of the value of a maximally mis-specified dependency parameter. Through our study, we have explicated the relationship between correlation, pool size and MSE within the context of a particular probability model to provide insight as to whether gains from MSE reduction are worth the cost of elicitation.

Empirical Bayes is a rich methodology offering the opportunity to gain the benefits of error reduction enjoyed by the Bayesian methodology but without the same elicitation burden for subjective assessment with its recognised associated biases. Empirical Bayes relies on pooling relevant data together. It is well known that the more homogeneous this data pool, then the stronger the inference in the sense that the estimation error will be smaller. One way of homogenising the pool is to assess correlations between processes and so discriminate between degrees of similarity for the events of interest. For example, in our motivating industry case, the events related to failures, the candidate pools were formed by data on events experienced by earlier design generations or on tests for the new system design, and the correlation was assessed by engineering experts via a structured elicitation. However, regardless of how well this elicitation was constructed and managed there still lurks the possibility that the dependency is mis-specified.

Our study has explored the impact of corrupting the inference through mis-specifying the correlation. We have derived a formula to explicate the MSE in relation to the parameters of the marginal distribution under cases of assumed assessed and true correlations. A more general derivation is shown in the Formulas 3.26 – 3.47. Such formulae can inform analytical choices about the incorporation of data from perceived correlated processes by aiding assessment of the consequences. For situations, where no information about possible underlying dependency exists, the analyst is suggested to assume a conservative estimate of the dependency parameter at 0.6. These findings can guide practical choices about the elicitation method selected to support inference about the reliability of a new system design or other applications where a multivariate Poisson-Gamma model is appropriate.

Further, the methodological approach we have adopted to assess an understanding of mis-specification could be applicable to examine the implications for estimation errors for a wider class of probability models.

Is expert judgement elicitation worth the effort? Rather anti-climatically, our answer is “it depends”. It depends on how accurate the results need to be so that the value of elicitation can be assessed in relation to a fuller consideration of the costs and benefits. Costs include not only the time and effort to plan and conduct an elicitation but also the cognitive burden to those providing subjective assessments. For the particular context of our study, the value depends on the potential of the candidate correlated processes. This potential is determined by both the correlation and the characteristics of the marginal distribution, since the benefits of eliciting the dependency are found to be proportional to the shape parameter in our study for a multivariate Poisson-Gamma model.

There are a number of limitations of our study even within the context of HPPs. First, we had a specific form to our multivariate prior distribution. Second, we had limiting data not allowing a full Bayesian solution to be found. Thirdly, we do not have an information framework to evaluate, whether more precise assessments of correlation are worth learning.

Chapter 4

Poisson Models with Copula Priors

4.1 Introduction

The study presents a homogeneous Poisson process in a hierarchical Poisson-Gamma model with copula dependency in the prior failure rate parameter, to examine the consequences of mis-specifying the underlying copula family. Little focus has been directed towards copula mis-specifications in hierarchical Bayesian models. Literature on count data with copula dependency in the prior distribution is found in Section 1.1

We set up our model with dependency structure in the latent variable and look at the copula combinations and percentiles when such copula structure mis-specifications matter. We want to test these results on sparse Poisson data to show the possible application of theory in praxis. We work specifically with a Poisson-Gamma model with varying time and prior shape parameters aiming to comprehend relationships that are affecting the Mean Squared Percentage Error (MSPE) as the precision and percentage bias increase ($\%bias$) estimation as the accuracy. As in theory straightforward, but in practice mostly challenging. The moment estimator, then the percentile estimators. All those parts of the simulation do not have analytical formulas. They need to be approximated numerically.

The bias is constructed in the conducted simulation by comparing the percentile of the posterior using the correct true underlying copula relatively to the posterior constructed adopting a proposed assumed copula. This is formulated as a $\%bias$. We measure likewise the square of this quantity, denoted as the MSPE (mean squared percentage error). The implications are that copulas matter especially in the tails. Next, that it is always better to assume a correct copula rather than a mis-specified one. Lastly, the combinations that matter the most are having $t = 10$ with combining the important pairing combinations. Literature on engineering applications and road incident analysis is found in Section 1.1.

The paper structure is as follows. We provide a thorough description of the simulation study and show the setup for technical insight. The work examines the results and draws appropriate conclusions.

4.2 Copula Choice

Two variables representing rates can be more dependent in one tail and less dependent in the other tail. That is, lower rates have one value of dependency and higher another one. For such situations are copulas well suited.

One case is that lower rates are more correlated and higher rates less. Trivedi and Zimmer (2007) describes the Clayton copula to have weak right tail with a strong left tail. This makes the copula according to the author suitable for applications in risk. The contrary case is a weak left tail, but a strong right tail. These characteristics resembles the Gumbel copula. Another option is that both tails behave similarly. One of the copula choices for our study is the Gaussian copula. It is the most widely used copula, (Hawas and Cifuentes, 2017, p. 94). Especially for finance and insurance risk problems (Fang and Madsen, 2013). For the case, where neither low nor high rates are much correlated. Modelling variables loosely dependent in the upper and lower percentiles, the Frank copula is a suitable choice.

Therefore, we have selected the Gaussian copula being an elliptical copula and 3 prominent Archimedean copulas, namely, the Frank, Gumbel and Clayton illustrated in Figure 2.2 with $\rho_S = 0.9$. Yan et al. (2007) writes that the 2 most used copula families (comprising of many copulas) are the elliptical copulas and the Archimedean copulas.

These copulas have been studied in related research by (Acar et al., 2019; Corbella and Stretch, 2013; Dorey and Joubert, 2005; Dupuis, 2007; Genest et al., 2006) and more. For the purpose of studying left and right tail percentiles of the posterior distribution the right and left tail behaviour of copulas plays a role. When comparing the copulas, Dupuis (2007) and Venter (2002) argue the Gaussian copula to have a lighter right tail than the Gumbel copula, however stronger than the Frank. The Gumbel copula stands opposed to the Clayton copula with respect to tail dependences. It displays a weak left tail and strong right tail dependence. The Frank copula is specified symmetrically, as the Gaussian copula. In comparison to the Gaussian copula does the Frank copula have a relatively weak tail dependence. So the Gaussian copula has heavier on both tails than the Frank copula (Venter, 2002).

To summarise, in the left that the copulas have the lightest tail to the strongest in the order of Frank, Gaussian, Gumbel, Clayton. The right tail is in the order Clayton, Frank, Gaussian, Gumbel (Dupuis, 2007, p. 382). We work with two symmetric copulas, the Gaussian and Frank.

4.3 Method

We create a simulation study for the investigation of the $\%bias$ and the $MSPE$ of the percentiles estimator. With the use of a full Bayesian model, we study the influence of shape parameter of the marginal distribution (α), the exposure time of the HPP and observed percentile (p). The HPP generates a pair with a unique dependency structure. Data are generated given their specified rates. The parameter values we control in the simulation as shown in Table 4.1. The percentile of the posterior distribution does not have an analytical form and hence a Monte Carlo estimator has been used. Without loss of generalisation, we set $\beta = 1$ to be the scale parameter. We choose the time horizon $t = 1$ as yearly value, $t = 5$ used as a short time horizon in investment decision and $t = 10$ accepted as an intermediate time horizon. The values of α are rationalised by properties of the gamma distribution, which changes with the watershed value 1. If the shape parameter is $\alpha < 1$, the mode is zero, and if $\alpha \geq 0$ the mode is positive.

Apart from the basic starting point 1, we choose one value bellow and the next natural choice above. For the percentiles p, a lower bound for failure rate estimation, a most likely value ($p = 0.5$) and an upper bound is examined. The values 0.1 and 0.9 and often used percentiles for tail values. We wish to model the scenario of a moderate to strong dependency value. This suggests the boundary value $\rho = 0.75$.

4.3.1 Algorithm Steps

For given loop in the algorithm, we take the next steps with a given value for α, exposure time t, selected percentile of interest (p) and a generating copula structure of form *cop_T* with a fixed underlying dependency parameter. In each for loop, we select a specific setting from all the available ones. For each unique setting 10,000 copula-dependent pairs following a gamma distribution are simulated. From there, Poisson samples $[N_{1i}, N_{2i}]$ are drawn. For each pair $[N_1, N_2]$ a Monte Carlo calculation of the marginal percentile is performed. Given, we have calculated the true percentile, we can move to the calculation of the assumed percentile under a given assumed copula *cop_A*. After we obtained all values, we can calculate the percentage error vector and from it the bias and the MSE.

Table 4.1: In the simulation study used parameter values

Input parameter	Specified Values
t	$1, 5, 10$
α	$0.5, 1, 2$
p	$0.1, 0.5, 0.9$
ρ_S	0.75

Algorithm 2 Copula algorithm:

Set: $\alpha, t, cop\_A, cop\_T, p$ and ρ_S
for $iq \leftarrow 1$ to N **do**
    for $i\alpha \leftarrow 1$ to N_α **do**
        for $cop\_T \leftarrow Gaussian$ to $Clayton$ **do**
            for $it \leftarrow 1$ to N_t **do**
                Simulate: $[\lambda_{1j},\ \lambda_{2j}] \sim C_{cop\_T}(G(\alpha_{i\alpha}, 1), G(\alpha_{i\alpha}, 1); \rho_S)$
                Simulate: $[N_{1j},\ N_{2j}] \sim Poisson(\lambda_{1j} \cdot t_{it}, \lambda_{2j} \cdot t_{it})$
                Calculate: $\underset{\sim}{q}_{p,cop\_t}[\Lambda_1 | \underset{\sim}{N}_1, \underset{\sim}{N}_2, \alpha_{i\alpha}, \beta, \rho_S]$
                for $cop\_A \leftarrow Gaussian$ to $Clayton$ **do**
                    Calculate: $\underset{\sim}{q}_{p,cop\_A}[\Lambda_1 | \underset{\sim}{N}_1, \underset{\sim}{N}_2, \alpha_{i\alpha}, \beta, \rho_S]$
                end for
            end for
        end for
    end for
end for
Calculate: $\frac{q_{p,cop\_A}[\Lambda_1|N_1, N_2, \alpha, \beta, \rho_S] - q_{p,cop\_T}[\Lambda_1|N_1, N_2, \alpha, \beta, \rho_S]}{q_{p,cop\_T}[\Lambda_1|N_1, N_2, \alpha, \beta, \rho_S]} = \mathsf{e}_1$
Calculate $\%bias = \left(\frac{\sum_{i=1}^{n} \mathsf{e}_{1i}}{n} \right) \cdot 100\%$ and $MSPE = \left(\frac{\sum_{i=1}^{n} \mathsf{e}_{1i}^2}{n} \right) \cdot 100\%$

4.4 Analysis of Copula Mis-specification in a Simulation Study

As anticipated, with the wrong copula the error increases. Therefore, it is best to choose the right copula, which is underlying the data. This claim is supported by the diagonal (Figures 4.4 and 4.5). The diagonal have errors of the smallest magnitude. When choosing the wrong copula, we shall draw conclusions from the analysis in terms of span decrease, symmetry, right tail bias decrease, some individual cases, important pairing combinations and the MSPE. Following the analysis of the *bias* (Figure 4.4):

Span decrease: For some prior setting and time span, the results converge - the dispersion decreases, when we move from the 10th percentile over the median to the 90th percentile across all copulas. The possible span of observed values is the largest in the 10th percentile. In the 50th percentile less as the values close in. Focusing on the

90th percentile observed values do not differ that much. This is especially visible from the combinations *Gaussian_A-Frank_T, Gaussian_A-Clayton_T, Frank_A-Gaussian_T, Gumbel_A-Frank_T, Gumbel_A -Clayton_T, Clayton_A-Frank_T, Clayton_A-Gumbel_T*.

Symmetry: We are able to observe some degree of symmetry. If we switch the assumed copula with the true copula, the pattern in the chart (Figure 4.4) seems to switch broadly upside down. This means, if an increase was observed moving across quantiles in the $\%bias$, then for the case of interchanging the true underlying copula with the assumed copula, a decrease will be observed. The behaviour might be suggesting that the combination of the copulas in combination with the underlying copula structures plays an important role. The symmetrical claim is stronger for the area between the 50th and 90th percentile rather than the 10th and 50th. Up to the 50th is generally no visible trend and from the 50th percentile, we observe an increase/decrease.

Some selected cases: *Gaussian_A-Frank_T* and *Frank_A-Gumbel_T* for $t = 1$, there is a lack of noteworthy trends, the $\%bias$ is around 0. For other t's in the *Gaussian_A-Frank_T* jointly with the *Gumbel_A-Frank_T*, the bias is gradually decreasing across percentiles in almost all cases. In $t = 10$ are the charts for *Gaussian_A,Frank_T*, *Gumbel_A,Frank_T* and *Gumbel_A-Gaussian_T* decreasing in the order (from the highest to the smallest) of $\alpha = 0.5, \alpha = 1, \alpha = 2$.

Important paring combinations: for the 10th percentile, a significant combination in large observed values is for $t = 10$ *Gaussian_A-Frank_T, Gaussian_A-Clayton_T*, *Gumbel_A-Frank_T* and especially *Gumbel_A-Clayton_T*. This combination can be expected, since the Clayton copula has a strong left tail and the Gumbel has a weak left tail. The cases can be followed in Figure 4.1, denoted by red boxes. In each square are values for three prior settings. The cases are colour coded. The absolute magnitude of the increase is separated into three categories $< 3\%, 3\% - 10\%$ and $> 10\%$ highlighted with a fill of green orange and red. With a red border for all cells containing a large increase. If an analyst does not know which copula of the 4 examined has been characterizing the generating process and would wish such a choice that maximal errors are possibly low. If we assume the worst case. Then the analyst would choose the Frank copula. Vice versa, if stating the question, which copula would have the largest maximal errors, the Gumbel copula would qualify. The green liming indicates the risk-averse of the Frank copula and orange liming the risky choice of the Gumbel copula as shall be explained in the following paragraph. The red liming shows combinations with highest increases. In particular, for *Gausian_A-Clayton_T* for $t = 5, 10$ from the 10th to the 50th percentile and for all remaining cases, the bias will decrease with increasing dependency. On half-way between medium and high importance situated the *Gaussian_A-Frank_T*. This is since for $t = 10$ in Figures 4.1 and 4.2 some values are above 10% and some bellow. Of medium importance are the cases *Frank_A-Gaussian_T, Frank_A-Gumbel_T, Frank_A-Clayton_T, Clayton_A-Gaussian_T* and *Clayton_A-Gumbel_T*.

The 10th percentile has the cases with the largest percentage increase, *Gumbel_A-Clayton_T* (29%), *Gaussia_A-Clayton_T* (21%), *Gumbel_A-Frank_T* (17%). All at $\alpha = 1, t = 10, p = 0.1$. For most cases, the maximal increase in the 10th percentile is showing a higher mis-specification impact than the 90th percentile (Figures 4.4 and 4.5).

Risk Averse Choices: For the 10th percentile, let us assume, the analyst chooses the Gaussian copula, the highest error can occur, if the underlying copula is Clayton. Similarly for the Frank copula is the largest absolute error in the Gumbel copula (7.634%). For the Gumbel copula in the Clayton copula (29.347%). If we assume a Clayton copula, the largest error occurs under the true Gumbel (8.198%) followed by a Gaussian copula (8.198%). Therefore, the safest choices for an analyst are Frank (green box in Figure 4.1) and then Clayton, as indicated in the earlier paragraph. Because the worst scenario will be a true Gumbel copula (orange box in Figure 4.1) followed by a true Gaussian copula. The Frank copula does not introduce high dependency is the left tail, whereas the Gumbel does.

Time aspect When the time t is changed, some features (titles of earlier paragraphs) prevail and some are not so imperative any more. We highlight of 4 important combinations for $t = 10$, 3 important combinations for $t = 5$ and that *Gaussian_A-Clayton_T*, *Gumbel_A-Frank_T* and *Gumbel_A-Clayton_T* case in Figure A.1. Again, we claim that the safest choice for an analyst is the *Frank_A* copula. The riskiest choice remains the *Gumbel_A* copula. Under $t = 5$ as the Figure A.2 shows, we can just advise to using the risk aversive choice of *Frank_A* copula.

MSPE: For the MSPE, as the percentile increases, the MSPE decreases. This is true for all cases, except *Gaussian_A-Gumbel_T, Frank_A-Gumbel_T.* This decrease is larger between the 10th and 50th percentile than between the 50th and 90th percentiles. Here is the symmetry present in the way that if we mention one combination, the second combination is likely to be mentioned. Between the 10th and the 50th percentile, all cases experience an MSPE decrease. For specific cases, it is worth looking at the *Clayton_A-Frank_T* combination. The MSPE decreased from the 10th percentile to the median without almost any further decrease. In contrast the *Gaussian_A-Clayton_T* case decreases in the same manner along the percentiles, being also among the largest absolute decreases. That is to say a maximum of 28.88% reducing to 7.16%. Most other cases are halving their descent at the median. The maximal absolute decrease of 60% to 14.49% in the *Gumbel_A-Clayton_T* case. Followed in the third place by *Gumbel_A-Frank_T* (24.67% to 6.31%). In some cases, the $MSPE$ is almost proportional to α. *Gaussia_A-Frank_T*, *Frank_A-Clayton_T*, *Gumbel_A-Gaussian_T*, *Clayton_A-Frank_T*, *Clayton_A-Gumbel_T.*

Important MSPE paring combinations, risk averse choices and time aspect: The conclusions are identical to the bias. For $t = 10$ of importance in the 10th percentile *Gaussian_A-Frank_T, Gaussian_A-Clayton_T*, *Gumbel_A-Frank_T* and especially *Gumbel_A-Clayton_T* following Figure 4.2. All other conclusions observed by the *%bias* hold here, by the case of the *MSPE.* Including the risk averse choice of a Frank copula (5.197%) and the risky pick of the Gumbel copula (57.545%).

Sensitivity on dependency In Figure A.5 the symmetry feature is present. That is under $\rho_S = 0.5$. It is not that explicit as by $\rho_S = 0.75$, but still we can predict changes between the 50th and 90th percentile.

The symmetry is still present by $\rho_S = 0.25$, nevertheless not that much as earlier. Some cases are already mismatching their assumed behaviour, Figure A.7.

For examining the span decrease under changing underlying dependency, we generally conclude that the less dependency is present the smaller the span decrease, when moving from the 10th, over the 50th to the 90th percentile. The case *Gaussian_A-Frank_T* is pretty much unchanged for $\rho_S = 0.5$ and $\rho_S = 0.25$. The very same can be claimed about *Frank_A-Gaussian_T*. *Gaussian_A-Clayton_T* curves are gradually flattening for $\rho_S = 0.5$, $\rho_S = 0.25$ as they move from the 10th over the 50th to the 90th percentile. So there is a span decrease happening for $\rho_S = 0.5$, but not any more for $\rho_S = 0.25$. *Gumbel-_A-Frank_T* for $\rho_S = 0.5$ is still the decrease present and for $\rho_S = 0.25$ it is gone and so by *Gumbel_A-Clayton_T*. By the case of *Clayton_A-Frank_T* and *Clayton_A-Gumbel_T* does the lingering tendency of the decrease remain even at $\rho_S = 0.25$.

In Figure 4.3 on the right-hand side for assumed copulas column-wise and true underlying copulas row-wise, the characterised features from the results mainly for the %*bias*. From preceding paragraphs this would be span decrease as 10-90 convergence, the behaviour between the 50th and 90th percentile as 50-90 decrease or increase and then large increases in the 10th percentile.

	t = 10	TRUE			
		Gaussian	Frank	Gumbel	Clayton
ASSUMED	Gaussian	0.710%	10.552%	-2.343%	20.988%
		0.971%	9.873%	-1.455%	12.424%
		0.893%	6.869%	-1.042%	10.185%
	Frank	-6.013%	0.711%	-6.436%	5.642%
		-5.713%	-0.219%	-7.634%	3.059%
		-6.023%	-1.250%	-5.369%	3.039%
	Gumbel	3.510%	17.086%	-0.507%	29.347%
		2.549%	13.309%	-1.007%	19.375%
		1.868%	10.235%	-2.115%	12.489%
	Clayton	-8.198%	-1.173%	-6.475%	0.613%
		-7.304%	-1.428%	-8.307%	-0.684%
		-7.714%	-2.307%	-6.712%	-0.042%

Figure 4.1: Important combinations for percentage increases of the $\%bias$ in the 10th percentile. Three values bellow each other in the cell are for $\alpha = [0.5, 1, 2]$ with $t = 10$. The assumed copula is specifying the row and a true copula indicates the column.

	t = 10	TRUE			
		Gaussian	Frank	Gumbel	Clayton
ASSUMED	Gaussian	0.047%	8.592%	1.605%	28.885%
		0.071%	4.229%	1.067%	7.932%
		0.054%	1.972%	0.577%	4.239%
	Frank	2.502%	0.118%	5.197%	2.432%
		1.915%	0.096%	4.238%	1.160%
		1.385%	0.062%	3.096%	0.406%
	Gumbel	2.275%	24.666%	0.020%	57.545%
		1.139%	10.659%	0.051%	21.140%
		0.764%	5.422%	0.135%	6.856%
	Clayton	6.139%	3.095%	12.237%	0.210%
		3.608%	1.641%	7.058%	0.053%
		2.328%	0.338%	3.955%	0.091%

Figure 4.2: The MSPE of the 10th percentile with colour coding comparable to Figure 4.1. The results for exposure time $t = 1, 5$ and shown in the Appendix in Figures A.4 and A.3.

	Gaussian_A	Frank_A	Gumbel_A	Clayton_A
Gaussian_T		10-90 convergence 50-90 increase	50-90 decrease	50-90 increase
Frank_T	10-90 convergence 50-90 decrease 10th perc. large increase	risk averse choice	10-90 convergence 50-90 decrease 10th perc. large increase	10-90 convergence 50-90 increase
Gumbel_T	50-90 increase	50-90 increase	risky choice	10-90 convergence 50-90 increase
Clayton_T	10-90 convergence 50-90 decrease 10th perc. large increase	50-90 decrease	10-90 convergence 50-90 decrease 10th perc. large increase	

Figure 4.3: The main analysis points from Chapter 4.4 on the $\%bias$.

4.5 Discussion

The work is characterized by mis-specification of the copula structure within the prior marginal distributions. With respect to other research, might be the closest to our research Park et al. (2021, p. 3), who create a dependent variable ν in the prior distributions and is using only the Gaussian copula. Whereas we work with 3 more additional copulas. In the area of dependency in the prior parameter for a hierarchical model with Poisson count data. The authors observe the advantage of the method especially in the variable fatal crashes, when compared to the univariate Poisson-Gamma mixture model (Park et al., 2021, p. 7).

Similarly Caliendo et al. (2021) models the log of the failure rate $\log(\lambda_i^m)$ with a model involving $\boldsymbol{\beta}_{j[i]}^m$ that is coupled in a Gaussian copula. The authors draw conclusions about the bivariate variables (severe and non-severe) with the use of independent regressors such as daily traffic per lane or tunnel length.

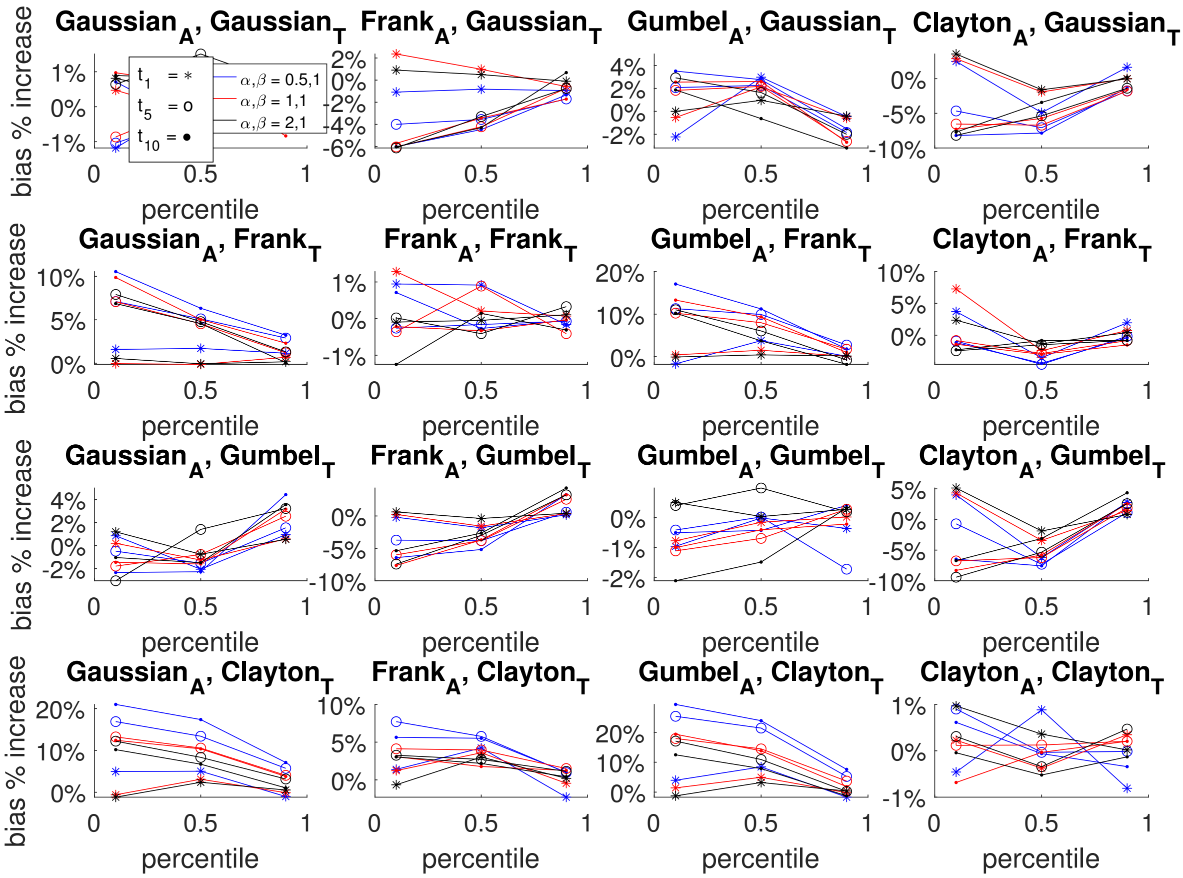


Figure 4.4: The simulation results for the $\%bias$: $(q_A - q_T)/q_T$

4.6 Summary

This chapter focused on the understanding of the mis-specification of copula functions. We aspired to understand better how $\%bias$ and $MSPE$ of the copula mis-specified percentiles will be affected. We show in this work that with the knowledge of the prior parameters, the knowledge of the dependence parameter and the specification of the underlying copula, we are able to estimate the percentage error that occurs under the mis-specification of the assumed copula.

The combination we judged as important in the left tail, namely the 10th percentile are the *Gaussian_A-Frank_T*, *Gaussian_A-Clayton_T*, *Gumbel_A-Frank_T* and

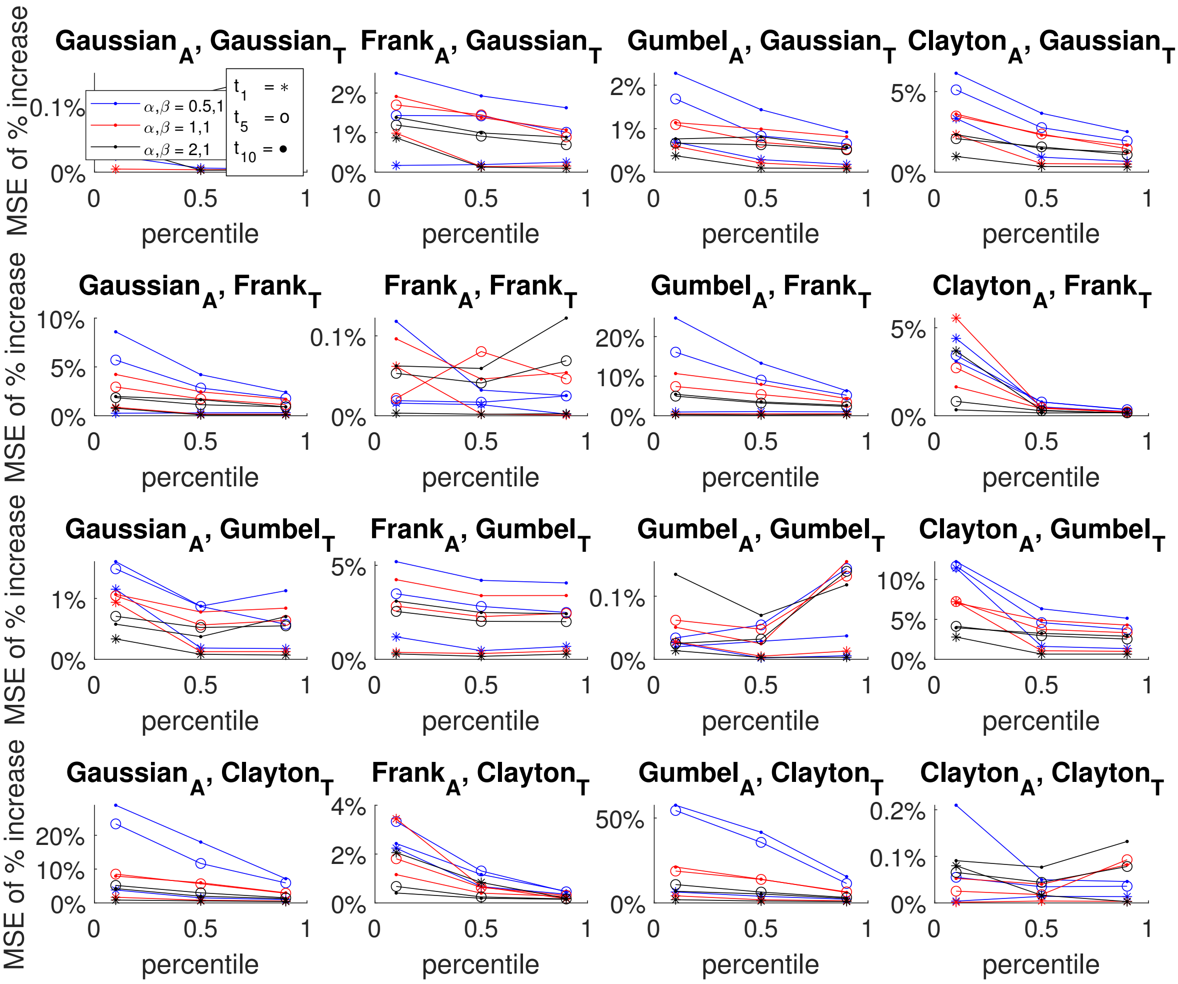


Figure 4.5: The simulation results for the MSPE: $((q_A - q_T)/q_T)^2 \cdot 100\%$

Gumbel_A-Clayton_T case. The *Gumbel_A-Clayton_T* combination has a strong justification in the Clayton copula having a strong left tail and the Frank a weak left tail. If we do not have any information about the underlying copula, the safest choice for the 10th percentile is to choose a Frank copula, followed by a Clayton copula. For the 10th percentile is a colour coded classification designed as shows Figure 4.1.

For an analyst, the safest choice is to select the Frank copula. Owing to the choice selection under which the least error would occur and having an absolute maximum error of 7.634% for the $\%bias$ and 5.197% for the $MSPE$, which is the maximal misspecification we may obtain while the true copula would be Clayton. In second place would be the Clayton copula (8.269% for $\%bias$ and 8.307% $MSPE$). With maximal absolute errors under a true underlying Gumbel copula.

Chapter 5

Accounting for Dependency in Failure Estimation: A Case Study of an Industrial Machine

5.1 Introduction

Spare part management plays a high-importance role across industries. Especially in turbulent time like this with global supply chain disruptions. We want to show that it is worthwhile including partially dependent data in an industrial setting for the proposed case to achieve error reduction, when taking into account data from processes with correlated failure rates. Forecasting models play a vital role in spare parts management. Given the large value of funds and material that a spare parts program can be requiring, precise prediction estimates are a valuable asset. The chapter aims to forecast the number of needed stored spare parts for servicing equipment. To show different scenarios under different applied estimation methods. With the main focus on predicting future production stops, as well the number of failures for supporting spare part management and decision makers.

Using data from industrial machines, we seek to improve estimation of failure rates; for given failure causes by taking advantage of similarities between two machines. We claim that the count of observed failures for a failure cause on one machine (Table

5.1) is similar to the count of failures for the same failure cause on the other machine. Precisely, their rates are similar. A copula based estimator will be applied and the findings will advocate the support for its usage. We want to show that it is worthwhile including partially dependent data in an industrial setting for the proposed case. The data is parted into two. The training set, used to estimate the prediction, and the test set, containing the actual data. We apply an inference method taking into account correlated event rates that is based on the empirical Bayes method. In addition, we implement a similar copula based method. Result are verified via cross-validation. The failure rate estimation impacts decision making. A decision problem is shown, how more accurate estimation leads to better management strategies. Real data with professional support regarding data clarification was kindly provided by Siemens Analyse.

In the final cross-validation, we take the original dataset, divide it into 5 parts and leave always one part out in order to predict it. This means dividing the dataset into 5 parts, using 4 parts for eliciting failure rates estimates, and comparing the predictions with the actual data in the remaining fifth part. The process is repeated five time and at the end averaged.

5.2 Description of Organisation and Data

The company is providing industrial partners with tools for four main areas. Increase of production productivity, reduction of production costs, saving resources and investment and online overviews. Increase of production productivity focuses on analysis of operating states, reduction of downtime, quick identification of problems in production. Reduction of production costs details control of production cycles, increase production quality, optimization of the production process. Saving resources and investment specifies monitoring the service life of machines and tools, human resources load analysis and energy monitoring. Online overviews provide automated reporting and applications for overviewing online environments.

The company specialised in the field of condition monitoring. That is to monitor the state of machine through a combination of sensors in real time with a focus on predictive maintenance. Meaning to provide better information of the state of condition and to advise, when maintenance should be conducted. The aim of condition monitoring for the company is detailed in six areas. OEE (overall equipment effectiveness) & production analysis, predictive maintenance, preventive maintenance, cutting tools, integration reporting and energy monitoring. OEE & production analysis is concerned with analysis of operating conditions, manufactured pieces, comparison of cycles and shifts and traceability of production. Especially their optimization. Preventive maintenance gives timely and realistic predictions of machines failures using modern calculation methods. In order to solve failures on time, order spare parts and reschedule production. Preventive maintenance analyses alarms on the machine, workloads, operating hours and is concerned with component and part exchanges. Cutting tools examine the use of individual machine tools. The manual correction by using tools to eliminate NOK pieces and machine breakdowns. Integration offers support for systems like SAP, MES, ERP. Energy monitoring oversees the consumption of electricity, compressed air, water and others.

This research aims to give a contribution to predictive maintenance and would be likely to have implications is addition on preventive maintenance in form of narrowing the strategies for stocking and re-planning scheduled maintenance check-ups.

5.3 Objectives

The purpose of the model is to estimate when failure may happen for a specific industrial machine in light of data about a similar machine. This supports reliability concerns about testing and analysis and also future development and decision planning. We wish to decrease the estimation error in future failure rate prediction for improved cost and reliability assessment.

We aim to improve future failure rate prediction. To create a good estimator that remains positive, given that this estimator produces sometimes estimates being equal to zeros, which is undesirable (Quigley et al., 2007, p. 619). The classical approach would take the ratio of the number of failures divided by the exposure time to failure. Despite having desirable asymptotic properties, as we suspect, if there are missing failures, the estimate equals 0. Characteristic for situations with few observations of failures in the data available. For the failure rate estimation of failure causes of machine A is used the data from machine A. Additionally, since rates of failure causes are dependent, data from machine B. This is done in an appropriate fashion using copula based estimators and the empirical Bayes estimator. In this way we include information generated from similar failure rates of machine B in the estimation process of machine A. The similarity of failure causes between machine A and machine B is well visible from Chart 5.2. The narrative suggests if on one machine has a failure cause just a low failure occurrence, then as well the failure occurrence on the other machine will be low. And vice versa, if a failure cause will have a high failure rate on one machine then a similarly high occurrence will be observed on the other machine as shows Table 5.1. For example, if a specific electrical failure on one machine occurs let's say more often, then as well it will occur more often on the other machine. And if a given hydraulic failure occurs less often on one machine, then it will be as well the case on the other machine.

The contribution would help to support the area of spares resourcing. There is need to be enough parts provided, as there is need to have enough parts in stock, however it is not desirable to have extensive "tied up capital in the inventory", (Lincoln, 2000, p. 1). We may ask ourselves how many spare parts should be ordered? Zahedi-Hosseini et al. (2017) show that policies can be adopted using data and simulations. As Qarahasanlou et al. (2018) point out, correct estimation is desired, in order to decrease business risk and estimate business and production performance. Qarahasanlou et al. (2018, fig. 1) highlights the concerned areas in the maintenance topic that are expected. Reaching from resources, tools, spares to accessibility. From an OR perspective (Hu et al., 2018) helps to reduce equipment downtime. The spares provisioning area is a complex problem. de Almeida (2001) applies a multiattribute utility function to make trade-offs between contract cost and system performance. To have a stocking strategy in place proves especially crucial for essential parts, "the stocking strategy is dependent on time and the actual occurrence of failures" (Geurts and Moonen, 1992, p. 43). The basis of the stocking strategy consists of finding the optimal number of repairmen and spares (Wang and Sivazlian, 1992).

Our objectives are:

1. To illustrate the process of effectively modelling dependency using empirical Bayes for a real industry data set.

2. To illustrate the effective modelling of dependency using a copula for this industry case.

3. To examine the effects of ignoring or wrongly characterising dependency for this industry case.

5.4 Data Description

The data entries originate from the machines used for the milling of blades that are forming the core of energy-generating gas turbines or producing general mechanical components. Specifically, from the PLC (Programmable Logic Controller) machine, controllable via a PC touchpad (touchscreen) with SAP software operating the machine. The operator himself has responsibility for the entries and signs into the machine with her/his card for manual entries. The data can be characterized as logs. The records are saved alerts about warning messages and mainly errors that resulted in a repair action. This study is limited to the failures resulting in repair action and production stop. We have many possible combinations that could be investigated to indicate dependency in the underlying rates as explained in Chapter 5.5. With this respect milling machines A and B show promising features as analysis will proceed. Mainly by examining underlying correlation (Quigley et al., 2013, p. 2218) analysis reveals in 5.5 the underlying rates of the machines to exhibit dependency. The main 10 failure causes are depicted in a Pareto chart in Figure 5.1. The data generating process is further discussed in Section 5.5, where the underlying generating process and HPP of the machines in visualizations of Figures 5.2 and 5.3 are explained.

The machines have in total 52 failure causes and are categorises by failure groups; Cooling, Electronic failures, Hydraulics, Machine failures, Mechanics, Pneumatic system, Electrical system. For each machine, we observe a “similar” count of failures for the failure causes being shown in Table 5.1. If on one machine are large failures, then we can anticipate large failures on the other machine. And vice versa. The two machines are summarized in the summary statistic in Table 5.2. We observe the data over the time period 23-Oct-2020 to 26-Jan-2022 equal to 461 days equal to 1.263014 years. There are in total 32 failures on machine A and 36 failures on machine B. To split the data into a training and validation set, we use failures that occurred before 14-Oct-2021 and exposure times that lead to these failures as training data and the remaining data as a validation set that we will compare against.

Information is collected about the interaction activities between the operator and the machine and also about maintenance activities. To repair a machine, an order is put in place by an operator. Requesting a certain operation to be done. The repair activity proceeds, while taking note of the workplace the machine is placed in. The equipment is noted that is needed for a repair like for example a crane, if some heavy equipment needs to be transported. Or simply just a hammer. Performing a certain operation is described in a variable called operation. The duration of the repair operation is noted. And so are noted the waiting times for the repair personnel to arrive and to put a machine back into working condition. Leading to the total time from flagging the failure to bringing the machine back into working condition. There are some workplaces that are more expensive to operate than others and the record of the workplace will help the management address the costs of the repairs. The machines in the investigated study group are manufacturing two main products, mechanical equipment and blades. Machine A focuses on mechanical equipment, whereas machine B focuses on blades.

Blades produced by machine B are used in power plants. Initially, water is heated until temperatures mount to a threshold, where water is converted to (superheated) steam. The most common heat sources are coal, nuclear, gas and sun. The speed of the steam entering the turbine needs to be regulated with a governor preventing damage to the turbine. The hot gas is pushing through the turbine blades giving them momentum and making them spin. The blades are attached to the drive shaft, which causes the generator to rotate and produce electricity.

Machine A and machine B are milling machines. A milling machine is a machine tool with a rotating cutter. Typically used to slot drill and form workpieces. The nozzle moves around the workpiece and removes material similarly to a sculptor creating a sculpture from a block of marble.

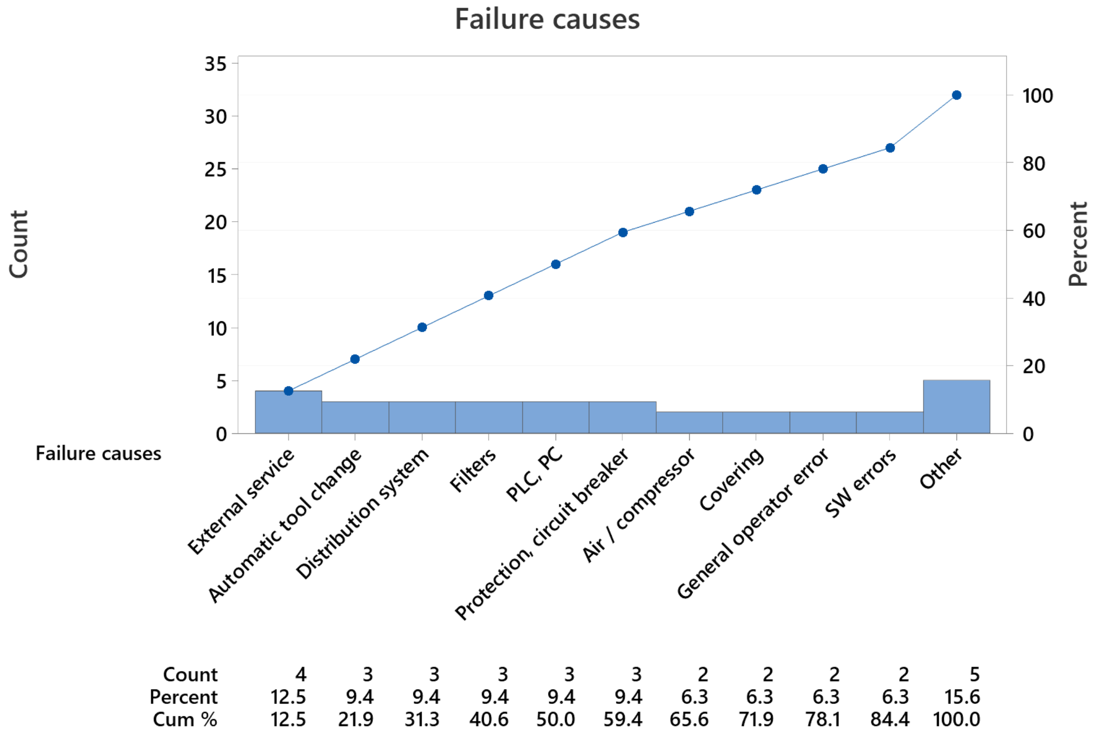


Figure 5.1: Proportion of main failure causes of machine A.

failure cause	machine A	machine B
Air / compressor	2	1
Automatic tool change	3	2
⋮	⋮	⋮
Spindles	0	1

Table 5.1: Count of failure causes of machine A & B.

Machine	Min.	1st Quantile	Median	Mean	3rd Quantile	Max
machine A	0	0	0	0.615385	1	4
machine B	0	0	0	0.692308	1	6

Table 5.2: Summary statistics of count of failure causes of machine A & B.

5.5 Assumptions and Inference Framework

Modelling problem statement: In this modelling framework we make use of the assumption that similar machines can expect a comparable mean failure rate across their failure causes. To introduce the underlying assumed data generating process that we will discuss, a bivariate Gamma distribution produces a pair of rates for a given failure cause. One rate for each homogenised machine. These generate failure counts following the assumed Poisson distribution. As tests show in Section 5.5.1 and 5.5.2 assuming a homogeneous Poisson process (HPP) is justified. These steps are repeated for each failure cause as shown is Figure 5.2. The equipment is assumed operating under normal similar conditions all the time.

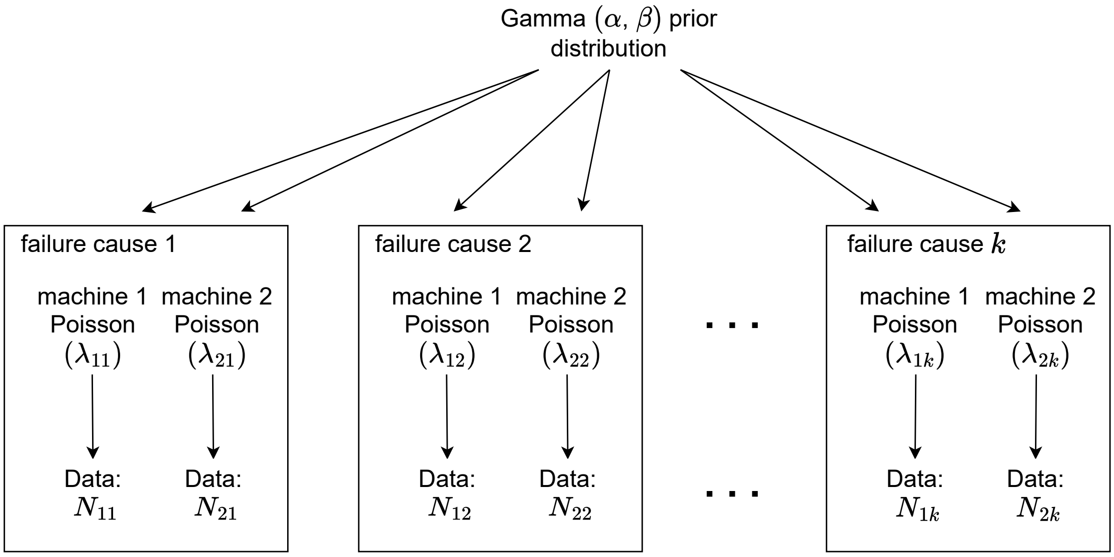


Figure 5.2: Underlying data generating process for support of the modelling framework.

Input data: The input for the analysis represents a list of potential failure causes for machine i that occurred or also did not occur. We index each fault type with a number k. Examining the expected failure rate and count of fault type k of a machine over the exposure time. The particular rate λ_{ik} of the ik-th process can be expressed with the conditional probability density for the realisations observed over the time window

t expressed as:

$$[N_{1k}, N_{2k}] \sim Poisson(\lambda_{1k}t_1, \lambda_{2k}t_2) \tag{5.1}$$

the failure rates λ_{ik} follow a prior distribution. Therefore N_{1k}, N_{2k} are conditionally independent given $\lambda_{1k}, \lambda_{2k}$. The number of events N_{1k} is a sum of observed failure occurrence accumulated over time t on the given machine of considered type k How failures occur is shown in Figure 5.3. Par example, after occurrence of the failure at t_1 that occurs on machine A of failure cause type k, t_2 happens on machine B with the same failure cause. And since these machines are having the same failure causes, which we are investigating, we can pool them into one Poisson process, if we assume the processes to be fully dependent.

HPP: For the study are events assumed to follow a homogeneous Poisson process (HPP) tested in Section 5.5.1. The assumption of a homogeneous Process implies that the machines do not age over the observed time period. The assumption implied, the rate of occurrence of events being constant over time for a given failure cause. We consider that the situation is represented by a collection of HPP. In theory possible, in practice rather rare. Since everything (except software) is in the real world in the long term ageing and therefore the homogeneous assumption does not hold. The assumption of HPP might be more and less suitable dependent on which type of failure cause we are looking at. We are unable to provide a practical check given the reasons at end of Section 5.5.2. Taking as an example a car water pump with a constant failure rate over the whole lifetime and a higher wear out phase (Bertsche, 2008, p. 27). Or an old steam engine type component with a typical bathtub curve.

Homogenisation: Homogenisation by expert opinion adjusts the failure rate so that failure rates between machines are comparable. $h_i > 0$ are used as homogenisation factors so that we assure to hold the assumption of identical marginals from a multivariate Gamma distribution. The homogenisation is to proportionally scale the rate of occurrence controlled by the α, β parameters. We consider that the homogenisation factors were elicited via expert judgement. See for further discussion (Quigley et al., 2011; Quigley et al., 2011, p. 2212).

Exposure Time: We are provided with time stamps on the occurrence of failures. To recall, failure occurrences happen on the timeline from 23-Oct-2020 to 26-Jan-2022, 461 days being 1.263014 years. Our interest lies in the time exposure until the machine fails. Figure 4.3 shows that a machine $X(t)$ can be either in the state of operation (1) or in the state of repair (0). At the end of each operation time window $U_1, U_2, \ldots$ is a failure that starts the repair time window $D_1, D_2, \ldots$ The black dots at the end of the operating time windows denote the occurrence of a failure and the transition into repair mode. The repair restores the machine into an as-good-as-new state. There is a 2 weeks production stop during Christmas. The total time is discounted by the time in repair leading to the exposure time. The exposure time of machine A is $t_1 =$ 443.188426 days and $t_2 =$ 441.036469 days for machine B; out of the 461 days time window.

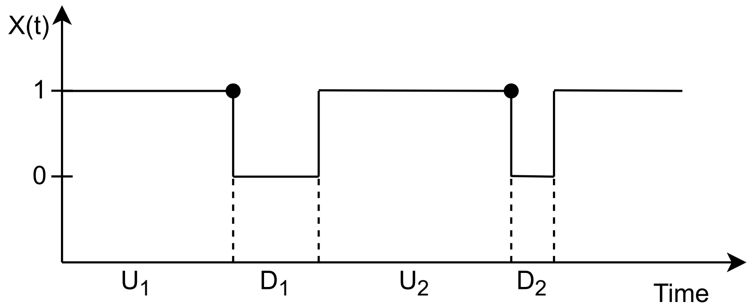


Figure 5.3: Illustration of the HPP of a machine, the comparison of operational time and time in repair.

Marginal distributions: The marginal distributions of the failure rates are assumed to be gamma. The assumption of a gamma distribution is chosen for multiple reasons. The choice for the Poisson data offers conjugate convenience. The distribution is easily scalable and is flexible for the modelling of various shapes and forms. We follow the established state of art of papers in the field, multiple studies claim this probability structure as the current state of art (Lord, 2006; Quigley et al., 2007). Practically has the first parameter, the shape parameter α, the interpretation of an expected failure count in the population pool, the second parameter, the rate parameter β, is then the exposure time, over which we normalize. The marginal distributions are as follows:

$$\pi(\lambda_i) = \frac{\left(\frac{\beta}{h_i}\right)^{\alpha} \lambda_i^{\alpha-1}}{\Gamma(\alpha)} e^{-\frac{\beta}{h_i}\lambda_i}, \lambda_i > 0, \alpha > 0, \beta > 0 \tag{5.2}$$

With λ_i modelling the non-negative failure rate on the prior. β/h_i stands for the rate parameter, where h_i represents the homogenisation factor.

Prior distribution: There are two considered forms of the prior distribution, the multivariate gamma distribution with dependencies between rate. The rates originate from:

1. A bivariate Gamma distribution developed by Minhajuddin et al. (2004) and described in (Quigley et al., 2013, p. 2211)

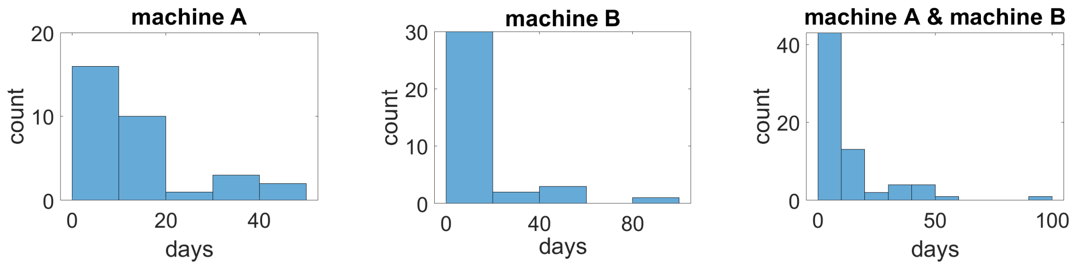
2. A copula constructed bivariate Gamma distribution, in detail in Equation 2.29 with the pdf in Equation 2.29 using the Gaussian, Frank, Gumbel and Clayton copula.

5.5.1 HPP Justification

We wish to test whether the occurrence of failures can be described with a HPP. This will support the argument that the failure rate is constant over time. To test for a HPP, we will test for the exposure time being exponentially distributed and the exposure time being stationary according to Section 2.1. For both features, we use two tests per feature. Examining machine A, machine B and the joint data set of machine A and B.

If the number of occurrences is described by the Poisson distribution, the length of time between occurences is described with an exponential distribution as Cooper (2005) notes. We can visually confirm this by looking at the time between failures distributed in a histogram on Figure 5.4 imitating the density of the exponential distribution. The average time to failure is for the combined machines 12.360369 days, machine A 13.675856 days and machine B 11.191046 days.

Exponential distribution in the data: For machine A, the KS test can not reject the null hypothesis (p=0.912193) that data originates from an exponential distribution, side by side the Lilliefors test ($p \geq 0.5$). For the joint data set of machine A and machine B, the KS test can not reject the null hypothesis (p=0.085474) that data origin from an exponential distribution. However, the Lilliefors test ($p \geq 0.013019$) is less optimistic.



(a) Exposure time between failures for machine A.

(b) Exposure time between failures for machine B.

(c) Exposure time between failures for both machines.

Figure 5.4: Three simple graphs supporting the exponential distribution argument. Having the count of the y-axes and the x-axes in days.

A TTT-plot in Figure 5.5 is used (Kvaløy and Lindqvist, 1998, p. 16) and will according to the authors have the plotted points near the diagonal if working with a HPP.

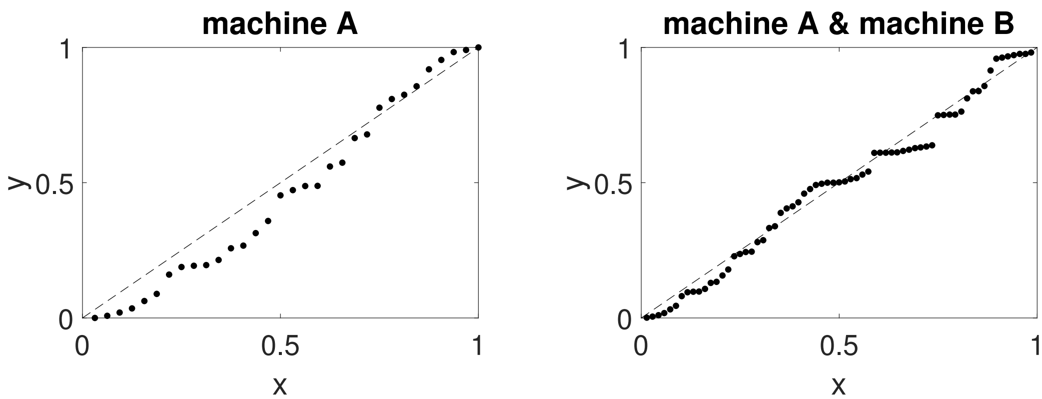


Figure 5.5: TTT plot for machine A and machine B.

S_k denotes the k-th arrival in the process $0 < S_1 \leq S_2 \leq \ldots \leq S_N \leq S$. $T(S_k)$ is the time of occurrence of the S_k event. The TTT plot consists of the points $(k/N, T(S_k)/T(S))$. The data for the joint two machines pass the Anderson-Darling test for trend (Kvaløy and Lindqvist, 1998, p. 18) with $p = 0.049$.

5.5.2 Stationarity Justification

By Section 2.1.1 are data assumed to be stationary, the mean is constant and its autocorrelation function is only time dependent. This means that the curves in Figure 5.6 should have a constant mean and variance. Taking data about machine A, both the KPSS test ($p = 0.1$) and ADF test ($p = 0.00450867$) indicate stationarity. Machine B shows stationarity with KPSS ($p = 0.1$) and ADF ($p = 0.001$). The joint data set of machine A and machine B proves also to be stationary with KPSS ($p = 0.1$) and ADF ($p = 0.001$).

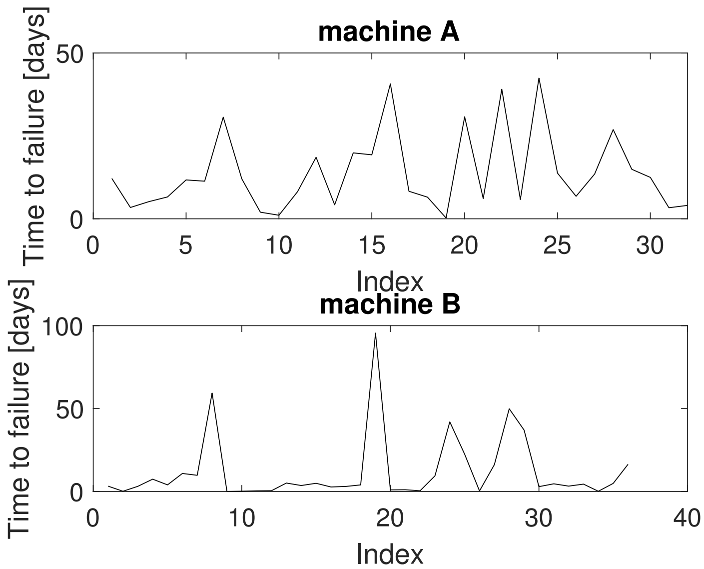


Figure 5.6: The graphical representation of the individual exposure times until failure. With on the x-axes the index of the individual subsequent time exposure and on the y-axes the magnitude of the exposure time.

Failure cause processes: The research tests only for all failure causes given the small data sets within each class. A single failure has in our case maximally 7 observations being inappropriate for any used test.

5.6 Method

For the calculation of the failure rate of failure causes, we consider multiple estimators. To evaluate a proposed estimator $\widehat{\lambda}_P$, we regard the data as random variables and calculate the KPI measures Mean Squared Error (MSE) and Mean Absolute Difference (MAD) with respect to the classical estimator. The proposed estimator $\widehat{\lambda}_P$ is a prediction F_k using data from the training set to evaluate the squared prediction error, the MSE. Done by evaluation to the actual data in the test set Y_k.

The evaluation of the MSE is given by

$$MSE = \frac{\sum_{k=1}^{K} (Y_k - F_k)^2}{K} = \frac{1}{K} \sum_{k=1}^{K} \left(\frac{N_{1k}}{t_1} - \widehat{\lambda}_P \right)^2 \tag{5.3}$$

similarly the MAD by

$$MAD = \frac{\sum_{k=1}^{K} |Y_k - F_k|}{K} = \frac{1}{K} \sum_{k=1}^{K} \left| \frac{N_{1k}}{t_1} - \widehat{\lambda}_P \right| \tag{5.4}$$

The explanation of the terms follows. Y_k denotes actual observation whereas F_k stands for the prediction of Y_k. Let us assume the classical estimator for the first machine and k-th failure cause used in (Schwarzenegger et al., 2021, eq. 2), in Formula 3.2. From a considered Poisson distribution for each failure cause, in Formula 3.1. This classical approach encompasses the computation of the ratio of the count of events to the observation duration period, here the count for the machine A for a specific failure cause. Such approach produces often and as in our case a large amount of estimates equal to zero (Quigley et al., 2007, p. 619), being inappropriate for many situations. Especially for rare observations and machines, where is not much data available. So analysts tend to move towards more conservative methods. There are methods for estimating failure rates without observing actual failure data. Welker and Lipow (1974) base their method on the χ^2 distribution. Whereas Bailey (1997) prefers a median estimate assigning the value 0.5, where no failures are observed yet and the failure rate estimate would be zero within the time window.

The rates in the classical estimator are not assumed to possess a prior distribution.

We would ideally desire an estimator close to the classical estimator that includes information about the overall data pool of all failures from the two machines. Next, we would like to highlight the observed failure counts of the target examined machine in each failure cause. Lastly, partially include failure counts of the target examined machine in each failure cause. The summarised sources of information for the inference:

1. Prior information
2. Data about machine A
3. Data about machine B

This summary can be further decomposed into a broad description. If we take apart the prior information and the partial inclusion of similar machines having included assumed knowledge about the dependency structure, in an overview as:

(i) Overall failure level

(ii) Uncertainty around the overall failure level

(iii) Data about machine A

(iv) Data about machine A that is being enriched with data about machine B.

(v) The appropriate Gaussian, Frank or (Minhajuddin et al., 2004, p. 602) dependency structure

5.6.1 The Proposed Estimators Justification

We shall now describe the proposed estimators $\widehat{\lambda}_P$ that will be evaluated via Equation 5.3 in Table 5.3 with respect to described environment to justify the suitability. Table 5.3 comprises of results sorted column-wise by the used parameters that are shown in Table 5.5 with the exception of the dependency estimator that is fixed at $\widehat{\rho}_{MME}$.

Empirical Bayes $\rho = 1$: The first estimator is the empirical with full dependency in Formula 3.19. It combines data from both machines together with prior information. The observed pair of failure counts and the prior knowledge. We assume in this case that data from machine A and machine B are indistinguishable. λ_i represents the i-th failure cause for which machine A has n_1 failures and machine B n_2 failures. t_1 the exposure of machine A, t_2 the exposure of machine B and h_i the homogenisation factors. We may substitute α with $\widehat{\alpha}$ as it is an estimate.

Empirical Bayes $\rho = 0$: The second estimator is very similar to the first estimator with the difference that we assume no similarity between the two machines. The estimation in Formula 3.18. We make use of the overall prior information. Then the data observations for each failure cause. The performance of the MSE improves as compared to the empirical estimator with full dependency.

Empirical Bayes $\rho = \widehat{\rho}$: The third estimator (Formula 3.16) is a combination of the two previous ones. The classification may be an empirical Bayes estimator with partial dependency. We assume that there is some degree of similarity between the machines, specified by $\widehat{\rho}$. In Table 5.3 is applied the value of the correlation parameter at $\widehat{\rho} = 0.321950$, the moment estimator. It incorporates the knowledge about the prior distribution. Then for each failure cause is produced a dedicated rate estimate based on the observed count of occurrences in each cause. Only a part of the informational value of machine B is included. The limiting cases $\rho \to 0$ and $\rho \to 1$ converge towards the empirical Bayes cases with full/no dependency (Quigley et al., 2013, sec. 2.1).

Copula based estimator Gaussian, Frank, Gumbel and Clayton In the fourth estimator, we combine prior information, value and its form between the marginal distributions, a Gaussian copula, Frank copula, Gumbel and Clayton. The Formula 2.30 shows similarities to the empirical Bayes estimator with partial dependency. In addition, judgments are made about the dependency structure. That is, deciding whether the marginals will be more correlated in one or the other tail. Leading to selecting an appropriate copula. We consider the same value as by the empirical Bayes estimator from the previous paragraph. Whereas for the Gaussian copula is the dependency estimate the exact copula parameter, for the Frank copula we approximate the rank correlation with a linear correlation estimate. The copula estimates provide us with even better results than the empirical Bayes estimator with partial dependency as seen in Table 5.3. The limiting cases $\rho \to 0$ and $\rho \to 1$ for the copulas converge towards the independence, resp. co-monotonicity copula (Ruppert, 2011, pp. 178-179).

5.6.2 Parameter Estimation of Marginal Distribution Prior Rates

We produce two types of prior estimators. The moment estimator and the maximum likelihood. Moment estimator (Quigley et al., 2007, p. 2218) provides us with point estimates of $\widehat{\alpha}, \widehat{\beta}, \widehat{\rho}$. The maximum likelihood estimator gives as well point estimates of these variables. In addition, we construct with the interval method the 95th percentile confidence region described in (Quigley et al., 2007). The interval estimates of the prior parameters $\widehat{\alpha}, \widehat{\beta}$ (Dias et al., 2017, p. 156) are obtained separately for all parameters. Pairing the individual lower estimates of $\widehat{\alpha}, \widehat{\beta}$ produces a conservative estimate of at least the 95-th percentile. The dependency and prior parameters in their point and interval estimates are listed in Table 5.5. How the MSE changes using the empirical Bayes estimator with partial dependency for the machine A changes is shown in Figure 5.7. On the x-axes being the assumed dependency ρ_A. The points for $0 < \rho_A < 1$ are being calculated with the empirical Bayes estimator with partial dependency. The shape of the curve imitates a convex function. The estimates for $\rho_A = 0$ and $\rho_A = 1$ are obtained with the empirical Bayes estimate under no dependency, respectively full dependency.

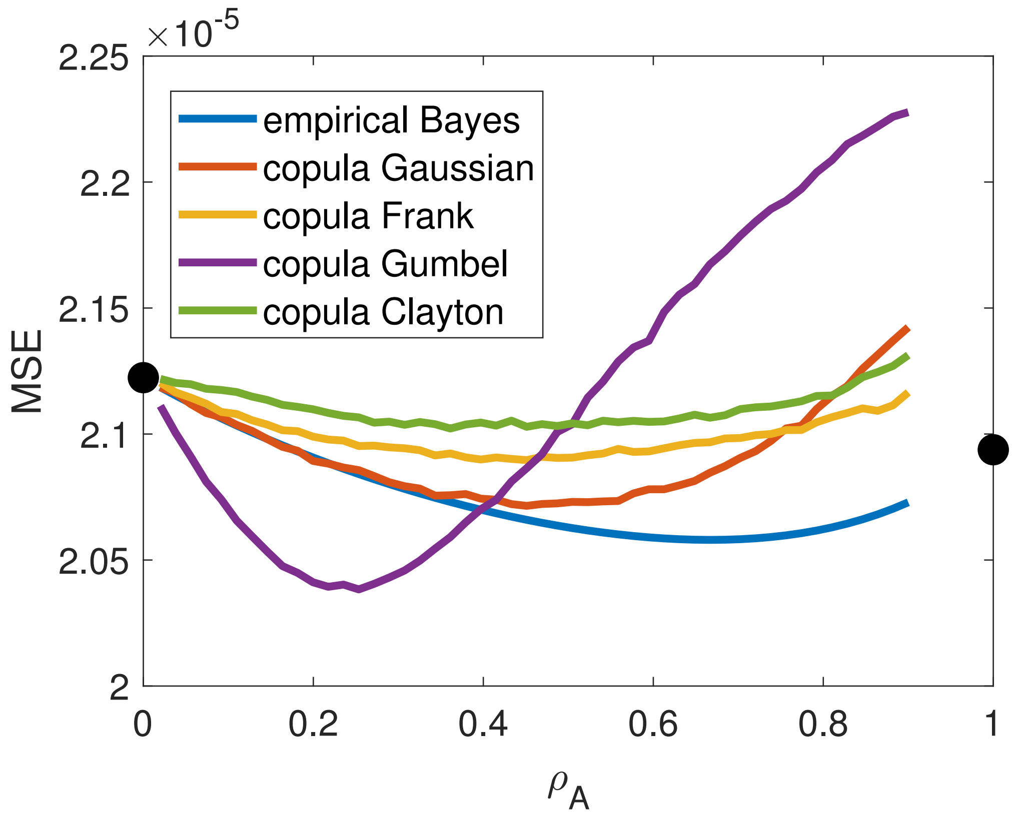


Figure 5.7: MSE of machine A given machine B using empirical Bayes estimation with full dependency, no dependency and partial dependency estimation. On the y-axes the MSE obtained by Formula 5.3. On the x-axes the assumed dependency ρ_A.

(to multiply by 1E-5)	MME	MLE	MLE lower	MLE upper
Empirical Bayes $\rho = 1$	2.093767	2.098799	2.108671	2.127661
Empirical Bayes $\rho = 0$	2.122369	2.133366	2.156022	2.179553
Empirical Bayes $\rho = \widehat{\rho}$	2.076769	2.079848	2.087128	2.128924
Copula estimate Gaussian	2.079697	2.085058	2.098107	2.112261
Copula estimate Frank	2.094549	2.045427	2.126419	2.146510
Copula estimate Gumbel	2.042552	2.045427	2.057802	2.093060
Copula estimate Clayton	2.105878	2.120270	2.146068	2.164637

Table 5.3: MSE results for machine A under parameters listed in Table 5.5 with a fixed $\widehat{\rho}$ under the MME.

(to multiply by 1E-5)	MME	MLE	MLE lower	MLE upper
Empirical Bayes $\rho = 1$	2.143223	2.149173	2.159995	2.178089
Empirical Bayes $\rho = 0$	2.106405	2.115268	2.137664	2.188715
Empirical Bayes $\rho = \widehat{\rho}$	2.073333	2.075810	2.087141	2.147872
Copula estimate Gaussian	2.082635	2.085880	2.094941	2.135661
Copula estimate Frank	2.104919	2.110459	2.124484	2.171466
Copula estimate Gumbel	2.049516	2.053288	2.077132	2.126029
Copula estimate Clayton	2.106962	2.116141	2.135460	2.184820

Table 5.4: MAD results for machine A under parameters listed in Table 5.5 with a fixed $\widehat{\rho}$ under the MME.

	MME	MLE	MLE lower	MLE upper
$\widehat{\alpha}$	0.519789	0.360426	0.198188	0.563273
$\widehat{\beta}$	351.466657	318.697482	307.487216	755.972456
$\widehat{\rho}$	0.321950	0.515458	-	0.902012

Table 5.5: Prior estimates of machine A & machine B consisting of point and interval estimates. The values are for the methods of moments (MME), the maximum likelihood method (MLE), the lower (MLE lower) and upper (MLE upper) bound for the maximum likelihood method.

5.7 Analysis

Examining the results we observe:

- For the metric of the MSE are included estimators working with partial dependency outperforming those who do not. Having the first two rows in Table 5.3 as reference values, the remaining estimators outdo the case of omitting dependency and very often case of considering full dependency. This is the case for parameters obtained by the MME and MLE.

- These results hold also for the MAD metric. Excluding the Clayton based estimator, but including next to the MME and MLE together with the MLE lower and MLE upper parameters. However, the results remain valid for the parameter values of MME, MLE and as well the interval estimates MLE lower and MLE upper.

- Both the MSE and MAD results indicate the use of the Gumbel based estimator followed by the empirical Bayes estimator with partial dependency together with the Gaussian based copula estimator.

- The MSE shows a 3.91% increase if using instead of a Gumbel the empirical Bayes estimator with no dependency.

- The MAD shows a 4.57% increase if using instead of a Gumbel the empirical Bayes estimator with full dependency.

- By not applying the Gaussian copula, but the empirical Bayes estimator with no dependency, the MSE by 2.05% and the MAD by 2.90% with full dependency. Similarly the partial dependency empirical Bayes estimators increases are 3.30% for the MSE and 3.37% for the MAD.

- If we would consider a most suitable dependency parameter for the MSE, the Gumbel based estimator under $\rho_A = 0.253469$ would have a 4.12% increase, Gaussian 2.43% under $\rho_A = 0.45102$ and the empirical Bayes estimator 3.13% under $\rho_A = 0.666531$. If comparing the ideal cases with the case of no dependency. That is if using the case of no dependency instead of applying the ideal cases with the most suitable dependency parameter.

From results in Chapter 4 are simulation results showings

- A 6.42% maximal MSPE increase if the generating process is Gumbel with $\rho_S = 0.5$ and the underlying structure is Clayton for the 10th percentile.

Chapter 3 shows

- A 16.93% increase if mistakenly assuming correlation $\rho_A = 1$ instead of the correct $\rho_A = 0.5$. However, if the underlying structure would have a Gumbel copula between failure rates, the increase would be just 8.85% under the same mistaken dependency assumption.

From Chapter 4 it is risky to take the Gumbel copula, in case there would be an underlying Gaussian structure. It would be safer to choose the partial dependency empirical Bayes estimator and Gaussian copula based estimator. The increases in the MSPE are 2.05 % when using instead of the Gaussian copula based estimator the empirical Bayes with no dependency and 2.20% with the partial dependency empirical Bayes estimator compared with the empirical Bayes estimator with no dependency. These values are in line with Yahav and Shmueli (2012) observing mostly a 3% difference for correlated rates in multivariate Poisson counts.

Dependencies are important. From Chapter 4 examining copula mis-specifications we can conclude what will be the true $\%bias$es and $MSPE$s when the underlying copula structure is known. The closest examined case is the case with $\alpha = 0.5, \beta = 1, t = 1$. However, if we assume an underlying true dependency $\rho_S = 0.75$. For the $\%bias$ in the 90th percentile, if we choose the Gaussian copula, the lowest possible absolute value will be 1.15% for the Frank copula. The largest absolute value of 2.76% occurs under choosing the Clayton copula, when the actual underlying true copula is Gumbel. For the time horizon $t = 5$ the best and the worst choice switches with 5.72% by choosing Gaussian and -1.76% choosing the Clayton. In the time horizon $t = 10$ the choices remain the same with 7.12% for Gaussian and 2.87% for Clayton. In the 90th percentile, are for all times ($t = 1$, $t = 5$, $t = 10$) the safest choice the Frank copula and the riskiest choice the Gumbel copula. The values of the $\%bias$ are for the Frank copula 1.39%, 7.71% and -5.44%. Compared to the Gumbel copula with 3.97%, 25.37% and 29.35%.

The biggest differences for the 10th percentile of MSPE. As by the previous case, the safest choice remains the Frank even at $t = 10$ copula and the riskiest choice the Gumbel copula for all times. The Frank copula is ranging in values 2.50%, 5.20%, 2.43% and the Gumbel copula 2.28%, 24.67% and 57.55%. More precisely, for $t = 1$ and $\rho_S = 0.25$ are these values for the Frank copula 3.01%, 2.35%, 2.14% and for the Gumbel copula 5.13%, 5.69%, 11.67%.

So we can support the argument that a safe choice is not the Gumbel copula and that differences are sometimes just in single digits. Rather then in the 90th and 5th percentile, the difference is in the 10th percentile. Further, it is important to consider partial dependency between failure rates.

5.8 Cross-Validation

To validate results, an evaluation of the prediction squared error Err (squared-error loss) by (Hastie et al., 2009, p. 223) will be performed in form of a cross-validation set up in Section B.2.2. Given that we have 32 observed failures in the machine A 36 failures in the machine B, we perform a K-fold cross validation to evaluate the prediction error. We treat the data as panel data and divide the 461 available days into five sets of 92 days. Take the observed failures in these days and the realized exposure time. Then we keep always one set aside and use the other four to predict this set. Hastie et al. (2009, p. 242) advice typical choices of $K = 5, 10$. We split the data into five sets $A_i, i \in \{1, \ldots, 5\}$ (validation data). And estimate the squared error loss by using the left out set A_i for the classical estimator C and all other sets $\neg A_i$ (training data) for the proposed estimator P as shows Equation 5.5. The training data have more observed failures and a larger exposure time than the validation data. Both prior parameters will be larger than those generated ones obtained from the validation set, however the mean as their fraction of them (α/β) will be the same, adding to the justification of the approach that estimates obtained from the validation data can be used for prediction of the training data.

Hastie et al. (2009, p. 242) argue "cross-validation only estimates effectively the average error". We can confirm that the overall mean has the desired properties.

$$Err = E\left(\underset{\sim}{\widehat{\lambda}}_{C,A_i} - \underset{\sim}{\widehat{\lambda}}_{P,\neg A_i}\right)^2 \tag{5.5}$$

When examining the results in Table 5.6, we compare the partial estimators mainly towards the empirical Bayes estimator with no dependency. 3/5 sets (2nd, 4th and 5th) have all partial dependency estimators performing satisfactory in terms of concluding that considering dependency is better than omitting it. These are the second, fourth and fifth sets. For the other two sets, the first and third have a good performing Clayton estimator.

The left out set for validation A_i is constructed following Figure 5.3, let us assume, we take for validation the time window beginning in the middle of the interval U_1 and ending in the middle of the interval U_2. Then the exposure time is all the exposure between the beginning and end of the validating time window. The number of failures are the observed failures in this time window.

(to multiply by 1E-5)	1st	2nd	3rd	4th	5th	mean
Empirical Bayes $\rho = 1$	1.401967	**1.820986**	1.637668	**1.433732**	**1.729563**	**1.604783**
Empirical Bayes $\rho = 0$	1.117471	**1.807235**	1.660637	**1.470667**	**1.688114**	**1.548825**
Empirical Bayes $\rho = \widehat{\rho}$	1.151788	**1.759115**	1.659922	**1.439148**	**1.664188**	**1.534832**
Copula estimate Gaussian	1.147046	**1.741885**	1.66394	**1.462265**	**1.672010**	**1.537430**
Copula estimate Frank	1.121289	**1.752446**	1.648675	**1.448365**	**1.688451**	**1.531331**
Copula estimate Gumbel	1.290728	**1.714261**	1.724079	**1.552428**	**1.663897**	**1.589079**
Copula estimate Clayton	1.115519	**1.778378**	1.628885	**1.452949**	**1.685815**	**1.532309**

Table 5.6: Cross-validation results.

5.9 Decision Problem

We apply the results on spares parts resourcing. To estimate what is the expected number of parts that will fail. Let us consider that when an failure occurs resulting in a repair action, the machine stops producing. The resulting costs for the company in the repair action itself. The costs are as well and mainly in the opportunity costs, which is the loss of profits represented by a production stop. We assign the following costs for production stop given by a failure as £10,000. We designate £1,000 per failing part for servicing and replacing in spare parts allocation.

Different methods give different failure rates and we compare the outcomes. The estimate of number of failures for a year ahead $t = 1$ is obtained as the rounded outcome of the failure rate multiplied by the time horizon, we are predicting ahead. That is one year.

$$\lfloor \widehat{\lambda}_P \cdot t \rceil = \widehat{N}_P \tag{5.6}$$

Table 5.7 presents the number of estimated failures for a given estimator. Each method assigns a specific anticipated count per failure cause (from 52 failure causes). Given that we we concluded some methods to be better than other, we will be selecting the preferred ones.

	Estimated Failures
Empirical Bayes $\rho = 1$	28
Empirical Bayes $\rho = 0$	16
Empirical Bayes $\rho = \widehat{\rho}$	19
Copula estimate Gaussian	22
Copula estimate Frank	23
Copula estimate Gumbel	24
Copula estimate Clayton	16

Table 5.7: Estimated number of failures in a decision problem under a given estimator.

The differences in the number of productions stops as compared to the empirical Bayes estimator with full dependency and no dependency are detailed in Table 5.8. The preferred estimators will be as in the preceding case selected.

	Empirical Bayes $\rho = 0$	Empirical Bayes $\rho = 1$
Empirical Bayes $\rho = 1$	12	0
Empirical Bayes $\rho = 0$	0	12
Empirical Bayes $\rho = \widehat{\rho}$	3	9
Copula estimate Gaussian	6	6
Copula estimate Frank	7	5
Copula estimate Gumbel	8	6
Copula estimate Clayton	0	12

Table 5.8: The difference of estimated failures resulting in a production stop for estimators considering dependency in comparison to estimators considering no dependency and full dependency.

So if we use the Gaussian method that provides accordingly to the MSE and MAD better values, we save £6,000 compared to Empirical Bayes $\rho = 0$ estimation method and £6,000 in comparison to Empirical Bayes $\rho = 1$ in wrong spare parts allocation. If we use the Empirical Bayes $\rho = \widehat{\rho}$ method which is well advisable by the MSE and MAD, the saving is £3,000 compared to Empirical Bayes $\rho = 0$ estimation method and £9,000 in comparison to Empirical Bayes $\rho = 1$ for mismatching spare parts allocation. The difference in production stop given Copula estimate Gaussian by £60,000 (compared to Empirical Bayes $\rho = 0$) and £90,000 ($\rho = 1$) and Empirical Bayes $\rho = \widehat{\rho}$ mismatches by £30,000 ($\rho = 0$) and £90,000 ($\rho = 1$).

5.10 Summary and Conclusion

We applied methods for the failure estimation of the machine A. In terms of the MSE, it showed that it is better to assume dependency rather than omitting it.

The empirical Bayes methodology with partial dependency, Gaussian and Gumbel copula approach perform well. Since choosing the Gumbel copula is slightly risky, we prefer the remaining two. The results hold across the prior confidence regions.

In Table 5.3 we draw conclusions about the MSE and Table 5.4 about the MAD when comparing the predicted failure rates with the actual observed rates. All partial dependency estimators are better than no or full dependency. Especially, using the Gumbel copula is better than using the empirical Bayes estimation with full dependency. There are improvements by applying the empirical Bayes method with partial dependency. Finally, the Gaussian copula based method leads to a very good performance. A 5-fold-cross-validation is applied to verify the proposed method. We use always 4 sets to predict the remaining 5th set. The results hold for 3 out of 5 sets and for the overall mean.

We provided point and interval estimates for the prior parameters using the moment estimator and the maximum likelihood estimator. The point estimates of $\widehat{\alpha}, \widehat{\beta}, \widehat{\rho}$ in Table 5.5 are as anticipated within the 95-th credible interval of these estimates.

We described the data to give the reader a picture of the generating process and statistical features of the data set. The argument of a homogeneous Poisson process and stationarity was supported with visual charts and tests. The examined machine and the joint data of the machines pass the exponential test. Stationarity is indicated by the separate and joint data about the machines.

The objectives in Section 5.3 are met in proposing a way to estimate failure rates under partial dependency. We show the error reduction in the MSE for multiple estimators using partial dependency. The most complex formulas with partial dependency seem as a sensible middle way between full and no dependency. We incorporate in them the three sources of information for calculating a given failure cause rate; the overall population data, the specific target data and similar supporting data. The results are cross-validated to increase their validity.

The application shows that instead of omitting dependency if we include dependency, we are able to assign better a total of £33,000 with Empirical Bayes $\rho = \widehat{\rho}$ and £66,000 Copula estimate Gaussian.

Chapter 6

Conclusion and Future Work

This thesis mainly aimed to investigate the performance of an analytical approach based on a multivariate Poisson-Gamma probability model with parameter dependency.

Our first contribution demonstrated a significant potential reduction in mean squared error (MSE) by about half by employing structured expert judgement to accurately assess rate dependencies. We presented a mathematical expression to evaluate the impact of correct dependency assessment on the MSE and further guidance of challenges that might be experienced. Based on examined scenarios, we concluded that it depends whether it is actually worthwhile to consider and elicit dependency within the examined probability model.

In the second contribution we experienced that mis-specification of the copula structure effected the error differently in its percentiles. The largest impacts were experienced in the lower percentiles for the percentage bias as well the mean squared percentage error. Our findings indicate that the selection of an assumed underlying copula structure, in conjunction with accurate knowledge of the true copula, plays a critical role. We identified four combinations of these factors that hold particular importance in determining the overall outcomes.

The third contribution combined gained knowledge from both preceding chapters. By considering similarity between machine failure causes, we showed the advantage of including dependency in failure rate estimation. We scrutinized the data with appropriate tests to indicate the suitability and applicability of employed methods. Furthermore, we emphasized the significance of considering potential risks that analysts should be mindful of during the estimation process. The analysis showed that including partial dependency decreases estimation error for this data set. The results were cross-validated with a 5-fold cross-validation to ensure the robustness of the findings.

6.1 Research Objectives

R.O. 1: How was the estimation accuracy investigated using empirical Bayes inference when there is correlation between the mean rates in a multivariate Poisson-Gamma probability model?

The work in our first contribution shows that in the first place, the MSE displays proportionality to the shape parameter. Using the logarithm of the pool size and the square of the correlation parameter, we are able to demonstrate a good fit. In alignment with our anticipation, the bigger the pool size is, the smaller the MSE. As we obtain more data and the estimates become more precise. The larger the dependency parameter, the smaller the MSE. We are able to anticipate also this relationship. Since the more dependence we observe, the more data and information for estimation we may include.

We find that

$$\lim_{\rho \to 1, m \to \infty} MSE = \frac{\alpha}{3} \tag{6.1}$$

$$\lim_{\rho \to 0, m \to \infty} MSE = \frac{\alpha}{2} \tag{6.2}$$

and we can anticipate the value around

$$\lim_{\rho \to 0.5, m \to \infty} MSE = 0.456\,\alpha \tag{6.3}$$

R.O. 2: What were the changes in estimation accuracy when the correlation between the mean rates in a multivariate Poisson-Gamma probability model is mis-specified?
Our research showed that we are able to reduce the MSE by up to 53.8% in the specific case of $\rho_T = 0$, but elicited as $\rho_A = 1$ ($\alpha = 50, m = 60$). This means that if the analyst

using the model invests in structured expert judgement to obtain perfect assessments of the dependency between rates, then it is possible to achieve an MSE that is more than half of the value of a maximally mis-specified dependency parameter. Hence we have provided the analyst who might use this reliability model in an application to understand the consequences of dependency mis-specification and the factors that drive the estimation error reduction when accounting for the correlation between event rates. This information should help inform the analyst to decide whether accounting for correlation in analysis is worth the effort.

R.O. 3: To examine the effect on estimation accuracy and precision when the dependency in a multivariate Poisson-Gamma probability model is represented using a copula, including when the copula family is mis-specified.

When assuming a data generating process with a copulas structure, we examine how with mis-specification of the copula function, the bias and the MSE will be affected. Given assumed dependency, we examine tendencies of mis-specification in the 10th, 50th and 90th percentile. Out of 16 possible investigates combinations in a simulation, we are identifying 4 combinations that do matter the most, *Gumbel_A-Frank_T*, *Gaussian_A-Frank_T*, *Gumbel_A-Clayton_T* and *Gaussian_A-Clayton_T*. For our case study in the last contribution, we show that by including partial dependency, especially using the Gaussian and Frank copula, we achieve less uncertain estimators.

R.O. 4: How was the application of methods to account for dependency in the parameters of a Poisson-Gamma model illustrated using both empirical Bayes inference and copulas through an industry case? Further:

1. *How was the relative performance of the different methods for this case compared?*
2. *How were the results of modelling explored that might be translated to a decision-making context?*

For the final application, we conclude based on test the results and estimates are consistent and unchanging throughout time (time invariant), since we are working with a HPP. The data are devided into a training set and validation set. A final cross-validation confirms the results and is performed on the total dataset. It shows that it is almost always better to assume partial dependency for this data. The best results are obtained taking the Gumbel copula. However, we take rather a more conservative approach due to possible mis-specifications and prefer the Gaussian copula, leading to a possible saving of £66,000.

6.2 Limitations and Future Research

We acknowledge that we operate within several limitations. Relaxing those limitations might give opportunities for future research. Within the thesis, we used just Poisson data and Gamma priors. Exploring Poisson-Lognormal cases or other prior distribution possibilities. This would offer fewer limitations in situations, where these marginals would be more appropriate. In addition, the exact formulas for the MSE for the limiting cases under full and no dependency between failure rates are available. However, analytical exact formulas for partial dependency would be of great advantage. Additionally, the research is limited to the bivariate case. A pair of rates that has a given correlation. Future research may include the derivation of the estimators for dependency and prior parameters using the copula predictive density. The quantification of the individual contribution towards the posterior mean of the prior and both Poisson observations in a copula constructed posterior to improve traceability. For non-Gaussian copulas, where the linear correlation coefficient can not be used, suitable methods could be incorporated for finding the dependency parameter. For the non-bivariate case incorporate vine copulas and start exploring more complex structures. Future research can extend to other copulas. Apply the method in relation to creating a regressive model to take into account other variables like Park et al. (2021) or Caliendo et al. (2021). Use multivariate Poisson-Lognormal mixture model and extend to other copulas (Park et al., 2021). To include other marginals to model different severities in crashes. It can be the use of zero-inflated models or the use of negative binomials distribution of levels of severity.

Next, all methods are valid only for and operating under positive dependencies. Negative dependencies are not considered. They could be started to be explored by the Gaussian and Frank copulas, since their formulas permit negative dependencies.

Fantazzini (2009) shows they can have a larger impact that positive values. Further, we are interested in finding the for non-Gaussian copulas the rank correlation parameter or an equivalent quantity. Dimitriou (2022, p. 57) developed prior parameter estimation using data assuming a Clayton copula. This can be further developed for more copulas. We are interested in the derivation of the estimators using the predictive density, as done in Quigley et al. (2007, p. 622), for the copula dependency cases. In the study are used data from accident analysis and industrial production. There are other fields, like software development, where the research could be applied.

Appendix A

Sensitivity on Copulas in Bayesian Hierarchical Model

	t = 5	TRUE			
		Gaussian	Frank	Gumbel	Clayton
ASSUMED	Gaussian	-1.039%	7.104%	-0.488%	16.883%
		-0.864%	7.087%	-1.790%	13.228%
		0.641%	7.921%	-3.061%	12.210%
	Frank	-3.997%	-0.261%	-3.729%	7.714%
		-6.038%	-0.359%	-5.998%	4.119%
		-6.131%	0.018%	-7.431%	3.262%
	Gumbel	2.070%	11.288%	-0.420%	25.373%
		1.828%	10.227%	-1.116%	17.940%
		2.922%	11.020%	0.401%	17.063%
	Clayton	-4.661%	-0.929%	-0.712%	0.900%
		-6.545%	-0.847%	-6.758%	0.117%
		-8.179%	-2.457%	-9.421%	0.309%

Figure A.1: Results of $\%bias$ alike to Figure 4.1 with the time of $t = 5$.

	t = 1	TRUE			
		Gaussian	Frank	Gumbel	Clayton
ASSUMED	Gaussian	-1.179%	1.587%	0.846%	4.951%
		0.472%	-0.051%	0.196%	-0.597%
		0.814%	0.541%	1.191%	-1.166%
	Frank	-1.081%	0.945%	-0.192%	1.392%
		2.361%	1.282%	0.241%	1.228%
		0.904%	-0.094%	0.625%	-0.653%
	Gumbel	-2.229%	-1.691%	-0.963%	3.965%
		-0.539%	0.418%	-0.779%	1.417%
		0.001%	-0.083%	0.502%	-1.256%
	Clayton	2.498%	3.708%	4.000%	-0.458%
		2.888%	7.278%	4.324%	0.227%
		3.564%	2.363%	5.131%	0.967%

Figure A.2: Results of $\%bias$ alike to Figure 4.1 with the time of $t = 1$.

	t = 5	TRUE			
		Gaussian	Frank	Gumbel	Clayton
ASSUMED	Gaussian	0.081%	5.704%	1.487%	23.371%
		0.054%	2.946%	1.043%	8.513%
		0.032%	1.853%	0.707%	5.178%
	Frank	1.433%	0.019%	3.488%	3.332%
		1.700%	0.022%	2.832%	1.808%
		1.191%	0.053%	2.548%	0.675%
	Gumbel	1.683%	16.058%	0.034%	54.446%
		1.094%	7.392%	0.062%	18.759%
		0.666%	4.947%	0.025%	10.818%
	Clayton	5.108%	3.438%	11.661%	0.055%
		3.482%	2.721%	7.240%	0.025%
		2.077%	0.816%	4.136%	0.065%

Figure A.3: Results of $MSPE$ alike to Figure 4.1 with the time of $t = 5$.

	t = 1	TRUE			
		Gaussian	Frank	Gumbel	Clayton
ASSUMED	Gaussian	0.024%	0.263%	1.152%	3.822%
		0.005%	0.863%	0.938%	1.677%
		0.040%	0.767%	0.335%	0.845%
	Frank	0.170%	0.016%	1.195%	2.243%
		0.985%	0.062%	0.370%	3.446%
		0.867%	0.003%	0.288%	2.058%
	Gumbel	0.691%	0.939%	0.026%	6.423%
		0.595%	0.458%	0.028%	4.185%
		0.376%	0.298%	0.014%	1.888%
	Clayton	3.335%	4.394%	11.435%	0.004%
		2.327%	5.551%	7.212%	0.001%
		0.973%	3.678%	2.794%	0.080%

Figure A.4: Results of $MSPE$ alike to Figure 4.1 with the time of $t = 1$.

Sensitivity on dependency

Additional to Figure 4.4 and 4.5, is laid out results for the option of the dependency ρ_S equalling 0.5 and 0.25.

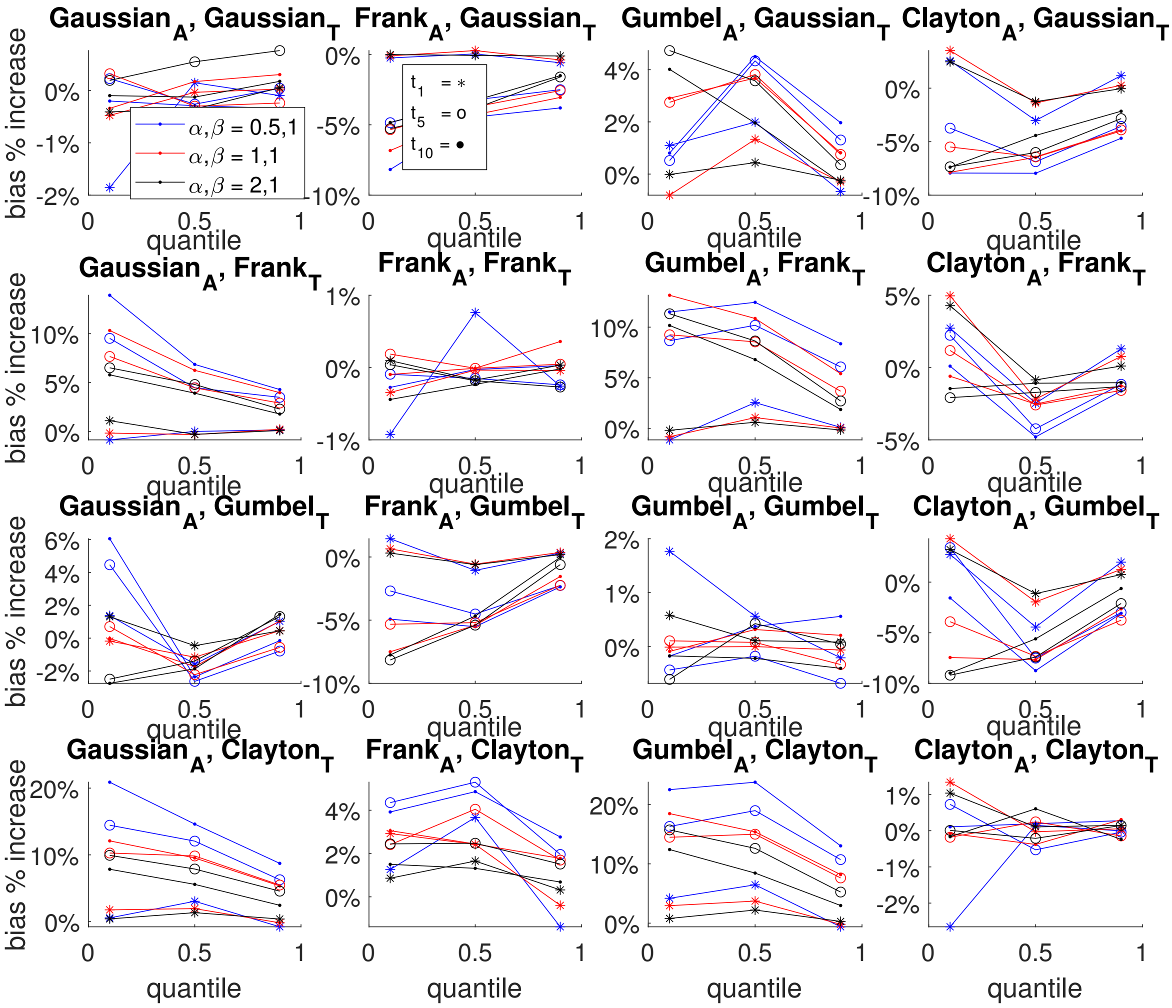


Figure A.5: The simulation results for the $\%bias$ for $\rho_S = 0.5$. $\%bias : (q_A - q_T)/q_T \cdot 100\%$. Extending Figure 4.4.

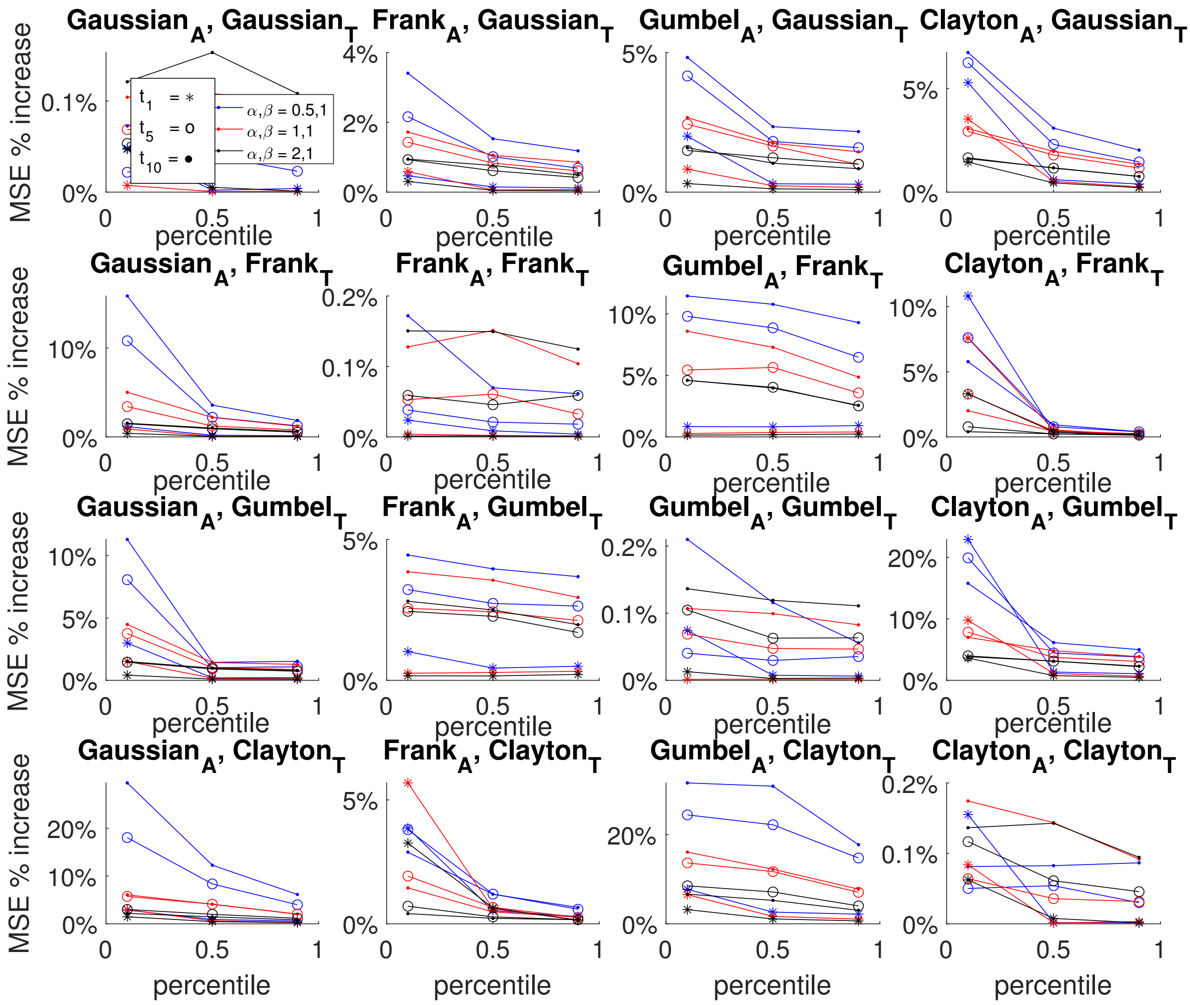


Figure A.6: The simulation results for the $MSPE$ for $\rho_S = 0.5$. $MSPE$: $((q_A - q_T)/q_T)^2 \cdot 100\%$. Extending Figure 4.5.

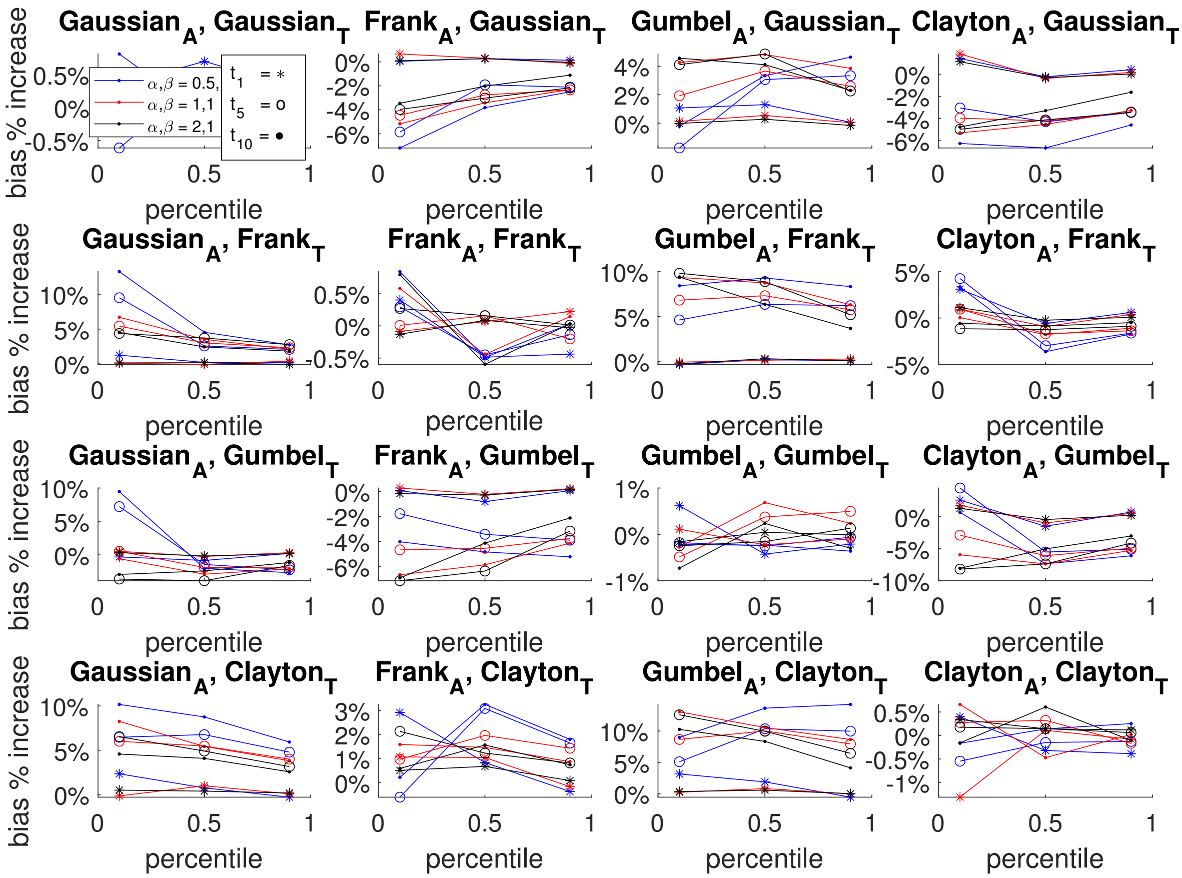


Figure A.7: The simulation results for the $\%bias$ for $\rho_S = 0.25$. $\%bias : (q_A - q_T)/q_T \cdot 100\%$. Extending Figure 4.4.

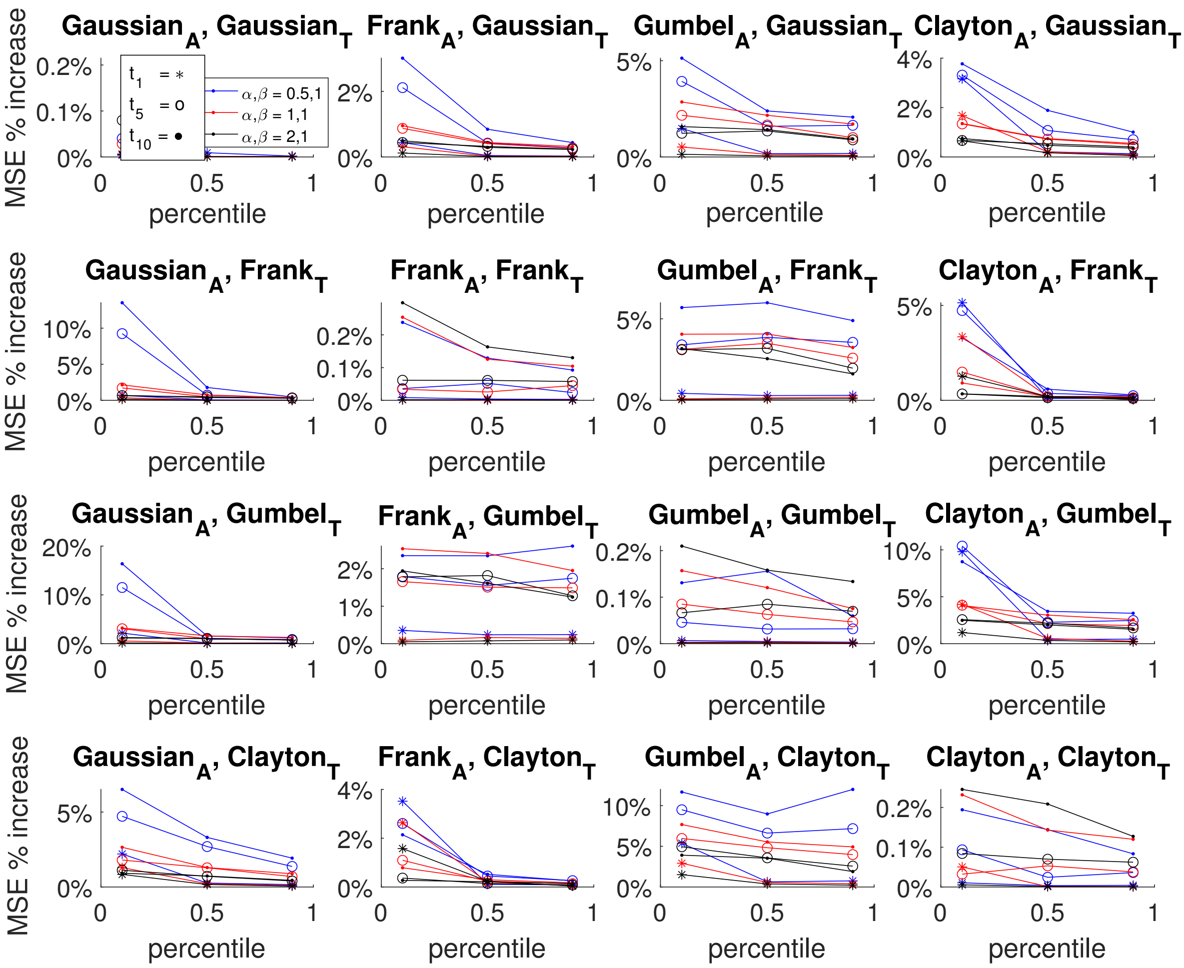


Figure A.8: The simulation results for the $MSPE$ for $\rho_S = 0.25$. $MSPE$: $((q_A - q_T)/q_T)^2 \cdot 100\%$. Extending Figure 4.5.

Appendix B

Technical Details for Used Methods

B.1 Monte Carlo Integration

Since the posterior mean in Formula 2.30 does not have an analytical expression, an estimation method need hence to be taken. We apply the Monte Carlo methods to obtain estimates of the true value. By Givens and Hoeting (2012, p. 151) let $\mu = E\{h(\boldsymbol{X})\}$ be the expectation of a function and f the density of $\boldsymbol{X}$. Then using random sampling $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ we can approximate

$$\widehat{\mu}_{\mathrm{MC}} = \frac{1}{n} \sum_{i=1}^{n} h(\boldsymbol{X}_i) \to \int h(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} = \mu \qquad \text{(B.1)}$$

as $n \to \infty$ by the strong law of large numbers.

B.2 Assessing Models

A model is always an approximation of reality. To address the suitability of a model, the most common measures for accessing the accuracy of model data are the r^2, Bayes factor and likelihood ratio test. For validation is used the leave-one-out validation, k-fold cross validation or likelihood comparison (Hastie et al., 2009, ch. 7).

B.2.1 Partial Variable Plot

The partial variable plot (in Chapter 3) by (Velleman and Welsch, 1981) is removing by the Frisch–Waugh–Lovell theorem the model all variables with exception of the constant term. Broadly speaking, the plot is showing the constant term after having controlled for all other variables.

We create a new variable of all variables except the intercept: we have the plan matrix X with the coefficient vector β, having 0 in the first column (the intercept). Using $s = \text{norm}(\beta)$, $u = \beta/s$, $X\beta = (Xu)s$. Xu is our new variable with coefficient scalar s.

We select all variables, except the intercept x_1 as follows:

$$y_i = g_y(x_{1i}) + r_{yi}$$

$$x_i = g_x(x_{1i}) + r_{xi}$$

g_y is the fit of y against the intercept. g_x is the fit of all variables, against the intercept. We denote all variables except the intercept (xu) as previously.

$$\tilde{y}_i = \bar{y} + r_{yi}$$

$$\widetilde{(xu)}_i = \overline{(xu)} + r_{xi}$$

Matlab command `plotAdded` plots a scatter plot of $(\widetilde{(xu)}_i, \tilde{y}_i)$ and also as a fitted line $s\widetilde{(xu)}$. s is the same coefficient as before. After all variables are removed, the final adjusted variables are x_0, y_0.

B.2.2 Cross-Validation

To validate results (in Chapter 5), we evaluate the prediction squared error Err with a K-fold cross-validation. The ideal case of validation as (Efron and Hastie, 2016, p. 213) states would be to have an independent validation set (test set).

Let us define:

- *training set*

$$\boldsymbol{d}_{\text{val}} = \{(x_{0j}, y_{0j}), j = 1, 2, \ldots, N_{\text{val}}\} \tag{B.2}$$

- *prediction rule*

$$\widehat{y} = r_{d(x)} \quad \text{for} \quad x \in X \tag{B.3}$$

- discrepancy

$$D(y, \widehat{y}) = (y - \widehat{y})^2 \tag{B.4}$$

is quantifying the prediction error

- validation set $\boldsymbol{d}_{\text{val}}$

$$\boldsymbol{d}_{\text{val}} = \{(x_{0j}, y_{0j}), j = 1, 2, \ldots, N_{\text{val}}\} \tag{B.5}$$

giving an unbiased Error estimate Err

$$\widehat{\text{Err}}_{\text{val}} = \frac{1}{N_{\text{val}}} \sum_{j=1}^{N_{\text{val}}} D(y_{0j}, \widehat{y}_{0j}), \quad \widehat{y}_{0j} = r_d(x_{0j}) \tag{B.6}$$

The cross-validation is mimicking the Error-estimate with a cross-validation estimate.

$$\widehat{\text{Err}}_{\text{cv}} = \frac{1}{N} \sum_{i=1}^{N} D(y_i, \widehat{y}_{(i)}), \quad \widehat{y}_{(i)} = r_{\boldsymbol{d}(i)}(x_i) \tag{B.7}$$

$\boldsymbol{d}(i)$ being the reduced training set, $r_{\boldsymbol{d}(x_{0j})}$ the prediction rule constructed without the omitted set.

If we would split of from the data a validation set, the number of observations would fall below 30. Statistical test and theory prefer the number of samples being at least 30 Montgomery and Runger (2013). The number, where the central limit theorem is starting to work and where the t-distribution approximates a normal distribution. There are 32 failures on one machine, 36 failures on the other. We assume homogeneity, if in 5 cases we observe same or very similar results, we have better confidence.

Bibliography

Abid, R., Kokonendji, C. C., and Masmoudi, A. (2020). Geometric Tweedie regression models for continuous and semicontinuous data with variation phenomenon. *Asta-Advances in Statistical Analysis*, 104(1):33–58.

Acar, E. F., Azimaee, P., and Hoque, M. E. (2019). Predictive assessment of copula models. *Canadian Journal of Statistics*, 47(1):8–26.

Achibi, M., Broniatowski, M., Duveau, C., and Marboeuf, A. (2012). Bivariate Cox models and copulas. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 226(5):476–487.

Allella, F., Chiodo, E., and Pagano, M. (2002). Dynamic discriminant analysis for predictive maintenance of electrical components subjected to stochastic wear. *COMPEL - The international journal for computation and mathematics in electrical and electronic engineering*, 21(1):98–115.

Arnold, B. C. and Strauss, D. J. (1991). Bivariate distributions with conditionals in prescribed exponential families. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):365–375.

Bailey, R. T. (1997). Estimation from zero-failure data. *Risk Analysis*, 17(3):375–380.

Baxora, M. (2021). Processes with independent increments the Poisson process and its properties. *European Journal of Molecular and Clinical Medicine*, 8(2):2124–2128.

Bertsche, B. (2008). *Reliability in automotive and mechanical engineering: determination of component and system reliability.* Springer Science & Business Media.

Bourezaz, H., Merabet, H., and Druilhet, P. (2020). Bayesian prediction modeling for two-stage experimental trials for Poisson or gamma distributed data. *Electronic Journal of Applied Statistical Analysis*, 13(1):268–286.

Bunea, C., Charitos, T., Cooke, R. M., and Becker, G. (2005). Two-stage Bayesian models—application to zedb project. *Reliability Engineering & System Safety*, 90(2-3):123–130.

Cabras, S. (2021). A Bayesian-Deep Learning Model for Estimating COVID-19 Evolution in Spain. *Mathematics*, 9(22).

Caliendo, C., Guida, M., Postiglione, F., and Russo, I. (2021). A Bayesian bivariate hierarchical model with correlated parameters for the analysis of road crashes in Italian tunnels. *Statistical Methods & Applications*, pages 1–23.

Carlin, B. P. and Louis, T. A. (2010). *Bayes and empirical Bayes methods for data analysis*. Chapman and Hall/CRC.

Chen, J., Zhang, Y., and Zhao, P. (2019). Comparisons of order statistics from heterogeneous negative binomial variables with applications. *Statistics*, 53(5):990–1011.

Chen, Y.-C. and Leong, Y.-Y. (2022). *The Effect of Cyber Risk Management Services in Insurance Policies*, volume 10 of *Advances in Pacific Basin Business, Economics and Finance*, page 159–172. Emerald Publishing Limited.

Coolen, F. P. A., Goldstein, M., and Munro, M. (2001). Generalized partition testing via Bayes linear methods. *Information and Software Technology*, 43(13):783–793.

Cooper, J. C. (2005). The Poisson and exponential distributions. *Mathematical Spectrum*, 37(3):123–125.

Corbella, S. and Stretch, D. D. (2013). Simulating a multivariate sea storm using Archimedean copulas. *Coastal Engineering*, 76:68–78.

Corotis, R. B. (2015). An overview of uncertainty concepts related to mechanical and civil engineering. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, 1(4).

de Almeida, A. T. (2001). Multicriteria decision making on maintenance: spares and contracts planning. *European Journal of Operational Research*, 129(2):235–241. publisher: Elsevier.

Dias, L. C., Morton, A., and Quigley, J. (2017). Elicitation: State of the art and science. *Elicitation*.

Diers, D., Eling, M., and Linde, M. (2013). Modeling parameter risk in premium risk in multi-year internal models. *The Journal of Risk Finance*, 14(3):234–250.

Dimitriou, A. (2022). *Empirical Bayesian inference on Poisson processes with a Clayton prior distribution*. Phd thesis, University of Strathclyde.

Dorey, M. and Joubert, P. (2005). Modelling copulas: an overview. *The Staple Inn Actuarial Society*, pages 1–27.

Dupuis, D. (2007). Using copulas in hydrology: Benefits, cautions, and issues. *Journal of Hydrologic Engineering*, 12(4):381–393.

Durrleman, V., Nikeghbali, A., and Roncalli, T. (2000). Which copula is the right one? *Available at SSRN 1032545*.

Dutfoy, A. and Lebrun, R. (2009). Practical approach to dependence modelling using copulas. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 223(4):347–361.

Efron, B. (2019). Bayes, oracle Bayes and empirical Bayes. *Statistical Science*, 34(2):177–201. Zbl: 1420.62023.

Efron, B. and Hastie, T. (2016). *Empirical Bayes*, page 75–88. Cambridge University Press.

Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American statistical association*, 96(456):1151–1160.

Fang, Y. and Madsen, L. (2013). Modified Gaussian pseudo-copula: Applications in insurance and finance. *Insurance: Mathematics and Economics*, 53(1):292–301.

Fantazzini, D. (2009). The effects of misspecified marginals and copulas on computing the value at risk: A Monte Carlo study. *Computational Statistics & Data Analysis*, 53(6):2168–2188.

Frahm, G., Junker, M., and Szimayer, A. (2003). Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286.

Frost, P. A. and Savarino, J. E. (1986). An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, 21(3):293–305.

Furman, E. and Landsman, Z. (2010). Multivariate Tweedie distributions and some related capital-at-risk analyses. *Insurance Mathematics & Economics*, 46(2):351–361.

Gabbiani, F. and Cox, S. J. (2010). Chapter 16 - stochastic processes. In Gabbiani, F. and Cox, S. J., editors, *Mathematics for Neuroscientists*, pages 251–266. Academic Press, London.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Gelman, A. and Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician*, 60(4):328–331.

Genest, C., Quessy, J.-F., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33(2):337–366.

Gerhold, S., Schmock, U., and Warnung, R. (2010). A generalization of Panjer's recursion and numerically stable risk aggregation. *Finance and Stochastics*, 14(1):81–128.

Geurts, J. H. J. and Moonen, J. M. C. (1992). On the robustness of 'insurance type' spares provisioning strategies. *Journal of the Operational Research Society*, 43(1):43–51.

Ghosh, S. (2010). *Copula Based Hierarchical Bayesian Models*. Phd thesis, Texas A & M University.

Givens, G. H. and Hoeting, J. A. (2012). *Computational statistics*, volume 703. John Wiley & Sons.

Gonzales-Barron, U. and Butler, F. (2011). A comparison between the discrete Poisson-gamma and Poisson-lognormal distributions to characterise microbial counts in foods. *Food Control*, 22(8):1279–1286.

Gonzales-Barron, U., Zwietering, M. H., and Butler, F. (2013). A novel derivation of a within-batch sampling plan based on a Poisson-gamma model characterising low microbial counts in foods. *International Journal of Food Microbiology*, 161(2):84–96.

Green, J., Bert, F., and Tukey, J. (1960). Complex analyses of variance: general problems. *Psychometrika*, 25(2):127–152.

Hasan, M. M. and Dunn, P. K. (2011). Two Tweedie distributions that are near-optimal for modelling monthly rainfall in Australia. *International Journal of Climatology*, 31(9):1389–1397.

Hassan, J., Khan, F., and Hasan, M. (2012). A risk-based approach to manage non-repairable spare parts inventory. *Journal of Quality in Maintenance Engineering*, 18(3):344–362.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Hawas, F. and Cifuentes, A. (2017). Valuation of projects with minimum revenue guarantees: A Gaussian copula–based simulation approach. *The Engineering Economist*, 62(1):90–102.

Hemrit, W. and Ben Arab, M. (2012). The determinants of frequency and severity of operational losses in Tunisian insurance industry. *The Journal of Risk Finance*, 13(5):438–475.

Hu, Q., Boylan, J. E., Chen, H., and Labib, A. (2018). Or in spare parts management: A review. *European Journal of Operational Research*, 266(2):395–414.

Jiang, W. and Zhang, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *The Annals of Statistics*, 37(4):1647–1684.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts.* CRC press.

Khosravi, F., Müller, M., Glaß, M., and Teich, J. (2018). Simulation-based uncertainty correlation modeling in reliability analysis. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 232(6):725–737.

Kokonendji, C. C., Mizere, D., and Balakrishnan, N. (2008). Connections of the Poisson weight function to overdispersion and underdispersion. *Journal of Statistical Planning and Inference*, 138(5):1287–1296.

Kvaløy, J. T. and Lindqvist, B. H. (1998). Ttt-based tests for trend in repairable systems data. *Reliability Engineering & System Safety*, 60(1):13–28.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178.

Lawless, J. F. (2011). *Statistical models and methods for lifetime data.* John Wiley & Sons.

Leng, N. and Kendziorski, C. (2015). Ebseq: an empirical Bayes hierarchical model for inference in rna-seq experiments. *Bioinformatics*, 29(8):1035–1043.

Lincoln, H. (2000). Spares parts provisioning and management. In *Reliability, Maintenance and Logistic Support*, pages 281–344. Springer.

Lord, D. (2006). Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. *Accident Analysis & Prevention*, 38(4):751–766.

Louis, T. A. (1991). Using empirical bayes methods in biopharmaceutical research. *Statistics in Medicine*, 10(6):811–829.

Martz, H. F. and Waller, R. (1982). Bayesian Reliability Analysis. *JOHN WILEY & SONS, INC., 605 THIRD AVE., NEW YORK, NY 10158, 1982, 704.*

Minhajuddin, A. T., Harris, I. R., and Schucany, W. R. (2004). Simulating multivariate distributions with specific correlations. *Journal of Statistical Computation and Simulation*, 74(8):599–607.

Montgomery, D. C. and Runger, G. C. (2013). *Applied Statistics and Probability for Engineers*. Wiley, Hoboken, NJ, 6th ed. edition edition.

Mouatassim, Y. (2012). Zero-modified discrete distributions for operational risk modelling. *The Journal of Risk Finance*, 13(5):476–490.

Mukhopadhyay, S. and Fletcher, D. (2018). Generalized empirical Bayes modeling via frequentist goodness of fit. *Scientific Reports*, 8(1):1–15.

Nelsen, R. (2007a). *An Introduction to Copulas*. Springer Series in Statistics. Springer New York.

Nelsen, R., Krickeberg, K., and Fienberg, S. (1999). *An Introduction to Copulas*. Lecture notes in statistics. Springer.

Nelsen, R. B. (2007b). *An introduction to copulas*. Springer Science & Business Media.

Olatubosun, S. and Zhang, Z. (2018). Conditional subjective probability density approach to dependency consideration in passive system reliability. *Nuclear Engineering and Design*, 337:108–115. WOS:000441759300010.

Pan, J., Bai, G., and Chen, W. (2018). Lifetime estimation of nitrile butadiene rubber o-rings under storage conditions using time-varying copula. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 232(6):635–646.

Park, E. S., Oh, R., Ahn, J. Y., and Oh, M.-S. (2021). Bayesian analysis of multivariate crash counts using copulas. *Accident Analysis & Prevention*, 149:105431.

Patil, G. P. and Joshi, S. W. (1968). Dictionary and bibliography of discrete distributions.

Percy, D. F. (2004). Subjective priors for maintenance models. *Journal of Quality in Maintenance Engineering*, 10(3):221–227.

Pörn, K. (1990). *On empirical Bayesian inference applied to Poisson probability models.* Linkoping studies in science and technology. Dissertations ; no.234. Div. of Quality Tech., Dept. of M.E., Linkoping U., Linkoping.

Pratt, J. W. (1976). FY Edgeworth and RA Fisher on the efficiency of maximum likelihood estimation. *The Annals of Statistics*, pages 501–514.

Qarahasanlou, A. N., Barabadi, A., and Ayele, Y. Z. (2018). Production performance analysis during operation phase: A case study. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 232(6):559–575. publisher: SAGE Publications.

Quigley, J., Bedford, T., and Walls, L. (2007). Estimating rate of occurrence of rare events with empirical bayes: A railway application. *Reliability Engineering & System Safety*, 92(5):619–627.

Quigley, J., Hardman, G., Bedford, T., and Walls, L. (2011). Merging expert and empirical data for rare event frequency estimation: Pool homogenisation for empirical bayes models. *Reliability Engineering & System Safety*, 96(6):687–695.

Quigley, J. and Walls, L. (2011). Mixing Bayes and empirical Bayes inference to anticipate the realization of engineering concerns about variant system designs. *Reliability Engineering & System Safety*, 96(8):933–941.

Quigley, J. and Walls, L. (2018). A methodology for constructing subjective probability distributions with data. In *Elicitation*, pages 141–170. Springer.

Quigley, J. and Walls, L. (2021). *Characteristics of a process for subjective probability elicitation*, pages 287–318. International Series in Operations Research and Management Science. Springer.

Quigley, J., Walls, L., Demirel, G., MacCarthy, B. L., and Parsa, M. (2018). Supplier quality improvement: The value of information under uncertainty. *European journal of operational research*, 264(3):932–947.

Quigley, J., Wilson, K. J., Walls, L., and Bedford, T. (2013). A Bayes linear Bayes method for estimation of correlated event rates. *Risk analysis*, 33(12):2209–2224.

Ross, S. (2006). *Introduction to Probability Models*. Elsevier Science, 10th edition edition.

Ross, S. M. (1995). *Stochastic processes*. John Wiley & Sons.

Ruppert, D. (2011). *Statistics and data analysis for financial engineering*, volume 13. Springer.

Savolainen, P. T., Mannering, F. L., Lord, D., and Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. *Accident Analysis & Prevention*, 43(5):1666–1676.

Schwarzenegger, R., Quigley, J., and Walls, L. (2021). Is eliciting dependency worth the effort? A study for the multivariate Poisson-Gamma probability model. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability.*

Schweizer, B. (1991). *Thirty years of copulas*, page 13–50. Springer.

Scott, J. G. and Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, 38(5):2587–2619.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Song, J.-S. (2002). Order-based backorders and their implications in multi-item inventory systems. *Management Science*, 48(4):499–516.

Trivedi, P. K. and Zimmer, D. M. (2007). *Copula modeling: an introduction for practitioners*. Now Publishers Inc.

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., and Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(11):1–26.

Velleman, P. F. and Welsch, R. E. (1981). Efficient computing of regression diagnostics. *The American Statistician*, 35(4):234–242.

Venter, G. G. (2002). Tails of copulas. In *Proceedings of the Casualty Actuarial Society*, volume 89, pages 68–113.

Wali, B., Khattak, A. J., and Xu, J. (2018). Contributory fault and level of personal injury to drivers involved in head-on collisions: application of copula-based bivariate ordinal models. *Accident Analysis & Prevention*, 110:101–114.

Walls, L. and Quigley, J. (1999). Learning to improve reliability during system development. *European Journal of Operational Research*, 119(2):495–509.

Wang, J., Gupta, M., and Rao, H. R. (2015). Insider threats in a financial institution: Analysis of attack-proneness of information systems applications. *Mis Quarterly*, 39(1):91–U491.

Wang, K. H. and Sivazlian, B. D. (1992). Cost analysis of the m/m/r machine repair problem with spares operating under variable service rates. *Microelectronics Reliability*, 32(8):1171–1183.

Wang, X., Wang, B., Chang, M., and Li, L. (2020). Reliability and sensitivity analysis for bearings considering the correlation of multiple failure modes by mixed copula function. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 234(1):15–26.

Welker, E. L. and Lipow, M. (1974). Estimating the exponential failure rate from data with no failure events. In *Annual Reliability and Maintainability Symposium, Los Angeles, Calif*, pages 420–427.

Werner, C., Bedford, T., Cooke, R. M., Hanea, A. M., and Morales-Napoles, O. (2017). Expert judgement for dependence in probabilistic modelling: A systematic literature review and future research directions. *European Journal of Operational Research*, 258(3):801–819.

Werner, C., Colson, A., Morton, A., and Bedford, T. (2021). Risk assessment of future antibiotic resistance—eliciting and modelling probabilistic dependencies between multivariate uncertainties of bug-drug combinations. *Frontiers in Applied Mathematics and Statistics*, 7.

Yahav, I. and Shmueli, G. (2012). On generating multivariate Poisson data in management science applications. *Applied Stochastic Models in Business and Industry*, 28(1):91–102.

Yan, J. et al. (2007). Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software*, 21(4):1–21.

Yasmin, S., Momtaz, S. U., Nashad, T., and Eluru, N. (2018). A multivariate copula-based macro-level crash count model. *Transportation research record*, 2672(30):64–75.

Zahedi-Hosseini, F., Scarf, P., and Syntetos, A. (2017). Joint optimisation of inspection maintenance and spare parts provisioning: a comparative study of inventory policies using simulation and survey data. *Reliability Engineering & System Safety*, 168:306–316.

Zhang, C.-H. (2003). Compound decision theory and empirical Bayes methods: invited paper. *The Annals of Statistics*, 31(2):379–390.

Zhang, X.-D. (2022). Modern signal processing. In *Modern Signal Processing*. De Gruyter.

Zhou, Y., Lu, Z., Shi, Y., and Cheng, K. (2019). The copula-based method for statistical analysis of step-stress accelerated life test with dependent competing failure modes. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 233(3):401–418.