Eindhoven University of Technology

MASTER

Measurement system analysis techniques for non-normal processes parameter estimation of the Conway-Maxwell-Poisson latent variable model

Paus, M.H.J.

Award date:
2018

Link to publication

Technische Universiteit
Eindhoven
University of Technology

Department of Mathematics and Computer Science

Measurement system analysis techniques for non-normal processes

Parameter estimation of the Conway-Maxwell-Poisson latent variable model

Master thesis
by
Michiel Hans Joost Paus

Supervisors:
Prof. dr. E.R. van den Heuvel   TU/e
dr. R.A. Ion   ASML
ir. F. van Beek   ASML

In cooperation with:

ASML

Eindhoven, November 2017

Contents

Introduction

In the last decades, the amount of collected measurement data has grown exponentially and businesses have become more and more data driven. All types of industries, from manufacturing, pharmacology, medical to financial, use measurement data to make data-based decisions on the quality of materials, products or processes. For example, measurement data is evaluated against standards or criteria such as statistical control limits or used to investigate relationships between responses and factors. The quality of these decisions is positively correlated with the quality of the measurement data, which is a strong determinant for the quality of the measurement system. To make high quality decisions, measurement data should be collected from a high quality measurement system. The definition of a measurement system used by the (AIAG) (2010) is "the collection of instrument or gages, standards, operations, methods, fixtures, software, personnel, environment and assumptions to quantify a unit of measure or fix assessment to the feature characteristic being measured; the complete process used to obtain measurements".

Measurement system analysis (MSA) is a vital part of various quality management methodologies and standards that have been developed over the years (Li & Al-Refaie, 2008). Among others we have ISO 9001 from the International Standards Organization, Total Quality Management (TQM) with many tools and techniques for quality control and Six Sigma, the management strategy developed by Motorola in 1986. All methodologies have their own focus and ideology, but are not mutually exclusive and can be used alongside each other as they are complementary and synergistic. In specific, Six Sigma is a methodology that provides business with the tools to improve the capability of their business processes. Here, a process could be a product or a service that a company provides to outside customers, or it could be an internal process within the company. It distinguishes five phases, namely Design, Measure, Analyze, Improve, and Control (DMAIC) (Basem, 2008).

In order to improve the quality of decisions based on the measurement data, measurement systems should be thoroughly evaluated. In a measurement system analysis, the measurement system is assessed by means of an experiment that aims to both identify and quantify the different sources of variation in the product or process. In the controlled setting of an experiment, randomness avoids confounding effects to influence the results of the analysis. The key statistical properties on which a measurement system is evaluated are bias, linearity, stability, repeatability and reproducibility ((AIAG), 2010). Where the first three characteristics are location related, the latter two describe the variation (precision) of the measurement system. In particular, reproducibility represents the variability with different units (operators/set-ups/time-periods), repeatability is the variability from measurement instrument (on same units). Moreover, when the variation of the measurement system is small relative to the total variation, the measurement system is considered capable.

1.1 MSA within ASML

ASML is the world leading manufacturer of lithography systems for the semiconductor industry and is currently in a major technology transition by introducing extreme ultraviolet (EUV) lithography for high volume chip production. The high importance of quality at ASML makes that the Six Sigma methodology is widely used within the organization and that measurement system analyses are performed for many ASML projects. An example of measurement system analysis for count data is the qualification of EUV pellicles.

One of ASML's key priorities in the manufacturing of EUV machines is having reticles, which contain the blueprint of the pattern for the chips, that are free of particles. To keep the reticles clean of particles and bring them out of focus, an (EUV-)pellicle is placed in front of the reticle, see Figure 1.1. The challenge for EUV is to manufacture clean and defect-free films (pellicles) with high EUV transmission. In the manufacturing process of pellicles, defects such as holes can be created and particles can be added or lost.
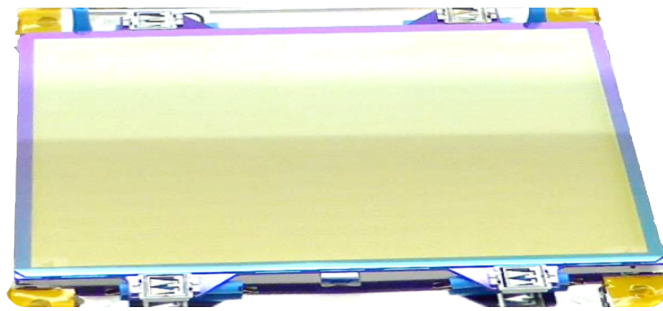
Figure 1.1: An extreme ultra violet (EUV) pellicle

These impurities on the surface are captured in an image that is constructed by a digital microscope and processed by software into a set of classified defects. Finally, a summarizing report of a pellicle, with for example total number of defects per size category and defect type, is automatically produced. When the reported number of defects of two consequent measurements performed at different steps in the process differs, a manual analysis will follow to determine the cause of the difference. The main focus is on defects with a diameter size larger than 10 arbitrary units (a.u.).[1]

In order to check the quality of a pellicle and the production process, pellicles are repeatedly measured during the process. As a result of variation due to measurement system, discrepancies between the true and observed number of defects might appear. The smaller this difference, the better the measurement system is. In the ideal case, if the measurement system is perfect, the number of defects measured will be the same for all repeats on that particular part, and equal to the actual number of defects. In practice, the actual number of defects on a pellicle (larger than a certain size) is unknown, however, by nature of a count, this number should follow a discrete distribution. Besides, the conditions under which measurements are done could strongly vary, due to the use of different operators, measurement systems etcetera. In order to quantify the sources of variation and determine the contribution of the measurement system variation to the total variation, an experiment should be performed. This experiment should involve repeated measurements on different tools with various (randomly selected) parts and operators. The design that is planned to be executed is as follows

6 parts × 3 tools × 2 operators × 3 repeats.

In absence of the data of a full sized gauge repeatability and reprodicubility (GR&R) experiment, three historical datasets from ASML were analyzed to define a suitable underlying statistical

---

[1]For the sake of confidentiality of ASML technology, we use arbitrary units.

model. The first dataset contains five repeated measurements on two tools each using the same part. So conditional on an expected count on a tool, variation in the measurements is observed. Due to the effect of the light intensity of the microscope on the size of the defects, the observed number of defects larger than a certain diameter size can deviate from the actual number of defects on a pellicle in both directions. A discrete distribution where the actual number of defects is represented by the mean, such as a Poisson distribution or a binomial distribution, would be straightforward. The binomial distribution was not considered as it requires the selection of a maximum value.

When fitting a Poisson regression model on this data, a lack of fit was observed since the Poisson distribution was not capable of capturing the severe under-dispersion present in the data. This under-dispersion indicates that the variance is much smaller than one would expect based on the (mean of a) Poisson distribution. A straightforward extension to the Poisson distribution such that it can cope with under-dispersion is the quasi-Poisson model, that includes an additional dispersion parameter. For the quasi-Poisson regression model, a dispersion of 0.019 was observed, which implies that the variance is more than a factor 50 smaller than the mean. The second and third dataset consist of a cross-sectional analysis of the number of defects larger than 10 a.u. of 141 pellicles. The hypothesis is that the distribution of the counts of defects on pellicles is described by an (over-dispersed) Poisson (e.g. a negative binomial). More background on these findings and a more thorough analysis of both datasets can be found in Chapter 4.

1.2 Statistical techniques for MSA

The statistical techniques that are mainly used for MSA in the Six Sigma framework to assess the gauge's reproducibility and repeatability are the analysis of variance (ANOVA) GR&R and Average and Range ($\bar{X}$ and R) ((AIAG), 2010). The ANOVA GR&R uses the ANOVA theory that was developed by Sir Ronald Fisher to study agricultural experiments. Here, the significance of model effects is assessed by means of an F-test using the mean squares that are obtained by dividing the sum of squares with the corresponding degrees of freedom. Moreover, the %-Contribution, %-Study Variation and %-Tolerance of the part to part and total gage, consisting of all R&R model components, are presented. Three types of ANOVA can be distinguished, namely fixed effects ANOVA, random effects ANOVA and mixed effects ANOVA. Note that all ANOVA types strongly depends on normality of the residuals and fall under the category of linear (mixed) models. The $\bar{X}$ and R method only allows for a traditional two-factor design, does not provide confidence intervals, and is more reliable for GR&R analysis when residuals are non-normal (Osma, 2011).

In case the underlying distribution of the response variable is not normally distributed, the more sophisticated generalized linear models (GLMs) or generalized linear mixed models (GLMM) could be applied (McCullagh & Nelder, 1989). These models that are not included in the Six Sigma methodology, connect the mean via a link function to a linear combination of the model parameters. If all model parameters are considered fixed constants we have a GLM and in presence of random effects, we have a GLMM. Here, it is assumed that the random effects follow a normal distribution. By relaxing the restricting of the random effect in the linear predictor of the GLM to any distribution, a hierarchical generalized linear model (HGLM) is obtained (Lee & Nelder, 1996).

1.3 Thesis objectives and overview

Thesis objectives

In this thesis we aim to extend the existing framework for assessing the repeatability and reproducibility for univariate count data, to the setting where the mean is significantly higher than the variance, also called severely under-dispersed. Whereas a straightforward (quasi-)Poisson regression model would suffice when individual parts are analyzed, the estimation of the maximum

likelihood estimates becomes less trivial in case of multiple parts. Additional complexity is obtained when the distribution of counts on parts is not normally distributed, but from some discrete distribution. This connects to the concept of hierarchical generalized linear models. Now, conditional on the true counts on these parts, which is considered a latent variable, a distribution for under-dispersed count data is demanded. To model the severe under-dispersion within parts, we mainly focus on the Conway Maxwell Poisson (CMP) distribution, due to some undesirable properties of alternatives that will be explained in the next chapter. The likelihood of this probability model with a CMP distribution contains a normalization constant in terms of an infinite sum that is dependent on the latent variable. This makes the estimation of the maximum likelihood estimates a non-trivial problem that cannot be evaluated analytically. We will present various estimation methods to obtain approximations for the parameter estimates of the probability model. By means of an extensive simulation study, these estimation techniques are compared and assessed on their estimating and testing performance. Moreover, the adequacy of selecting a continuous prior distribution, such as a normal or gamma distribution, is investigated. Furthermore, recommendations on the experimental design are provided.

Outline thesis

The content of the remainder of this thesis is as follows. Chapter 2 starts with a thorough description of the literature on distributions for under-dispersed count data. Subsequently, the Conway-Maxwell Poisson (CMP) and Generalized Poisson (GP) distribution are extensively described by means of their parametrizations and (approximated) moments. For the CMP, also various approximations of the normalization constant are listed. At the end of this chapter, this existing theory is brought together to define two statistical models for count data with severe under-dispersion in a repeated measurement setting. Here, the probability models with their likelihoods are presented and the corresponding issues are addressed.

In Chapter 3, we consider estimation techniques based on (pseudo-)likelihood optimization and the method of moments. Both approaches utilize the different approximations of the CMP normalization constant. For the likelihood methods, a simplified model with a quasi-Poisson repeated measurement and a normally distributed latent variable is included. This to prove the added value of the new methods compared to this existing approach. For the methods of moments, we assess the performance of the F-test, as used in ANOVA, on data that is not normally distributed, but comes from a CMP distribution.

Chapter 4 discusses the case study where three datasets are considered: one with repeated measurements of a single part on two tools and a two cross-sectional datasets. On the former dataset the lack of fit of a Poisson distribution to model the repeated measurements is shown, a CMP distribution is fitted and the tool effects are determined. The latter datasets are used to derive the distribution of counts on parts and demonstrate a the heterogeneity of variances. The main reason for this separate analysis was the absence of a full R&R experiment with repeated measurements on multiple parts.

To examine the performance of the estimation techniques developed in Chapter 3, an extensive simulation study is performed in Chapter 5. The generated datasets use the key findings of the case study as input. The additional simulation settings and essential function used will be discussed in detail. In particular, over 20 simulation scenarios are examined, involving different experimental setups, main effects, latent variable distributions and levels of under-dispersion. Furthermore, the analysis of these simulated experiments will be explained, as well as the selection of the starting values. After this, all results are listed.

The conclusions and recommendations can be found in Chapter 5. In Appendix A, an overview of all proofs of lemmas, propositions and extensive derivations of equations is provided. In addition, some customized algorithms and functions written in the statistical software [R] are listed in Appendix B. Appendix C is dedicated to the CMP normalization constant. Here, the original forms

of the approximations of the CMP normalization, as well as the performance of the approximations in case of severe under-dispersion are presented.

Underlying statistical model

In this chapter, we describe the characteristics of the statistical model for repeatedly measuring counts on different subjects (parts). Firstly, we describe the Conway-Maxwell-Poisson and Generalized Poisson distribution to model the situation that the expected mean count is significantly higher than the expected variance of this count. After introducing these distributions with its characteristics, the full probability model and the corresponding likelihoods will be presented. Note that these models form the basis of the simulation study that will be performed in Chapter 5.

2.1 Literature overview on statistical count models

One of the most extensively studied and exploited distributions for count data is the Poisson distribution (Shmueli et al., 2005). An important characteristic of the Poisson distribution is its equi-dispersion, implying that the variance equals the mean. In practice however, this equivalence does often not hold and the variation in the data is larger or smaller relative to the Poisson distribution, respectively over-dispersed and under-dispersed. The negative binomial distribution, which is a Gamma-Poisson mixture, is a popular alternative to the Poisson distribution in case of over-dispersion, but is not able to capture under-dispersion. The same holds for some other mixed Poisson models such as the Poisson-inverse Gaussian and Poisson-lognormal. Models that are capable of handling both under- and over-dispersion are the Generalized Poisson (GP) (Consul & Jain, 1973) and the Conway-Maxwell-Poisson (CMP) (Conway & Maxwell, 1962). Both models have an additional parameter to model the dispersion and have the Poisson as a special case. In addition, there is the quasi-Poisson, where dispersion parameter is not part of the likelihood but derived by the ratio of the residual deviance and the residual degrees of freedom.

An appealing characteristic of the GP distribution is that the mean can directly be expressed in terms of both parameters. Zamani & Ismail (2012) rewrite the GP to a more suitable form for a regression model, with the mean is expressed by a single parameter. A disadvantage of the GP however, is that for some levels of under-dispersion the distribution is truncated, making it not a true probability model (Famoye, 1993). The data dependent support makes the estimation of the GP parameters in a regression setting particularly problematic (Czado et al., 2007; Huang, 2016). The CMP distribution was initially developed for queuing models with state dependent service rates. In contrast to the GP, CMP can cover a wide range of under-dispersion levels. A downside of CMP is that due to the normalization constant, the expected value does in general not have a simple form. However, various link functions that connect the linear combination of the covariates with the expectation are suggested (Sellers et al., 2012). In addition, approximations for the normalization constant are derived, see Paragraph 2.2.

Numerous examples exist of generalized linear models (GLM) for under-dispersed count data using a Conway-Maxwell-Poisson distribution (Sellers & Shmueli, 2010; Jowaheer & Khan, 2009).

Among others, in Jowaheer & Khan (2009), a CMP GLM was fitted using maximum likelihood (ML) optimization and quasi-likelihood (QL) optimization using the first order approximations of the moments. Next to the fact that the ML approach was significantly slower than QL, convergence issues occurred for both methods, although more frequent for the ML. The non-convergence was attributed to a combination of under-dispersion and small sample sizes. For large sample sizes this convergence problem did not appear. Combined with a minor efficiency loss (1%) on the estimates, the QL was considered the superior method. Even though that the maximum likelihood estimates of a GLM do not have a closed form, the computational complexity remains gentle as various bounded nonlinear minimization or optimization procedures could be employed. The ML estimates can obtained with great accuracy using for example the procedures *nlminb*, *nlm* or *optim* from the statistical software [R] (Sellers & Raim, 2016). Moreover, for under-dispersion the summation of the CMP normalization constant quickly converges, requiring numerically only a small number of terms to obtain accurate estimates.

However, despite a thorough literature review, nothing was found on the combination of GLMMs and under-dispersed count data.

2.2 Conway-Maxwell-Poisson distribution

Parameterization

The probability mass function of the CMP distribution, as defined by Shmueli et al. (2005) is

$$P(X = x) = \frac{1}{Z(\lambda, \nu)} \frac{\lambda^x}{(x!)^\nu}, \qquad x = 0, 1, 2, \dots \tag{2.1}$$

with normalization constant

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}, \tag{2.2}$$

for a random variable X, where $\nu \geq 0$ is the dispersion parameter. For $\nu = 1$, CMP reduces to the Poisson distribution, since $Z(\lambda, 1) = \exp\{\lambda\}$. Moreover, $\nu < 1$ and $\nu > 1$ denote over-dispersion and under-dispersion, respectively. Two properties of the CMP distribution are that it belongs to the exponential family and is part of the two parameter power series distributions. The moments defined by Shmueli et al. (2005) are of the form

$$\mathbb{E}(X^{r+1}) = \begin{cases} \lambda[\mathbb{E}(X+1)]^{1-\nu} & r = 0 \\ \lambda \frac{\partial}{\partial \lambda} \mathbb{E}(X^r) + \mathbb{E}(X)\mathbb{E}(X^r) & r > 0. \end{cases}$$

In particular, the expected value and variance are

$$\mathbb{E}[X] = \lambda \frac{\partial \log Z(\lambda, \nu)}{\partial \lambda} \quad \text{and} \quad \text{Var}(X) = \frac{\partial \mathbb{E}(X)}{\partial \log \lambda}. \tag{2.3}$$

Moreover, the first order approximations are $\mathbb{E}[X] \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu}$ and $\text{Var}(X) \approx \frac{1}{\nu} \lambda^{1/\nu}$. The approximation for the mean is particularly accurate for $\nu \leq 1$ or $\lambda > 10^\nu$.

Reformulation of the CMP

Another parameterization of CMP that has proven to be particularly useful in the generalized linear model setting was proposed by Guikema & Goffelt (2008). Here, we substitute $\zeta = \lambda^{1/\nu}$ in (2.1) and (2.2) and obtain

$$P(X = x) = \frac{1}{S(\zeta, \nu)} \left(\frac{\zeta^x}{x!} \right)^\nu, \quad x = 0, 1, 2, \dots \tag{2.4}$$

with normalization constant

$$S(\zeta, \nu) = \sum_{j=0}^{\infty} \left(\frac{\zeta^j}{j!} \right)^{\nu}. \tag{2.5}$$

Note that we again have that $S(\zeta, 1) = \exp\{\zeta\}$. The mean and variance of this reformulation are defined by

$$\mathbb{E}(X) = \frac{1}{\nu} \frac{\partial \log S(\zeta, \nu)}{\partial \log \zeta} = \frac{\sum_{j=0}^{\infty} j \left(\frac{\zeta^j}{j!} \right)^{\nu}}{\sum_{j=0}^{\infty} \left(\frac{\zeta^j}{(j!)} \right)^{\nu}}$$

$$\text{Var}(X) = \frac{1}{\nu^2} \frac{\partial^2 \log S(\zeta, \nu)}{\partial \log^2 \zeta} = \frac{\left[\sum_{j=0}^{\infty} \left(\frac{\zeta^j}{j!} \right)^{\nu} \right] \cdot \left[\sum_{j=0}^{\infty} j^2 \left(\frac{\zeta^j}{j!} \right)^{\nu} \right] - \left[\sum_{j=0}^{\infty} j \left(\frac{\zeta^j}{j!} \right)^{\nu} \right]^2}{\left(\sum_{j=0}^{\infty} \left(\frac{\zeta^j}{j!} \right)^{\nu} \right)^2}.$$

The advantage of this representation is that the parameters ζ and ν have a direct link to the mean and the variance, since their asymptotic approximations are $\mathbb{E}[X] \approx \zeta - \frac{\nu-1}{2\nu}$ and $\text{Var}(X) \approx \zeta/\nu$. The region for which this first order approximation of the mean is particularly accurate is $\nu \leq 1$ or $\zeta > 10$. This parameterization will be hereafter used as the standard representation of CMP.

Approximation normalization constant

Over time, various approximation methods are proposed for the normalization constant defined in (2.2). The first, which is the most simple version and that is suggested by Minka et al. (2003), is a truncation of the infinite series at the mth term. For the reformulated normalization constant (2.5), this gives

$$S(\zeta, \nu) = \sum_{j=0}^{m} \left(\frac{\zeta^j}{j!} \right)^{\nu} + R_m, \tag{2.6}$$

where $R_m = \sum_{j=m+1}^{\infty} \left(\frac{\zeta^j}{j!} \right)^{\nu}$ is the absolute truncation error.

Secondly, an asymptotic approximation for $Z(\lambda, \nu)$ involving Laplace approximation of a $(\nu-1)$-dimensional integral representation was proposed by Shmueli et al. (2005). This integral representation only applies to positive integer values of ν. After substituting $\zeta = \lambda^{1/\nu}$ in the approximation for $S(\zeta, \nu)$ we obtain

$$S(\zeta, \nu) = \frac{\zeta^{(1-\nu)/2} \cdot e^{\nu\zeta}}{(2\pi)^{(\nu-1)/2}\sqrt{\nu}} \{1 + \mathcal{O}(1/\zeta)\}. \tag{2.7}$$

In Gillispie & Green (2015), this first order approximation was compared with the truncated summation (2.6). For small values of $\zeta = \lambda^{1/\nu}$ the truncated summation was found to be most accurate way to calculate $S(\zeta, \nu)$. However, for large ζ it was stated that the first order approximation might be better than the truncated variant, that requires a summation of a relatively high number of terms, which could be incorrect due to aggregating multiple rounding errors.

Thirdly, in addition to the first order expansion in (2.7) that is restricted to integer values of ν, Gaunt et al. (2016) provided an entire asymptotic expansion for $Z(\lambda, \nu)$ for all real non-negative values of $\nu > 0$. An important note on asymptotic expansions is that the approximations not necessarily have to get better when the number of included terms gets higher. By substituting $\zeta = \lambda^{1/\nu}$ in the full asymptotic expansion for $Z(\lambda, \nu)$ we obtain

$$S(\zeta, \nu) = \frac{\zeta^{(1-\nu)/2} \cdot e^{\nu\zeta}}{(2\pi)^{(\nu-1)/2}\sqrt{\nu}} \sum_{k=0}^{\infty} c_k (\nu\zeta)^{-k}, \qquad \text{as } \zeta \to \infty \quad (\text{actually } \zeta^{\nu} \to \infty) \tag{2.8}$$

where the c_j are uniquely determined by the expansion

$$(\Gamma(t+1))^{-\nu} = \frac{\nu^{\nu(t+1)/2}}{(2\pi)^{(\nu-1)/2}} \sum_{j=0}^{\infty} \frac{c_j}{\Gamma(\nu t + (1+\nu)/2 + j)}. \tag{2.9}$$

In particular, $c_0 = 1$, $c_1 = \frac{\nu^2 - 1}{24}$, $c_2 = \frac{\nu^2 - 1}{1152}(\nu^2 + 23)$. More coefficients derived from the expansion (2.9) can be found in Appendix D.5. This appendix also contains all the above mentioned approximation methods of the original for of the normalization constant $Z(\lambda, \nu)$.

Moment approximations

By using the expansion of the approximation constant (2.8), Gaunt et al. (2016) derived more accurate approximations of the expectation and variance of a Conway-Maxwell-Poisson random variable. More specific, it uses the exact expressions for the expectation and variance of a CMP random variable in (2.3) and substitutes the normalization constant with an order n asymptotic expansion. The provided third order approximation of the moments in terms of ζ and ν are

$$\mathbb{E}[X] = \zeta \left(1 - \frac{\nu - 1}{2\nu} \zeta^{-1} - \frac{\nu^2 - 1}{24\nu^2} \zeta^{-2} - \frac{\nu^2 - 1}{24\nu^3} \zeta^{-3} + \mathcal{O}(\zeta^{-4}) \right),$$

$$\text{Var}(X) = \frac{\zeta}{\nu} \left(1 + \frac{\nu^2 - 1}{24\nu^2} \zeta^{-2} + \frac{\nu^2 - 1}{12\nu^3} \zeta^{-3} + \mathcal{O}(\zeta^{-4}) \right).$$

2.3 Generalized Poisson distribution

Parameterization

The concept of Generalized Poisson distributions (GP) was first introduced by Consul & Jain (1973) with a probability distribution given by

$$P(X = x) = \begin{cases} \frac{\theta(\theta + \lambda x)^{x-1}}{x!} \exp\{-\theta - \lambda x\} & \text{if } \lambda \geq 0 \wedge x = 0, 1, 2, \ldots \\ & \text{or } \lambda < 0 \wedge x = 0, 1, 2, \ldots, m \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 0$, $\max(-1, -\theta/m) \leq \lambda < 1$ and $m(\geq 4)$ is the largest positive integers for which $\theta + m\lambda > 0$ when λ is negative. In this probability mass function, negative values of λ correspond to under-dispersion, positive values to over-dispersion and GP reduces to a Poisson distribution for $\lambda = 0$.

The GP parameters θ and λ are independent but the lower limits on λ and $m \geq 4$ are imposed to ensure that there are at least five classes with nonzero probability when λ is negative. Note that truncation is applied for under-dispersion ($\lambda < 0$), causing the sum of all GP probability masses to be slightly less than 1. This truncation error can be eliminated by a normalization where we multiply $P(X = x)$ with $P(X \leq m)^{-1}$ (Consul & Famoye, 2006; Consul & Shoukri, 1985). This will be denoted by $P^*(X = x)$. In case of under-dispersion the moments are

$$\mathbb{E}^*[X] = \frac{\sum_x x P(X = x)}{\sum_x P(X = x)},$$

$$\text{Var}^*(X) = \frac{\sum_x x^2 P(X = x)}{\sum_x P(X = x)} - \left(\frac{\sum_x x P(X = x)}{\sum_x P(X = x)} \right)^2. \tag{2.10}$$

For non-negative λ the mean and variance are $\mathbb{E}(X) = (1 - \lambda)^{-1}\theta$ and $\text{Var}(X) = (1 - \lambda)^{-3}\theta$, respectively.

In Consul & Shoukri (1985), it was stated that the effect of the truncation on the moments was substantial when the number of non-zero classes is only 3 or 4 and $0.7 \leq \theta \leq 4.5$. Therefore, by using the lower limit $\lambda = -1$ in combination with an expected high count, we see that

$$\text{Var}(X) = \mathbb{E}[X](1 - \lambda)^{-2} \geq 1/4\mathbb{E}[X].$$

Finally, note that for all $\lambda \neq 0$, the GP distribution does not belong to the exponential family even if the dispersion parameter λ is known.

Reformulation of the GP

Another parameterization for the GP distribution was suggested by Zamani & Ismail (2012) and is obtained using $\theta = (1+\varphi)^{-1}\tau$ and $\lambda = (1+\varphi)^{-1}\varphi$. The mean and variance for the untruncated GP distribution are $\mathbb{E}(X) = \tau$ and $\text{Var}(X) = (1+\varphi)^2\tau$, where φ denotes the dispersion parameter. We have under-dispersion for $\varphi < 0$ and over-dispersion for $\varphi > 0$. The probability mass function of the GP distribution is described by

$$P(X = x) = \begin{cases} \frac{\tau(\tau+\varphi x)^{(x-1)}}{(1+\varphi)^x x!} \exp\{-\frac{\tau+\varphi x}{1+\varphi}\} & \text{if } \varphi \geq 0 \wedge x = 0, 1, 2, \ldots \\ & \text{or } \varphi < 0 \wedge x = 0, 1, 2, \ldots, m \\ 0 & \text{otherwise} \end{cases}$$

where $\tau > 0$, $\varphi > \max\{-1/2, -\tau/m\}$, and $m(\geq 4)$ the largest integer x such that $\tau + \varphi x > 0$. The truncated moments are defined by (2.10).

2.4 Statistical models for repeated measurements

In this paragraph, we aim to define a suitable statistical model that captures the behavior of the full measurement setting that is used to detect a certain number of defects on EUV pellicles (parts). The measurement system analysis consist of an experiment on the parts $i = 1, \ldots, I$ and repeat $j = 1, \ldots, J$, where a repeat represents a certain measurement setting of tool and operator.

Probabilistic models

Let Z_i be a random variable representing the true count of defects on the parts $i = 1, \ldots, I$ and Y_{ij} be the random variable that describes the measured count on part i in repeat $j = 1, \ldots, J$. Here, the repeat actually consist of various measurement settings of tools and operators that are each measured k times. Note that these tools and operators are considered fixed effects. Moreover, the true count on a part is unobserved and therefore a latent variable. Although, the true count is a discrete random variable, we will use a general formulation such that Z_i could represent any random variable. We now have that

$$P(Y_{ij} = y; \boldsymbol{x}_{ij}) = \int_{z=0}^{\infty} P(Y_{ij} = y | Z_i = z, \boldsymbol{x}_{ij}) dF_{Z_i}(z) dz, \tag{2.11}$$

where $Y_{ij}|Z_i$ denotes the conditional random variable representing the repeated measurements given a true count and all Z_i are independently and identically distributed. Note that $\boldsymbol{x}_{ij} = x_{ij1}, \ldots, x_{ijp}$ is the vector of binary values, that is used to indicate which covariates (or measurement settings) are used to measure part i and repeat j. Hereafter, we will suppress this vector in the notation of $P(Y_{ij} = y)$ and $P(Y_{ij} = y | Z_i = z)$.

Based on historical data, see Chapter 4, we assume that repeated measurements are either distributed with a Conway-Maxwell-Poisson or Generalized Poisson distribution, denoted by $\text{CMP}(\zeta, \nu)$ and $\text{GP}(\tau, \varphi)$ respectively. Moreover, heterogeneity of the variances is detected, indicating a multiplicative effect where more variation in the repeated measurements is expected when the true count is higher. Assuming that the measurement system is capable, we should have that true count on a part equals the expected count on a certain measurement setting. These characteristics are incorporated in the following probabilistic models for a CMP and GP distribution, defined by

$$P(Y_{ij} = y) = \int_{z=0}^{\infty} \text{CMP}(\zeta_{ij}(z), \nu_{ij}) dF_{Z_i}(z) dz,$$
$$\zeta_{ij}(z) = z \cdot \exp\{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}\} \qquad \text{and} \qquad \nu_{ij} = \exp\{\boldsymbol{x}_{ij}^T \boldsymbol{\gamma}\}. \tag{2.12}$$

for a CMP distribution and

$$P(Y_{ij} = y) = \int_{z=0}^{\infty} \text{GP}(\tau_{ij}(z), \varphi_{ij}) dF_{Z_i}(z) dz,$$
$$\tau_{ij}(z) = z \cdot \exp\{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}\} \qquad \text{and} \qquad \varphi_{ij} = \exp\{\boldsymbol{x}_{ij}^T \boldsymbol{\gamma}\}. \tag{2.13}$$

for a GP distribution. Here, $\boldsymbol{x}_{ij} \in \mathbb{R}^{p \times 1}$, $\boldsymbol{\beta} \in \mathbb{R}^{p \times 1}$, $\boldsymbol{\gamma} \in \mathbb{R}^{p \times 1}$, with p the number of covariates that are included in the model. Here, $\boldsymbol{\beta}$ is the vector with tool and operator effects on the location and $\boldsymbol{\gamma}$ the vector with tool effects on the dispersion. Note that the latent variable value z is inherited in the location parameters $\zeta_{ij}(z)$ and $\tau_{ij}(z)$ of the CMP and GP, respectively. In addition, the location parameter is connected with the systematic part of the model by means of a logarithmic link function. This logarithmic link function allows for a model that is based on the proportional difference of tools and operators with the true count.

When $\boldsymbol{\gamma} = \boldsymbol{0}$, with $\boldsymbol{0} \in \mathbb{R}^{p \times 1}$, we obtain equi-dispersion for both the CMP and GP distribution, which transforms the probabilistic models (2.12) and (2.13) into a Poisson regression model with the latent variable Z_i.

Likelihood

Given the probability model for repeated measurements following a CMP distribution in (2.12), the likelihood in the most general form is defined by

$$L(\theta; y_{11}, \ldots, y_{IJ}) = \prod_{i=1}^{I} \int_{z=0}^{\infty} \prod_{j=1}^{J} \frac{1}{S(\zeta_{ij}(z), \nu_{ij})} \left(\frac{\zeta_{ij}(z)^{y_{ij}}}{y_{ij}!} \right)^{\nu_{ij}} dF_{Z_i}(z) dz, \tag{2.14}$$

$$\zeta_{ij}(z) = z \cdot \exp\{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}\} \qquad \text{and} \qquad \nu_{ij} = \exp\{\boldsymbol{x}_{ij}^T \boldsymbol{\gamma}\},$$

where $\theta = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$. Here, $\boldsymbol{\delta}$ denotes the vector of parameters corresponding to the distribution of the latent variable Z_i. Since the counts are actually described by discrete distribution for the latent variable Z_i, we have that (2.14) reduces to

$$L(\theta; y_{11}, \ldots, y_{IJ}) = \prod_{i=1}^{I} \sum_{z=0}^{\infty} \prod_{j=1}^{J} \frac{1}{S(\zeta_{ij}(z), \nu_{ij})} \left(\frac{\zeta_{ij}(z)^{y_{ij}}}{y_{ij}!} \right)^{\nu_{ij}} \mathbb{P}(Z_i = z). \tag{2.15}$$

Note that the normalization constant S, as defined in (2.5), is not a closed form expression for $\nu \neq 1$ and a function of the latent variable value z. Accordingly, we have that the (log-)likelihood, and therefore also the maximum likelihood estimates, can not be evaluated in a closed form either.

In order to obtain parameter estimates of the probability model, various approximation methods could be employed. Firstly, a numerical approach, however, the downside is that the computational complexity of this likelihood is high, due to the summation of z from 0 to infinity and the infinite sum of the normalization constant. Secondly, a pseudo-likelihood approach in which a close approximation of the true likelihood is used that allows for simpler or even closed forms. This pseudo-likelihood is than optimized as an approximation for the maximum likelihood estimates. A more extensive overview of estimation methods will be provided in the next chapter.

When the GP distribution is employed to model under-dispersion, we will obtain the following likelihood

$$L(\theta; y_{11}, \ldots, y_{IJ}) = \prod_{i=1}^{I} \int_{z=0}^{\infty} \prod_{j=1}^{J} \frac{\mu_{ij}(z)(\mu_{ij}(z) + \varphi_{ij} y_{ij})^{(y_{ij}-1)}}{(1 + \varphi_{ij})^{y_{ij}} y_{ij}!}$$
$$\cdot \exp\{-\frac{\mu_{ij}(z) + \varphi_{ij} y_{ij}}{1 + \varphi_{ij}}\} \mathbb{1}\{y_{ij} \leq m\} f_{Z_i}(z), \tag{2.16}$$

with $\theta = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$. Note that this likelihood has a data dependent support as a result of the truncation. In particular, there is a restriction on the range of means which depends on the level of underdispersion, and vice versa (Huang, 2016). Given the severe under-dispersion that was observed in the case study (with a mean variance ratio of approximately 1:70) in the repeated measurement setting, this dependence is especially unattractive in a regression setting. In this thesis, we therefore solely use the GP to simulate data, but do not use this distribution for estimation purposes.

Parameter estimation of the CMP latent variable model

In this chapter, we distinguish two major estimation techniques to determine the estimates of the probability model. The first technique is based on likelihood maximization and the second uses the method of moments. Among the likelihood based methods, a situation with a different repeated measurement distribution than the CMP is considered.

All estimation methods will be affected by the selection of the latent variable distribution. Although the latent variable is discrete by nature, continuous forms are considered aiming to obtain explicit estimates for the model parameters or computationally more efficient estimation procedures.

3.1 Maximum likelihood estimation

In this section we aim to maximize the likelihood function as described in (2.14) and find the corresponding parameters. The presence of the latent variable Z_i in the normalization constant, that itself has no closed form expression, results into a non-trivial estimation problem, where the likelihood has no closed form either and cannot be evaluated analytically. Therefore, various pseudo-likelihoods that are close approximations of the real likelihood will be examined. Among others, we have iterative numerical methods using truncations of infinite sums (integrals) and closed form solution using approximations of the normalization constant $S(\zeta, \nu)$ in (2.5), with $\zeta = \zeta_{ij}(z) = z \cdot \exp\{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}\}$ and $\nu = \nu_{ij} = \exp\{\boldsymbol{x}_{ij}^T \boldsymbol{\gamma}\}$. Note that these approximations can also be an algorithmic solution, where the normalization constant is not equally defined for all z. We denote the pseudo-log-likelihood by

$$l^* = l^*(\boldsymbol{\theta}; y_{11}, \ldots, y_{IJ}) = \sum_{i=1}^{I} \log \left(\int_{z=0}^{\infty} \prod_{j=1}^{J} \frac{1}{\widetilde{S}_{ij}^z(\zeta_{ij}(z), \nu_{ij})} \left(\frac{\zeta_{ij}(z)^{y_{ij}}}{y_{ij}!} \right)^{\nu_{ij}} f_{Z_i}(z) dz \right), \quad (3.1)$$

$$\zeta_{ij}(z) = z \cdot \exp\{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}\} \qquad \text{and} \qquad \nu_{ij} = \exp\{\boldsymbol{x}_{ij}^T \boldsymbol{\gamma}\},$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$ and $\widetilde{S}_{ij}^z(\zeta_{ij}(z), \nu_{ij})$ is some to be defined approximation of the normalization constant $S(\zeta_{ij}(z), \nu_{ij})$. Here, $\boldsymbol{\beta}$ is the vector with tool and operator effects on the location, $\boldsymbol{\gamma}$ the vector with tool effects on the dispersion and $\boldsymbol{\delta}$ the parameters of the distribution of the latent variable Z_i. Note that the tool and operator effects are considered fixed.

The first method is a numerical solution, where we use a finite instead of an infinite upper bound of the summation (or integral) of z and truncate the infinite series of the normalization constant at the kth term as defined in (2.6). For both truncations, the term at which we truncate is based on some relative contribution criteria, such that the relative improvement of adding an extra term is lower than a certain threshold. Note that for this method only discrete latent variable distributions are covered.

In the second approach, we aim to find closed form expressions by combining an order n approximation of the normalization constant with selecting a latent variable distribution, such that the latent variable integrates out of the likelihood. In case we do not find a closed form solution when using a discrete latent variable distribution, we will additionally use a finite upper bound for the summation of the latent variable in the likelihood, based on some relative contribution criteria, to maximize the likelihood. For the continuous case, this approach involving the relative contribution criteria will not be employed.

Method 1 - Truncated sums for the latent variable and normalization constant

The first method is an algorithmic procedure, involving a numerical approximation of the pseudo-log-likelihood (3.1) and a discrete latent variable distribution, that aims to find the set of parameter θ that maximizes the pseudo-log-likelihood

$$
\begin{aligned}
l^{*(k_t)}(\boldsymbol{\theta}_t; y_{11}, \ldots, y_{IJ}) &= \sum_{i=1}^{I} \log \left(\sum_{z=0}^{k_t} \prod_{j=1}^{J} \frac{1}{\widetilde{S}^{(m_t(z))}(\zeta_{ij}(z), \nu_{ij})} \left(\frac{\zeta_{ij}(z)^{y_{ij}}}{y_{ij}!} \right)^{\nu_{ij}} \mathbb{P}(Z_i = z) \right) \\
&= A_{\theta_t} + \sum_{i=1}^{I} \log \left(\sum_{z=0}^{k_t} \prod_{j=1}^{J} \frac{1}{\widetilde{S}^{(m_t(z))}(\zeta_{ij}(z), \nu_{ij})} \left(z^{y_{ij}\nu_{ij}} \right) \mathbb{P}(Z_i = z) \right),
\end{aligned}
\tag{3.2}
$$

where $\nu_{ij} = \exp\{\gamma_j\}$ and

$$
A_{\theta_t} = \sum_{i=1}^{I} \sum_{j=1}^{J} -\nu_{ij} \log(y_{ij}!) + \boldsymbol{x}_{ij}^T \boldsymbol{\beta} y_{ij} \nu_{ij}.
\tag{3.3}
$$

Provided that some starting values $\boldsymbol{\theta}_0$ are selected, some unconstrained optimization algorithm (e.g. the quasi-Newton based *L-BFGS-B* method) could be employed to maximize this pseudo-log-likelihood. Here, in each iteration t, the value for the likelihood $l^{*(k_t)}(\boldsymbol{\theta}_t; y_{11}, \ldots, y_{IJ})$ is determined for this new set of parameters θ_t. This procedure keeps increasing t until convergences has taken place.

For each single iteration t, the value for $l^{*(k_t)}(\boldsymbol{\theta}_t; y_{11}, \ldots, y_{IJ})$ is determined, by truncating the infinite series of the latent variable at a constant k_t, such that the relative contribution of the non-constant term $R^{(a)}$ is smaller than a minimal relative contribution threshold ϵ_1. This threshold should be selected with care, taking into account the magnitude of the value for $R^{(a)}$ and the required precision for the optimization algorithm to function properly. The non-constant term $R^{(a)}$ is defined by

$$
R^{(a)} = \sum_{i=1}^{I} \log \left(\sum_{z=0}^{a} \prod_{j=1}^{J} \frac{1}{\widetilde{S}^{(m_t(z))}(\zeta_{ij}(z), \nu_{ij})} \left(z^{y_{ij}\nu_{ij}} \right) \mathbb{P}(Z_i = z) \right).
\tag{3.4}
$$

Now, the upper bound k_t of the summation of the latent variable value z, which is taken equal for all parts i, is determined by

$$
k_t := \arg \min_{a > \bar{Y}_{..}} \left(\frac{R^{(a)} - R^{(a-1)}}{R^{(a-1)}} < \epsilon_1 \right),
\tag{3.5}
$$

where $\bar{Y}_{..} = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij}$ is the average of the data. Note that the summation of z is performed in an increasing order and the relative contribution term is only checked for values of z larger than $\bar{Y}_{..}$. When this contribution criterion is met, the summation is stopped. The purpose of the constraint $a > \bar{Y}_{..}$, that requires the truncation term of the summation of the latent variable z to be larger than the average, is twofold. Firstly, it is to reduce the number of checks in order to save computation time and secondly to prevent the algorithm from stopping too soon when parameters are selected poorly. Moreover, observe that the truncation term k_t corresponds to iteration t and is therefore allowed to differ from any other truncation term in iteration s, with $t \neq s$.

Within each iteration t, the approximated normalization constant $\widetilde{S}^{(m_t(z))}(\zeta_{ij}(z), \nu_{ij})$ is determined for each value of the latent variable z, with $z = 0, \dots, k_t$. This truncated normalization constant, that strongly relies on the concept described in equation (2.6), is described by

$$\widetilde{S}^{(m_t(z))}(\zeta_{ij}(z), \nu_{ij}) = \sum_{s=0}^{m_t(z)} \left(\frac{\zeta_{ij}(z)^s}{s!} \right)^{\nu_{ij}}. \tag{3.6}$$

where $m_t(z) = \{\arg\max_j m_{t,j}(z)\}$ and $m_{t,j}(z)$ is defined by

$$m_{t,j}(z) := \arg \min_{a > 1, a \in \mathbb{Z}} \left(\frac{\widetilde{S}^{(a)}(\zeta_{ij}(z), \nu_{ij}) - \widetilde{S}^{(a-1)}(\zeta_{ij}(z), \nu_{ij})}{\widetilde{S}^{(a-1)}(\zeta_{ij}(z), \nu_{ij})} < \epsilon_2 \right). \tag{3.7}$$

Here, $\zeta_{ij}(z)$ and ν_{ij} are described by $\boldsymbol{\theta}$ and $m_t(z)$ is an integer value. Note that we write $m_{t,j}(z)$, since the upper bound of the finite sum of the approximated normalization constant is dependent on the value of the latent variable $z = 0, \dots, k_t$ and the effect of the repeats $j = 1, \dots, J$. However, instead of selecting an upper bound for each repeat j, we take a single upper bound $m_t(z)$ for all repeats $j = 1, \dots, J$ to simplify the numerical computation. More specific, $m_t(z)$ is the smallest argument such that the contribution criterion ϵ_2 is met for all repeats $j = 1, \dots, J$.

Recall from Section 2.3, that for equi-dispersion ($\nu = 1$) the normalization constant is convergent and equal to $\exp\{\zeta_{ij}(z)\}$.

Proposition 3.1.1. *Let $\nu \geq 1$ and $z < \infty$. Then the normalization constant $\sum_{s=0}^{\infty} \left(\frac{z^s}{s!} \right)^n$, which is an infinite series, is a bounded, monotonic increasing series and therefore convergent.*

A full proof of Proposition 3.1.1, that uses Stirling's formula to define an upper bound, can be found in Appendix A.1. Due to convergence of this normalization constant for $z < \infty$ and equi-dispersion or under-dispersion ($\nu \geq 1$), the upper bound of the summation $m_t(z)$ is selected based on a minimal relative contribution to the normalization constant ϵ_2.

Summarizing, for each iteration t of the pseudo-log-likelihood maximization, the truncation term k_t for the latent variable is determined. Within this iteration we determine $m_t z$ for every single $z = 0, \dots, k_t$. Due to this double loop, this method becomes very computationally intensive.

Method 2 - Order n approximation of the normalization constant

In this method, the normalization constant is approximated by using the first n terms of the asymptotic expansion of the normalization constant as described in (2.8). This order n approximation is defined by

$$\widetilde{S}^{(n)}(\zeta_{ij}(z), \nu_{ij}) = \frac{\zeta_{ij}(z)^{(1-\nu_{ij})/2} \cdot e^{\nu_{ij}\zeta_{ij}(z)}}{(2\pi)^{(\nu_{ij}-1)/2} \sqrt{\nu_{ij}}} \sum_{k=0}^{n} c_k (\nu_{ij}\zeta_{ij}(z))^{-k}, \tag{3.8}$$

where the constants c_j are uniquely determined by the expansion in (2.9) and independent of $\zeta_{ij}(z)$. Now the general form of the pseudo-log-likelihood with latent variable Z_i is defined in Lemma 3.1.1.

Lemma 3.1.1. *Assume an order n expansion of the normalization constant* (3.8) *and take any latent variable Z_i, then the corresponding pseudo-log-likelihood* (3.1) *changes to*

$$l^* = A + \sum_{i=1}^{I} \log \left(\int_{z=0}^{\infty} z^{B_i} e^{-zC_i} \cdot \frac{1}{\prod_{j=1}^{J} \left(\sum_{k=0}^{n} c_k (\nu_{ij}\zeta_{ij}(z))^{-k} \right)} f_{Z_i}(z) dz \right), \qquad (3.9)$$

where

$$A = \sum_{i=1}^{I} \sum_{j=1}^{J} -\nu_{ij} \log(y_{ij}!) + \boldsymbol{x}_{ij}^T \boldsymbol{\beta} y_{ij} \nu_{ij} + \frac{\nu_{ij}-1}{2} \log(2\pi) + \frac{1}{2}\log(\nu_{ij}) + \boldsymbol{x}_{ij}^T \boldsymbol{\beta} \frac{(\nu_{ij}-1)}{2},$$

$$\tag{3.10}$$

$$B_i = \sum_{j=1}^{J} y_{ij}\nu_{ij} + \frac{\nu_{ij}-1}{2} \qquad and \qquad C_i = \sum_{j=1}^{J} \nu_{ij} e^{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}}.$$

A proof of this lemma can be found in Lemma A.1 in Appendix A.1. We can find a closed form expression for l^*, when for a certain combination of the latent variable Z_i and order of the normalization constant, the pseudo-log-likelihood does not involve z. Note that for $n = 1$, we have

$$\prod_{j=1}^{J} \left(\sum_{k=0}^{n} c_k (\nu_{ij}\zeta_{ij}(z))^{-k} \right) = \prod_{j=1}^{J} \left(1 + c_1 \cdot \left(\nu_{ij} z e^{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}} \right)^{-1} \right) = \prod_{j=1}^{J} \frac{z\nu_{ij} + c_1 e^{-\boldsymbol{x}_{ij}^T \boldsymbol{\beta}}}{z\nu_{ij}}. \qquad (3.11)$$

From this it can be concluded that for higher order asymptotic expansions ($n > 0$) and more than one repeat $J > 1$, the number of terms in the pseudo-log-likelihood will quickly grow to unworkable sizes when increasing the number of repeats J and/or order of the approximation of the normalization constant n.

In addition, increasing the order of the asymptotic expansion to infinity will not necessarily lead to a more precise approximation of the true normalization constant and therefore not necessarily to a more precise likelihood. Nevertheless, for an asymptotic expansion of the normalization constant with fixed order n, we obtain convergence as stated in Proposition 3.1.2. A proof of this Proposition can be found in Appendix A.1. Moreover, the effect of the order of the asymptotic expansion, in case of under-dispersion, on the performance of the approximation of the normalization constant is demonstrated in Chapter 6.1.

Proposition 3.1.2. *Let us take an order n approximation of the normalization constant, with $0 \leq n \leq \infty$. Then for finite $\nu_{ij} \geq 1$ the pseudo-log-likelihood* (3.9) *is absolute convergent.*

For a first order approximation of the normalization constant ($n = 0$), the product in (3.11) becomes 1. As a result, the pseudo-log-likelihood (3.9) from Lemma 3.1.1 converts to

$$l^* = A + \sum_{i=1}^{I} \log \left(\int_{z=0}^{\infty} z^{B_i} e^{-zC_i} f_{Z_i}(z) dz \right). \qquad (3.12)$$

In the following two paragraphs, we will consider both discrete and continuous prior distributions, in order to obtain a closed form expression for the pseudo-log-likelihood l^*.

(a) Discrete latent variable

When using a first order approximation of the normalization, the pseudo-log-likelihood (3.9) from Lemma 3.1.1 converts to

$$l^* := A + \sum_{i=1}^{I} \log \left(\sum_{z=0}^{\infty} z^{B_i} e^{-zC_i} \mathbb{P}(Z_i = z) \right), \qquad (3.13)$$

where A, B_i, C_i are defined in (3.10).

Among the possible discrete distributions for latent variable Z_i we considered the Poisson, negative binomial and geometric distribution, but did not find a form such that for all possible ν_{ij} the pseudo-log-likelihood (3.13) does not involve z. Since, the complexity of the pseudo-log-likelihood increases, the same was found for higher order approximations of the normalization constant. Note that for the special case where the CMP distribution has $\nu_{ij} = 1$, we actually describe a Poisson distribution. This implies that the for each order n expansion, the normalization constant is defined by $\exp\{\zeta_{ij}(z)\}$. Therefore, the pseudo-log-likelihood equals the true log-likelihood. However, the selection of $\nu_{ij} = 1$, still does not provide a closed form.

In order to solve the pseudo-log-likelihood (3.9) with any order $n > 0$ of the normalization constant and discrete latent variable, we could employ a numerical approach that uses the result of Proposition 3.1.2. The suggested numerical approximation of the likelihood

$$l^* = A + \sum_{i=1}^{I} \log \left(\sum_{z=0}^{\infty} z^{B_i} e^{-zC_i} \cdot \frac{1}{\prod_{j=1}^{J} \left(\sum_{k=0}^{n} c_k (\nu_{ij} \zeta_{ij}(z))^{-k} \right)} \mathbb{P}(Z_i = z) \right), \tag{3.14}$$

with A, B_i and C_i defined in (3.10) involves a truncation of the infinite summation of the latent variable distribution at the term k_t, described by

$$k_t := \arg\min_{a > \bar{Y}_{..}} \left(\left(R^{(a)} - R^{(a-1)} \right) / R^{(a-1)} < \epsilon_1 \right). \tag{3.15}$$

Here, $\bar{Y}_{..} = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij}$ is the empirical average of the data and

$$R^{(a)} = \sum_{i=1}^{I} \log \left(\sum_{z=0}^{a} z^{B_i} e^{-zC_i} \frac{1}{\prod_{j=1}^{J} \left(\sum_{k=0}^{n} c_k (\nu_{ij} \zeta_{ij}(z))^{-k} \right)} \mathbb{P}(Z_i = z) \right). \tag{3.16}$$

In this way, the relative contribution to the non-constant part $R^{(a)}$ of the pseudo-likelihood is controlled.

(b) Continuous latent variable distribution

Aiming to simplify the likelihood, various continuous distributions were considered for the latent variable Z_i. In particular, when selecting the Gamma as the prior distributions, a closed form was obtained, see Proposition 3.1.3. A proof of this derivation is provided in Appendix A.1.

Proposition 3.1.3 (Gamma prior)**.** *Let us assume a $Gamma(\kappa, \lambda)$ distribution for the latent variable Z_i. Now, given first order approximation of the normalization constant, we have the pseudo-log-likelihood (3.9) from Lemma 3.1.1 obtains the form*

$$\begin{aligned} l^* &= A + \sum_{i=1}^{I} \log \left(\frac{\lambda^{B_i} \Gamma(B_i + \kappa)}{(1 + C_i \lambda)^{B_i + \kappa} \Gamma(\kappa)} \right) \\ &= A + \sum_{i=1}^{I} B_i \log(\lambda) - (B_i + \kappa) \log(1 + C_i \lambda) + \log(\Gamma(B_i + \kappa)) - \log(\Gamma(\kappa)). \end{aligned} \tag{3.17}$$

with A, B_i and C_i defined in (A.2). Here, κ and λ are the shape and scale parameter of the Gamma distribution.

Method 3 - Alternative repeated measurements model

Instead of the CMP latent variable model, we now assume that $Y_{ij}|Z_i$ is quasi-Poisson distributed, with normally distributed random effects. Then we have

$$\begin{aligned} Y_{ij}|Z_i &\sim Pois(\mu_{ij}, \delta) \\ \mu_{ij} &= \exp(z_i + \boldsymbol{x}_{ij}^T \boldsymbol{\beta}) \\ z_i &\sim \mathcal{N}(\alpha_0, \sigma^2), \end{aligned} \tag{3.18}$$

with $\boldsymbol{\beta}$ the vector of tool and operator effects and $\boldsymbol{x}_{ij} \in \mathbb{R}^{p \times 1}$ a vector of the design matrix containing indicator values, representing the setting under which that single measurement is performed. Now, the moments of the conditional distribution are defined by

$$\mathbb{E}\left[Y_{ij}|Z_i\right] = \mu_{ij} \qquad \text{and} \qquad \text{Var}\left(Y_{ij}|Z_i\right) = \text{Var}\left(\mu_{ij}\right)\delta \tag{3.19}$$

Note that in this approach, the latent random variable Z_i has a continuous distribution, which slightly changes the interpretation. Instead of the actual count, which is discrete, it can now be explained as to the expected count under a specific tool and operator setting. Now for $\delta = 1$, the log-likelihood $l = l(\boldsymbol{\theta}; y_{11}, \ldots, y_{IJ})$ of this generalized linear mixed model (GLMM) can be written as

$$
\begin{aligned}
l &= \log\left(\prod_{i=1}^{I}\int_{z=-\infty}^{\infty}\prod_{j=1}^{J}\frac{\mu_{ij}^{y_{ij}}\exp\{-\mu_{ij}\}}{y_{ij}!}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{1}{2}\left(\frac{z_i}{\sigma}\right)^2\}dz_i\right)\\
&= A_3 + \sum_{i=1}^{I}\log\left(\int_{z_i=-\infty}^{\infty}\exp\{\sum_{j=1}^{J}\left(z_iy_{ij}+e^{z_i+\boldsymbol{x}_{ij}\boldsymbol{\beta}}\right)\}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{1}{2}\left(\frac{z_i-\alpha_0}{\sigma}\right)^2\}dz_i\right),
\end{aligned}
\tag{3.20}
$$

where

$$A_3 = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(-\log\left(y_{ij}!\right) + \left(\boldsymbol{x}_{ij}\boldsymbol{\beta}\right)\right). \tag{3.21}$$

Observe that this is the log-likelihood of a GLMM with a regular Poisson distribution. The log-likelihood of the model including the parameter δ, is defined by $l_\delta = l/\delta$. Here, δ is determined by the value of the residual deviance divided by the residual degrees of freedom. The residual deviance equals -2 times the likelihood-ratio. When considering a main effects model with only tool and operator effects, one of both effects is included in the intercept α_0, obtaining a residual degrees of freedom of $IJ - (J - 2 + p)$, where p is the number of parameters of the latent variable distribution.

Note that the parameter estimates $\boldsymbol{\beta}$, α_0 and σ are independent of the dispersion parameter δ. Whereas the point estimates are equal for the Poisson and quasi-Poisson GLMM, the standard errors of the estimates do differ for the quasi-Poisson GLMM. The standard errors for the quasi-Poisson GLMM can be obtained by multiplying the standard errors of the Poisson GLMM with a factor $\sqrt{\delta}$. Therefore, a generalized linear mixed model with a Poisson distribution can be fitted instead of quasi-Poisson distribution, after which the likelihood value and the standard errors should be scaled. Fitting a Poisson GLMM is straightforward as it is implemented in the *glmer* function of the *'lme4'* package of the statistical software [R] (Bates et al., 2015), whereas the quasi-Poisson is not. Here, the Gauss-Hermite quadrature is used to approximate the integral that is present in the log-likelihood.

Statistical inference

In the previous paragraphs of this chapter we defined various types of (pseudo-)log-likelihoods for the probability model (2.14). By maximizing these (pseudo-)log-likelihoods, we obtain their corresponding maximum (pseudo-)likelihood estimates. These are again approximations for the probability model's true ML estimates. Although it remains to be seen whether these (pseudo-)likelihood estimates are asymptotically normal, we assume this to approximate the confidence interval of the (pseudo-)ML estimates.

In the case that asymptotic normality of the ML estimates would hold, the confidence intervals of these estimates follow from the Fisher information matrix $I(\theta)$, which is the variance of score function. Here, the score function is the derivative of the log-likelihood. For the maximum likelihood estimates, $\hat{\theta}$, the score function $s(\hat{\theta}) = 0$. Then the confidence intervals of the asymptotically normal ML estimates can be derived using a first-order Taylor expansion of the score $S(\theta)$ around

its estimator $\hat{\theta}$, obtaining that $\hat{\theta} - \theta$ is approximately $N(0, I^{-1}(\theta))$. Then, the standard errors of the estimates are the square roots of the diagonal of inverse Fisher Information.

Only for a Gamma prior distribution we obtained a closed form expression of the estimates. In this situation, the ML estimates can be obtained by setting the score function to zero and deriving the Fisher information. However, for all other (pseudo-)likelihoods some optimization procedure needs to be employed to obtain the maximum likelihood estimates. To determine the confidence intervals, we use the observed Fisher Information matrix $\hat{I}(\theta) = -\frac{\partial^2 l(\theta)}{(\partial \theta)^2}$, which is the Hessian of the negative log-likelihood, as an approximation of the true Fisher information. The Hessian is some approximation of the variation around an estimator. The Hessian matrix is obtained by approximating the second partial derivatives evaluated at the maximum likelihood estimates using finite differencing and multiplying with -1. Now, the inverse of the Hessian is the estimated variance-covariance matrix, which can be used to establish the confidence intervals of the point estimates (Dobson & Barnett, 2008).

3.2 Method of moments

Probability model and its moments

For the moment based estimation, a different notation of the probability model (2.12) is preferred. Contrary to the old notation, where a repeat was any measurement on the same part (so all possible combinations of tool and operator), we now denote a repeat as number of repetitions on a part with the same tool and operator. The model is now defined as follows

$$
\begin{aligned}
P(Y_{ijkl} = y) &= \int_{z=0}^{\infty} P(Y_{ijkl} = y | Z_i = z) dF_{Z_i}(z) dz \\
&= \int_{z=0}^{\infty} \text{CMP}(\zeta_{ijk}(z), \nu_{ijk}) dF_{Z_i}(z) dz,
\end{aligned}
\tag{3.22}
$$

where Y_{ijkl} denotes the observed count on part $i = 1, \ldots, I$, tool $j = 1, \ldots, J$, operator $k = 1, \ldots, K$ and repeat $l = 1, \ldots, L$. Here, the true counts of the parts i are again assumed to be independently and identically distributed with dF_{Z_i}. For the location parameter of the CMP we again have a multiplicative form, and for the dispersion we either take a constant or random intercept. This gives

$$
\zeta_{ijk}(z) = z \cdot a_j \cdot b_k \qquad \text{and} \qquad \nu_{ijk} = \exp\{\gamma_0 + \gamma_j\}.
\tag{3.23}
$$

where $a_j = \exp\{\alpha_j\}$ and $b_k = \exp\{\beta_k\}$. Moreover, we denote $\nu_0 = \exp\{\gamma_0\}$ and $\nu_j = \exp\{\gamma_0 + \gamma_j\}$,

Let us introduce constraints for the fixed effects $\frac{1}{J} \sum_{j=1}^{J} a_j = 1$ and $\frac{1}{K} \sum_{k=1}^{K} b_k = 1$ such that the average effect of the tools and operators is 1. Without these constraints on the fixed effects, the model becomes unidentifiable. Note that for the selected multiplicative structure, no interaction term $(ab)_{jk} = \exp\{\alpha\beta\}_{jk}$ with a corresponding constraint $\frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} (ab)_{jk} = 1$ is included in $\zeta_{ijk}(z)$. The main reason is that by including this term it does *not* necessarily hold that $\frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} a_j b_k (ab)_{jk} = 1$ when $(ab)_{jk} \neq 1, \forall j, k$. This was preferred to enforce that z represents the average count over all tools.

Based on the full model described in (3.23), two scenarios that will be considered are described in Table 3.1.

Table 3.1: Model scenarios for the method of moments

Scenario	Location parameter $\zeta_{ijk}(z)$	Dispersion parameter ν_{ijk}
1	Main effects tool and operator	Constant dispersion ($\nu_{ijk} = \nu_0, \ \forall i, j, k$)
2	Main effects tool and operator	Dispersion per tool ($\nu_{ijk} = \nu_j, \ \forall i, j, k$)

Approximation conditional expectation and variance

Analogue to the maximum likelihood estimation of the CMP parameters, we also apply different approaches to determine the moment estimators. Since the moments of the CMP contain the normalization constant, which cannot be evaluated in a closed form, we again use the order n asymptotic expansion and the summation of a truncated series. Now, approximated moments of the conditional distribution look as follows

$$
\begin{aligned}
\mathbb{E}[Y_{ijkl} | Z_i] = \mu_{ijk} &\approx \frac{1}{\nu_{ijk}} \frac{\partial \log \widetilde{S}(\zeta_{ijk}(Z_i), \nu_{ijk})}{\partial \log \zeta_{ijk}} \\
\text{Var}(Y_{ijkl} | Z_i) &\approx \frac{1}{\nu_{ijk}^2} \frac{\partial^2 \log \widetilde{S}(\zeta_{ijk}(Z_i), \nu_{ijk})}{\partial \log^2 \zeta_{ijk}},
\end{aligned}
\tag{3.24}
$$

where $\widetilde{S}(\zeta_{ijk}(Z_i), \nu_{ijk})$ is the approximation of the true normalization constant $S(\zeta_{ijk}(Z_i), \nu_{ijk})$.

Mean squares

According to the CMP latent variable model (3.22) and the defined scenarios in Table 3.1, variation can be attributed to the measurement error, parts, tools or operators. For a balanced experiment, we distinguish the mean square error (MS_E), parts mean squares (MS_P), tool mean squares (MS_T) and operator mean squares (MS_O). These mean squares are characterized by

$$
\begin{aligned}
MS_E &= \frac{SS_{Error}}{IJK(L-1)} = \frac{1}{IJK(L-1)} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} (Y_{ijkl} - \bar{Y}_{ijk\cdot})^2, \\
MS_P &= \frac{SS_{Parts}}{I-1} = \frac{JKL}{I-1} \sum_{j=1}^{J} (\bar{Y}_{\cdot j\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot\cdot})^2, \\
MS_T &= \frac{SS_{Tool}}{J-1} = \frac{IKL}{J-1} \sum_{i=1}^{I} (\bar{Y}_{i\cdot\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot\cdot})^2, \\
MS_O &= \frac{SS_{Operator}}{K-1} = \frac{IJL}{K-1} \sum_{k=1}^{K} (\bar{Y}_{\cdot\cdot k\cdot} - \bar{Y}_{\cdot\cdot\cdot\cdot})^2,
\end{aligned}
\tag{3.25}
$$

where

$$
\bar{Y}_{\cdot\cdot\cdot\cdot} = \frac{1}{IJKL} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} Y_{ijkl},
$$

$$
\bar{Y}_{i\cdot\cdot\cdot} = \frac{1}{JKL} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} Y_{ijkl}, \qquad \bar{Y}_{\cdot j\cdot\cdot} = \frac{1}{IKL} \sum_{i=1}^{I} \sum_{k=1}^{K} \sum_{l=1}^{L} Y_{ijkl},
\tag{3.26}
$$

$$
\bar{Y}_{\cdot\cdot k\cdot} = \frac{1}{IJL} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{l=1}^{L} Y_{ijkl}, \text{ and} \qquad \bar{Y}_{ijk\cdot} = \frac{1}{IJK} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{ijkl}.
$$

Method 1 - Order n approximation of the normalization constant

In this paragraph, we aim to derive the moment estimates of the parameters for several model scenarios and the latent variable distributions. Since the order of the asymptotic expansion approximation of the normalization constant (2.8) strongly affects the complexity of the moment expressions, we first start with a first order approximation.

First order approximation (n=0)

Let us assume a first order approximation of the normalization constant (n=0). Then by the Law of Iterated Expectations we have that the expected value of Y_{ijkl} is described by

$$
\mathbb{E}[Y_{ijkl}] = \mathbb{E}\left[\mathbb{E}[Y_{ijkl}|Z_i]\right] \approx \mathbb{E}\left[\zeta_{ijk}(Z_i) - \frac{\nu_{ijk}-1}{2\nu_{ijk}}\right] = \mathbb{E}[Z_i]a_j b_k - \frac{\nu_{ijk}-1}{2\nu_{ijk}}.
\tag{3.27}
$$

Moreover, the expectation of the conditional variance, which is part of the expected mean squares is

$$
\mathbb{E}\left[\mathrm{Var}\left(Y_{ijkl}\right)\right] \approx \mathbb{E}\left[\frac{\zeta_{ijk}(Z_i)}{\nu_{ijk}}\right] = \frac{\mathbb{E}[Z_i]a_j b_k}{\nu_{ijk}}
\tag{3.28}
$$

We consider scenario 2, which is the most general scenario and contains scenario 1 as a special case ($\nu_j = \nu_0, \forall j$). Recall that this scenario has main effects for the location and a different dispersion per tool. By applying the Law of Iterated Expectations and the constraints for the

fixed effects for tools and operators defined in 3.2 to the expectations of the averages $\bar{Y}_{....}$, $\bar{Y}_{i...}$, $\bar{Y}_{.j..}$ and $\bar{Y}_{..k.}$ from (3.26) we obtain

$$
\mathbb{E}[\bar{Y}_{....}] = \mathbb{E}[\bar{Y}_{i...}] \approx \mathbb{E}[Z_i] \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} a_j b_k - \frac{\nu_{ijk} - 1}{2\nu_{ijk}} = \mathbb{E}[Z_i] - \frac{1}{J} \sum_{j=1}^{J} \frac{\nu_j - 1}{2\nu_j},
$$

$$
\mathbb{E}[\bar{Y}_{.j..}] \approx \mathbb{E}[Z_i] a_j \frac{1}{K} \sum_{k=1}^{K} b_k - \frac{\nu_{ijk} - 1}{2\nu_{ijk}} = \mathbb{E}[Z_i] a_j - \frac{\nu_j - 1}{2\nu_j}, \tag{3.29}
$$

$$
\mathbb{E}[\bar{Y}_{..k.}] \approx \mathbb{E}[Z_i] b_k \frac{1}{J} \sum_{j=1}^{J} a_j - \frac{\nu_{ijk} - 1}{2\nu_{ijk}} = \mathbb{E}[Z_i] b_k - \frac{1}{J} \sum_{j=1}^{J} \frac{\nu_j - 1}{2\nu_j}.
$$

To determine the expected mean squares, we apply the Law of Total Variance, which is defined in Appendix A.2 in Lemma A.2.1. From the mean squares in (3.25) we can derive the expected mean squares

$$
\mathbb{E}[MS_E] = \mathbb{E}[Z_i] \frac{1}{J} \sum_{j=1}^{J} \frac{a_j}{\nu_j}
$$

$$
\mathbb{E}[MS_P] = \mathbb{E}[MS_E] + JKL \cdot \mathrm{Var}(Z_i)
$$

$$
\mathbb{E}[MS_T] = \mathbb{E}[MS_E] + \frac{IKL}{J-1} \sum_{j=1}^{J} (a_j^2 - 1) \left(\frac{\mathrm{Var}(Z_i)}{I} \right)
$$

$$
\qquad\qquad + \frac{IKL}{J-1} \sum_{j=1}^{J} \left(\mathbb{E}[Z_i](a_j - 1) - \frac{\nu_j - 1}{2\nu_j} + \frac{1}{J} \sum_{j=1}^{J} \frac{\nu_j - 1}{2\nu_j} \right)^2 \tag{3.30}
$$

$$
\mathbb{E}[MS_O] = \mathbb{E}[MS_E] + \frac{IJL}{K-1} \sum_{k=1}^{K} (b_k^2 - 1) \left(\frac{\mathrm{Var}(Z_i)}{I} + \mathbb{E}[Z_i]^2 \right).
$$

A full derivation of the expected mean squares can be found in Appendix A.2. The expected first moments and mean squares for scenario 1 are presented in Proposition 3.2.1.

Remark 3.2.1. *Let us take scenario 2 with a constant dispersion $\nu_{ijk} = \nu_j = \nu_0$, such that we obtain the special case scenario 1. Now, the expected first moments in (3.29) reduce to*

$$
\mathbb{E}[\bar{Y}_{....}] = \mathbb{E}[\bar{Y}_{i...}] \approx \mathbb{E}[Z_i] - \frac{\nu_0 - 1}{2\nu_0},
$$

$$
\mathbb{E}[\bar{Y}_{.j..}] \approx \mathbb{E}[Z_i] a_j - \frac{\nu_0 - 1}{2\nu_0}, \tag{3.31}
$$

$$
\mathbb{E}[\bar{Y}_{..k.}] \approx \mathbb{E}[Z_i] b_k - \frac{\nu_0 - 1}{2\nu_0}.
$$

and the expected mean squares in (3.30) simplify to

$$
\mathbb{E}[MS_E] = \frac{\mathbb{E}[Z_i]}{\nu_0}
$$

$$
\mathbb{E}[MS_P] = \mathbb{E}[MS_E] + JKL \cdot Var(Z_i)
$$

$$
\mathbb{E}[MS_T] = \mathbb{E}[MS_E] + \frac{IKL}{J-1} \sum_{j=1}^{J} (a_j^2 - 1) \left(\frac{Var(Z_i)}{I} + \mathbb{E}[Z_i]^2 \right) \tag{3.32}
$$

$$
\mathbb{E}[MS_O] = \mathbb{E}[MS_E] + \frac{IJL}{K-1} \sum_{k=1}^{K} (b_k^2 - 1) \left(\frac{Var(Z_i)}{I} + \mathbb{E}[Z_i]^2 \right).
$$

Remark 3.2.2. *Let us assume the CMP latent variable model with model scenario 2 from Paragraph 3.2 with a first order approximation of the normalization constant and $a_j = b_k = 1$, $\forall j, k$. Then, in case $\nu_j = \nu_0$, $\forall j$ we have that $\mathbb{E}[MS_E] = \mathbb{E}[MS_T]$ and $\mathbb{E}[MS_E] = \mathbb{E}[MS_O]$.*

To obtain the moment estimators for the different scenarios of the probability model, we replace the expected moments $\mathbb{E}[\bar{Y}_{\ldots}]$, $\mathbb{E}[\bar{Y}_{.j.}]$, $\mathbb{E}[\bar{Y}_{..k.}]$ and the mean squares $\mathbb{E}[MS_E]$, $\mathbb{E}[MS_T]$, $\mathbb{E}[MS_O]$, $\mathbb{E}[MS_P]$ with their sample versions $\bar{Y}_{\ldots}$, $\bar{Y}_{.j.}$, $\bar{Y}_{..k.}$, MS_E, MS_T, MS_O and MS_P, respectively. However, first a latent variable distribution needs to be assumed. In the Proposition 3.2.1 we take a negative binomial (θ, k) prior for Z_i, which reduces to a $Pois(\theta)$ when the dispersion parameter of the negative binomial $k = 0$. The full derivation of this proposition can be found in Appendix A.2.

Proposition 3.2.1 (Moment estimates scenario 1)**.** *Let us assume model scenario 1 with a first order approximation of the normalization constant. Moreover, let the latent variable Z_i follow a negative binomial distribution, $NBinom(\theta, \kappa)$. Replacing the theoretical moments in Remark 3.2.1 with their sample versions, we obtain the moment estimators*

$$
\hat{\nu}_0 = (1 + 2\bar{Y}_{\ldots} - 2\hat{\theta})^{-1}, \qquad \hat{a}_j = \frac{\left(\bar{Y}_{.j.} - \bar{Y}_{\ldots} + \hat{\theta}\right)}{\hat{\theta}},
$$
$$
\hat{b}_k = \frac{\left(\bar{Y}_{..k.} - \bar{Y}_{\ldots} + \hat{\theta}\right)}{\hat{\theta}}, \qquad \hat{\kappa} = \frac{MS_P - MS_E - JKL \cdot \hat{\theta}}{JKL \cdot \hat{\theta}^2}, \tag{3.33}
$$

where

$$
\hat{\theta} = \frac{2\bar{Y}_{\ldots} + 1}{4} + \sqrt{\frac{-MS_E}{2} + \frac{(\bar{Y}_{\ldots} + 1)^2}{16}}. \tag{3.34}
$$

Note that the tool and operator mean squares are not used, since model scenario 1 has a constant dispersion. Although we formulate this paragraph in the more general setting of scenarios 2, a non-constant dispersion brings in heterogeneity of the variances. Whereas the first moment are rarely influenced by this assumption, the mean squares are strongly affected. The situation becomes particularly complex, when the design becomes unbalanced. For example, a weighted least squares could be employed. Please note that this is outside the scope of this research.

Higher order expansions (n>0)

In the previous paragraph we derived the moment estimates for a first order approximation of the normalization constant. Moreover, in Paragraph 2.2, the third order approximation of the mean and variance of a CMP distribution were presented. In the notation of the probability model, we have that

$$
\mathbb{E}[Y_{ijkl}|Z_i] \approx \zeta_{ijk}(Z_i) - \frac{\nu_{ijk} - 1}{2\nu_{ijk}} - \frac{\nu_{ijk}^2 - 1}{24\nu_{ijk}^2}\zeta_{ijk}(Z_i)^{-1} - \frac{\nu_{ijk}^2 - 1}{24\nu_{ijk}^3}\zeta_{ijk}(Z_i)^{-2}
$$
$$
\mathrm{Var}[Y_{ijkl}|Z_i] \approx \frac{\zeta_{ijk}(Z_i)}{\nu_{ijk}} + \frac{\nu_{ijk}^2 - 1}{24\nu_{ijk}^3}\zeta_{ijk}^{-1}(Z_i) + \frac{\nu_{ijk}^2 - 1}{12\nu_{ijk}^4}\zeta_{ijk}(Z_i)^{-2} \tag{3.35}
$$

Note that the derivation of the moment estimators requires calculating $\mathbb{E}[Y_{ijkl}] = \mathbb{E}[\mathbb{E}[Y_{ijkl}|Z_i]]$. For discrete distributions (i.e. Poisson or negative binomial distribution) with 0 in their domain, this implies solving $\mathbb{E}[1/Z_i]$ which is not possible. A higher order approximation is therefore not considered.

Statistical inference

In this paragraph, we will discuss the construction of confidence intervals of the moment estimators and an hypothesis test of the tool and operator effects based on the mean square ratios.

Confidence intervals

To obtain the confidence intervals of the moment estimators from Proposition 3.2.1, we have to derive the standard errors. Note that the moment estimators $\hat{\theta}, \kappa, \nu_0, a_j$ and b_k are functions of $\bar{Y}_{....}, \bar{Y}_{.j..}, \bar{Y}_{..k.}, MS_E$ and MS_P. Let $T = g(X_1, \ldots, X_n)$ denote the moment estimator that is a function of random variables $X_1, \ldots, X_n$. In order to determine the variances of these moment estimators T, we consider the univariate Taylor expansions around the means of the random variables $\mu_{X_1}, \ldots, \mu_{X_n}$. We then approximate the variance of this moment estimator by

$$\mathrm{Var}(T) \approx \sum_{i=1}^{n} \left(\frac{\partial g(\mu_{X_1}, \ldots, \mu_{X_n})}{\partial u_i} \right)^2 \sigma_{X_i}^2. \tag{3.36}$$

By taking a first order approximation of the normalization constant with model scenario 1, the variances of $\bar{Y}_{....}, \bar{Y}_{.j..}, \bar{Y}_{..k.}$ that are derived in Proposition A.2.2, convert into

$$
\begin{aligned}
\sigma_{\bar{Y}_{....}}^2 &= \mathrm{Var}(\bar{Y}_{....}) = \frac{1}{I}\mathrm{Var}\left(Z_i\right) + \frac{1}{IJKL}\frac{\mathbb{E}\left[Z_i\right]}{\nu_0} \\
\sigma_{\bar{Y}_{.j..}}^2 &= \mathrm{Var}(\bar{Y}_{.j..}) = \frac{a_j^2}{I}\mathrm{Var}\left(Z_i\right) + \frac{1}{IKL}\frac{a_j\mathbb{E}\left[Z_i\right]}{\nu_0} \\
\sigma_{\bar{Y}_{..k.}}^2 &= \mathrm{Var}(\bar{Y}_{..k.}) = \frac{b_k^2}{I}\mathrm{Var}\left(Z_i\right) + \frac{1}{IJL}\frac{b_k\mathbb{E}\left[Z_i\right]}{\nu_0}.
\end{aligned}
\tag{3.37}
$$

The variances of the mean squares are more complex and will be approximated by assuming that the underlying data is i.i.d. normally distributed. Under this assumption, we have that $SS_E/\sigma_E^2 \sim \chi_{v_1}^2$ and $SS_P/\sigma_P^2 \sim \chi_{v_2}^2$, where v_1 and v_2 denote the degrees of freedom of the sum of squares error and parts, respectively. This gives

$$
\begin{aligned}
\sigma_{MS_E}^2 &= \mathrm{Var}(MS_E) = \frac{\sigma_E^4}{(IJK(L-1))^2}\mathrm{Var}\left(\frac{SS_E}{\sigma_E^2}\right) = \frac{2\sigma_E^4}{IJK(L-1)} \\
\sigma_{MS_P}^2 &= \mathrm{Var}(MS_P) = \frac{\sigma_P^4}{(I-1)^2}\mathrm{Var}\left(\frac{SS_P}{\sigma_P^2}\right) = \frac{2\sigma_P^4}{I-1}
\end{aligned}
\tag{3.38}
$$

Now the Wald confidence interval of the moment estimators T, which is based on asymptotic normality of its parameter estimators, is defined as $\hat{T} \pm z_{1-\alpha/2}SE(\hat{T})$, where $z_{1-\alpha/2}$ is the $(1-\alpha)^{th}$ quantile of the standard normal distribution (Wald & Wolfowitz, 1939).

The F-test

Although the fact that in our data is not normally distributed, but a mixed model with a CMP distribution and a latent variable distribution Z_i, we have by Remark 3.2.2 that for scenario 1 with $a_j = b_k = 1, \forall\, j, k$ it holds that

$$\frac{\mathbb{E}[MS_T]}{\mathbb{E}[MS_E]} = 1 \qquad \text{and} \qquad \frac{\mathbb{E}[MS_O]}{\mathbb{E}[MS_E]} = 1.$$

Note that $\sigma_E^2 = \mathbb{E}[MS_E], \sigma_T^2 = \mathbb{E}[MS_T]$. To test the null hypothesis, we could evaluate the statistic $F_0 = MS_T/MS_E$ with an F-distribution as in the analysis of variance (ANOVA) framework. In this framework it is assumed that the data is independently and identically distributed (i.i.d.) with a normal distribution.

To what extent applying the F-test is valid due to the violation of the underlying distribution, will be shown in Section 5.4. In our case, the hypothesis test for the tool effect is $H_0 : a_j = 1, \forall j$ and $H_1 : \exists j : a_j \neq 1$, where $j = 1, \ldots, J$. Under H_0, we have $\sigma_T^2 = \sigma_E^2$ and therefore the following test statistic

$$F_0 = \frac{MS_T}{MS_E} = \frac{v_1\sigma_T^2 MS_T/(\sigma_T^2 v_1)}{v_2\sigma_E^2 MS_E/(\sigma_E^2 v_2)} = \frac{\sigma_T^2 SS_T/(\sigma_T^2 v_1)}{\sigma_E^2 SS_E/(\sigma_E^2 v_2)} = \frac{SS_T/(\sigma_T^2 v_1)}{SS_E/(\sigma_E^2 v_2)}. \tag{3.39}$$

When we assume that $SS_T/(\sigma_T^2) \sim \chi_{v_1}^2$ and $SS_E/(\sigma_E^2) \sim \chi_{v_2}^2$, we obtain that the test statistic F_0 follows a central F-distribution with $v_1 = J - 1$ numerator and $v_2 = IJK(L-1)$ denominator degrees of freedom. Now, for a level α the critical value is defined by $F_{crit,\alpha} = \{F_c \mid \mathbb{P}(F_{v_1,v_2} \geq F_c) = \alpha\}$. The null hypothesis H_0 is rejected when $F_0 \geq F_{crit,\alpha}$ and accepted if $F_0 < F_{crit,\alpha}$. An analogue hypothesis test to the one for the tool effect holds for the operator effect, where MS_T, σ_T^2, $v_1 = J - 1$ are replaced by MS_O, σ_O^2, $v_1 = K - 1$, respectively.

When the alternative hypothesis is true and we assume a normal distribution of the data, the tool and operator sum of squares follow a non-central chi-square distribution, $\chi_{J-1}^2(\lambda_{Tool})$ and $\chi_{K-1}^2(\lambda_{Operator})$, respectively. Then, for model scenario 1, the non-centrality parameter λ for the test on tool or operator effect would be defined by

$$\begin{aligned}
\lambda_{Tool} &= (J-1)\frac{\mathbb{E}[MS_T] - \mathbb{E}[MS_E]}{\mathbb{E}[MS_E]} = \frac{IKL}{\mathbb{E}[MS_E]} \sum_{j=1}^{J}(a_j^2 - 1)\left(\frac{\mathrm{Var}\,(Z_i)}{I} + \mathbb{E}[Z_i]^2\right), \\
\lambda_{Operator} &= (K-1)\frac{\mathbb{E}[MS_O] - \mathbb{E}[MS_E]}{\mathbb{E}[MS_E]} = \frac{IJL}{\mathbb{E}[MS_E]} \sum_{k=1}^{K}(b_k^2 - 1)\left(\frac{\mathrm{Var}\,(Z_i)}{I} + \mathbb{E}[Z_i]^2\right).
\end{aligned} \tag{3.40}$$

Now we obtain $MS_T/MS_E \sim F_{J-1,IJK(L-1)}(\lambda_{Tool})$ and $MS_O/MS_E \sim F_{K-1,IJK(L-1)}(\lambda_{Operator})$. Note that $\lambda = 0$ implies a central F-distribution. The power of this test is now described by $\mathbb{P}(F_{v_1,v_2,\lambda} \geq F_{crit,\alpha})$.

Case study

In this chapter we consider a case study about the qualification of EUV pellicles with microscopes. These microscopes provide a set of features of the observed defects (e.g. size and location) for each measurement. In absence of a full experiment, we analyze three datasets. The first dataset is about a single part that was measured five times on each of the two microscopes. This dataset is used to demonstrate the lack of fit of a Poisson distribution as repeated measurement distribution, motivate the application of the CMP distribution and show the performance of the pseudo-likelihood methods using the asymptotic expansion. The second dataset is a cross-sectional analysis of 141 parts measured a single time somewhere in time range T[1]. The aim of this dataset is to determine the underlying part to part distribution. The third dataset is about a series of three measurements on each part, that is performed during the production process. Each measurement is conducted by any available tool and operator, indifferent of where and by who the preceding measurement (if any) is performed. The aim here is to validate the assumption of a proportional effect and heterogeneity of the variances.

Note that for the first two datasets, two lower boundaries for the diameter size are used. The selected threshold of 10 arbitrary units (a.u.), is derived from the region of interest in practice, which are all defects larger than 10 a.u.. The second threshold of 4 a.u. originates from the fact that everything below this threshold is considered noise of the microscope. The amount of defects with a diameter size larger than this threshold, is taken as the output for each measurement. The underlying distributions will be used as input for the simulation study in Chapter 5.

4.1 Repeated measurements on single part

A single part from an earlier stage than the time range T was measured five times on both microscope A and microscope B. The observed counts of defects with a size larger than 4 a.u. and 10 a.u. is presented in Figure 4.1. These five measurements on each tool are performed without intermediate operator intervention. Operators were only involved in the placements on the beginning of each sequence of measurements, the removal and placement on the other tool. A fixed effects model was fitted to estimate the tool effects on a single part i with repeats j. Let us define the likelihood of the CMP distribution by

$$L(\theta; y_{11}, \ldots, y_{IJ}) = \prod_{i=1}^{I} \prod_{j=1}^{J} \frac{1}{S(\zeta_{ij}(z), \nu_{ij})} \left(\frac{\zeta_{ij}(z)^{y_{ij}}}{y_{ij}!} \right)^{\nu_{ij}} \tag{4.1}$$

$$\zeta_{ij}(z) = z \cdot \exp\{\boldsymbol{x}_{ij}\boldsymbol{\beta}\} \qquad \text{and} \qquad \nu_{ij} = \exp\{\boldsymbol{w}_{ij}^{T}\boldsymbol{\gamma}\},$$

---

[1]For sake of confidentiality, no specific time range is mentioned.
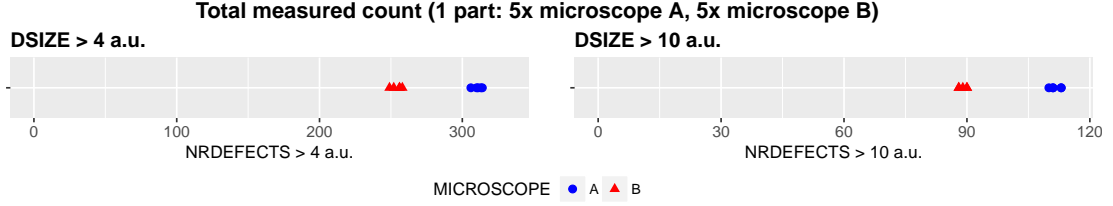
Figure 4.1: Observed number of defects on pellicles larger than a certain diameter size (DSIZE), measured in a certain time range T.

with $\boldsymbol{\beta} = (\beta_0, \beta_1)$ and constant dispersion $\nu_{ij} = \nu_0$. Moreover, β_0 and β_1 correspond to microscope A and B, respectively. Note that in contrast to Chapter 3, z is considered fixed instead of random as the dataset contains of only a single part. This implies that the normalization constant only needs to be determined for every combination of i and j of a single z. Moreover, β_0 is used as reference level and set to zero and $\exp\{\beta_1\}$ is the proportional difference of microscope B with A. We distinguish the case with pre-fixed $\nu_0 = 1$, where the CMP reduces to a Poisson, and a freely estimable $\nu_0 > 0$. Here, the Poisson distribution is a nested model of the CMP distribution. The log-likelihood is defined by

$$l(\theta; y_{11}, \ldots, y_{IJ}) = \sum_{i=1}^{I} \sum_{j=1}^{J} -\log\left(S(\zeta_{ij}(z), \nu_{ij})\right) + \nu_{ij} y_{ij} \log(\zeta_{ij}(z)) - \nu_{ij} \log(y_{ij}!). \tag{4.2}$$

To calculate the likelihood, we again use the convergence of the normalization constant for finite z and $\nu > 1$ as stated in Proposition 3.1.1. A finite summation until the nth term of the normalization constant is used, where n is defined such that the gain of adding the $(n+1)^{th}$ term is lower than 10^{-10}. Given a decimal precision of 10^{-5}, this approximation is equal to the true values of the log-likelihood and parameter estimates.

To assess the improvement of adding an extra estimable parameter ν, we check the Akaike information criterion with a correction for small sample sizes (AICc). This is preferred over the AIC given the low number of observations ($n = 10$). The difference in AICc between the Poisson and CMP shown in Table 4.1 strongly motivates the application of the CMP as a method to model repeated measurements that are severely under-dispersed. In addition, the pseudo-likelihoods that use the expansion of the normalization constant are considered. It can be concluded that in these cases the improvement of increasing the order of the expansion gets marginal as of order two.

Table 4.1: Maximum likelihood estimates (z, β_1, ν_0) for the fixed effect model Table 4.2 on the dataset in Figure 4.1. Note that $CMP(\zeta, 1) = Pois(\zeta)$.

Likelihood

Method	Norm constant	DSIZE	z	β_1	ν_0	LLH	AICc
Poisson	Closed form	> 4 a.u.	466.6	-0.16811	1	-39.6365	85.0
CMP	Finite sum	> 4 a.u.	467.1	-0.16792	36.2	-26.5568	63.1
Poisson	Closed form	> 10 a.u.	111.6	-0.22628	1	-32.2611	70.2
CMP	Finite sum	> 10 a.u.	112.1	-0.22516	91.8	-14.62	39.2

Pseudo-likelihoods

| Method	Norm constant	DSIZE	z	β_1	ν	LLH
CMP	Expansion O1	> 4 a.u.	467.1	-0.16792	36.2	-26.5216
CMP	Expansion O2	> 4 a.u.	467.1	-0.16792	36.2	-26.5567
CMP	Expansion O3	> 4 a.u.	467.1	-0.16792	36.2	-26.5568
CMP	Expansion O4	> 4 a.u.	467.1	-0.16792	36.2	-26.5568
CMP	Expansion O1	> 10 a.u.	112.1	-0.22516	99.5	-14.2237
CMP	Expansion O2	> 10 a.u.	112.1	-0.22516	91.8	-14.6163
CMP	Expansion O3	> 10 a.u.	112.1	-0.22516	91.8	-14.6236
CMP	Expansion O4	> 10 a.u.	112.1	-0.22516	91.8	-14.6237

4.2   Cross-sectional analyses

Single measurement per part

In a certain time range T, 141 pellicles were analyzed using microscopes. A density function of the diameter sizes of all defects combined is shown in Figure 4.2. The empirical cumulative density functions (ECDF) of the observed number of defects on pellicles with a size larger than 4 a.u. and 10 a.u. is visualized in Figure 4.3. Note that these counts originate from measurements performed by all microscopes and not corrected for the effect of variation caused by them. Observe that number of defects larger than 4 a.u. is much better described by a negative binomial distribution than a Poisson distribution. For the count with a diameter size larger than 10 a.u., the difference in fit between a Poisson and negative binomial is much smaller. Here, the small over-dispersion could be the effect of tools and operators.
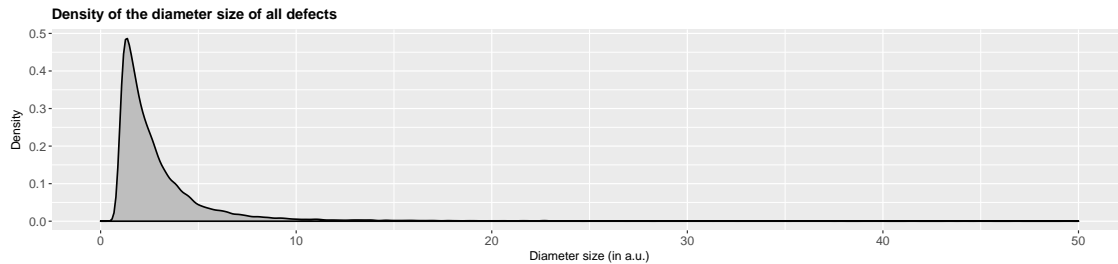

Figure 4.2: Density of diameter size ($DSIZE$) on all pellicles. (Density of defects with a $DSIZE > 50$ a.u. is approximately 0.002.)
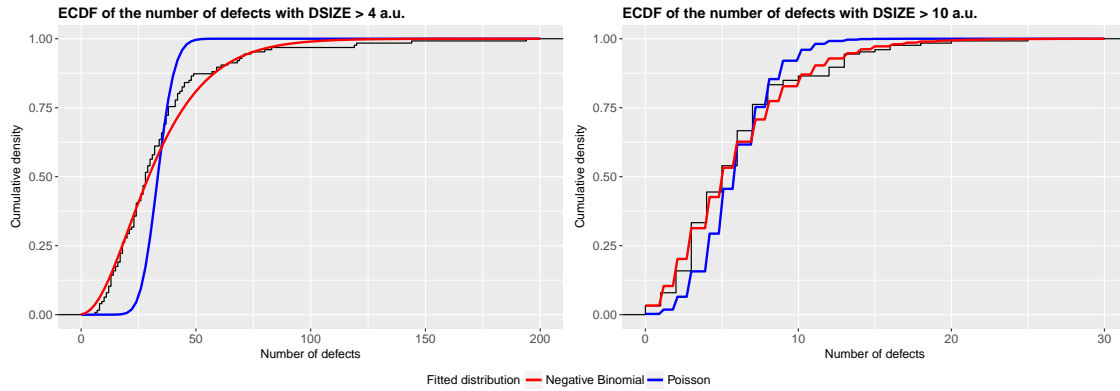
Figure 4.3: Empirical cumulative density function (ECDF) of the observed number of defects on pellicles exceeding a certain size. All initial assessments from a certain time range T.

Three measurements per part

In this dataset, each of the three measurements is performed at any available tool, indifferent of where the preceding measurement (if any) is performed. In Figure 4.4, empirical average of the number of defects larger than 10 a.u. is plotted against the sample variance of the different measurements. Note that the sample variance includes the effect of different tools and process steps. Given the positive slope of the regression line in this plot, it can be derived that assumption of a proportional effect seems reasonable.
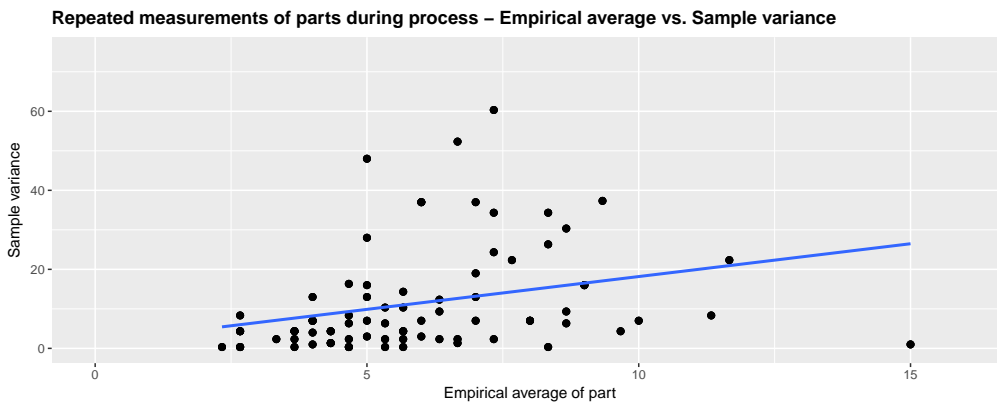
Figure 4.4: Empirical average of the number count on parts vs. the sample variance of the measured count in repeats on the same part.

Simulation Study

In this chapter we define all the characteristics of the simulation study. The goal of this study is to define which of the estimation techniques, that were described in Chapter 3, has the best estimating performance and is most suitable for hypothesis testing. To do this, we will focus on severely under-dispersed count data and simulate various kinds of experiments.

5.1 Context simulation study

In line with the planned experiment at ASML, we will take the experimental setup with *6 parts (P) × 3 tools (T) × 2 operators (O) × 3 repeats (R)* as a reference setting. This setup contains a lower number of operators than tools, since a relatively low contribution to the variation of operator is expected due to measurement setup. Besides, we also consider a larger setting with 10 parts, to determine the gain in estimating and hypothesis testing performance for all methods. In line with the case study in Chapter 4, two lower thresholds for the diameter size are considered, obtaining a mean count of approximately 10 defects for a lower threshold of 10 a.u. and on average 50 defects when considering a lower threshold of 4 a.u.. For the lower counts corresponding to the higher diameter size threshold, a Poisson or negative binomial distribution seemed a good fit. Therefore, the more simple Poisson distribution was selected in the simulation. For the higher average count coming from a lower minimal diameter size threshold, a negative binomial distribution was used since this distribution fitted much better than the Poisson. Additionally, the effect sizes of tool and operator are derived from the repeated measurements dataset of the case study. Finally, the robustness of the CMP based estimation methods to different underlying distributions is tested by simulating from a GP distribution. In the simulation study, a somewhat less severe dispersion level was used (mean = γ·variance, with $\gamma = 8, 20$) than the extreme level of under-dispersion that was observed in the case study (mean = γ·variance with $\gamma = 36, 92$).

5.2 Generating data

In this section, we elaborate on the way the simulated data is generated. Firstly, we present overview of all simulation scenarios that are considered in Table 5.1. Given a long analysis time of some methods, the number of simulations is limited to 500 for all simulation scenarios. For the method of moments analysis on the robustness of the F-test for data from a CMP repeated measurements distribution and some discrete prior distribution (Poisson or negative binomial), a total number of 10,000 simulations is performed.

In the explanation of the simulation setup, we use the same notation as was used for the method of moments in Section 3.2. Recall that we denote the parts $i = 1, \ldots, I$, the tools $j = 1, \ldots, J$,

Table 5.1: All simulation scenarios representing experiments with various settings. The experiments have a certain number of parts (P), tools (T), operators (O) and repeats (R) and latent variable describing the true count on the parts (Poisson(θ) and negative binomial NB(θ, κ), with mean θ and variance $\theta + \kappa\theta^2$ (so the size parameter is $1/\kappa$)).

Sim. NR.	Experiment Setup ($PxTxOxR$)	Part dist. (LV)	Repeat dist.	$Var/mean$	Tool effects	Operator effects	NSIMS
		$Pois(\theta)$			(a_1, a_2, a_3)	(b_1, b_2)	
1	6 x 3 x 2 x 3	$Pois(10)$	GP	0.05	(1,1,1)	(1,1)	500
2	6 x 3 x 2 x 3	$Pois(10)$	GP	0.05	(1,1.15,0.85)	(1.05,0.95)	500
3	6 x 3 x 2 x 3	$Pois(10)$	GP	0.125	(1,1,1)	(1,1)	500
4	6 x 3 x 2 x 3	$Pois(10)$	GP	0.125	(1,1.15,0.85)	(1.05,0.95)	500
5	10 x 3 x 2 x 3	$Pois(10)$	GP	0.05	(1,1,1)	(1,1)	500
6	10 x 3 x 2 x 3	$Pois(10)$	GP	0.05	(1,1.15,0.85)	(1.05,0.95)	500
7	10 x 3 x 2 x 3	$Pois(10)$	GP	0.125	(1,1,1)	(1,1)	500
8	10 x 3 x 2 x 3	$Pois(10)$	GP	0.125	(1,1.15,0.85)	(1.05,0.95)	500
9	10 x 3 x 2 x 3	$Pois(10)$	CMP	0.05	(1,1,1)	(1,1)	500
10	10 x 3 x 2 x 3	$Pois(10)$	CMP	0.05	(1,1.15,0.85)	(1.05,0.95)	500
11	10 x 3 x 2 x 3	$Pois(10)$	CMP	0.125	(1,1,1)	(1,1)	500
12	10 x 3 x 2 x 3	$Pois(10)$	CMP	0.125	(1,1.15,0.85)	(1.05,0.95)	500
		$NB(\theta, \kappa)$			(a_1, a_2, a_3)	(b_1, b_2)	
13	6 x 3 x 2 x 3	$NB(50, 1/2)$	CMP	0.05	(1,1,1)	(1,1)	500
14	6 x 3 x 2 x 3	$NB(50, 1/2)$	CMP	0.05	(1,1.15,0.85)	(1.05,0.95)	500
15	6 x 3 x 2 x 3	$NB(50, 1/2)$	CMP	0.125	(1,1,1)	(1,1)	500
16	6 x 3 x 2 x 3	$NB(50, 1/2)$	CMP	0.125	(1,1.15,0.85)	(1.05,0.95)	500
17	10 x 3 x 2 x 3	$NB(50, 1/2)$	CMP	0.05	(1,1,1)	(1,1)	500
18	10 x 3 x 2 x 3	$NB(50, 1/2)$	CMP	0.05	(1,1.15,0.85)	(1.05,0.95)	500
19	10 x 3 x 2 x 3	$NB(50, 1/2)$	CMP	0.125	(1,1,1)	(1,1)	500
20	10 x 3 x 2 x 3	$NB(50, 1/2)$	CMP	0.125	(1,1.15,0.85)	(1.05,0.95)	500

the operators $k = 1, \ldots, K$ and repeats $l = 1, \ldots, L$, where we use $I \in \{6, 10\}$, $J = 3$, $K = 2$ and $L = 3$. The simulation start by drawing the samples I from latent variable distribution Z, describing the true counts on part i. This is either a Poisson distribution, $Pois(\theta)$ with the probability mass function (PMF)

$$P(Z = z) = \begin{cases} \frac{\theta^z}{z!} \exp\{-\theta\}, & z = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases} \tag{5.1}$$

or a negative binomial distribution, $NB(\theta, \kappa)$ with the PMF

$$P(Z = z) = \begin{cases} \frac{\Gamma(z+1/\kappa)}{\Gamma(z+1)\Gamma(1/\kappa)} \frac{(\theta\kappa)^z}{(1+\theta\kappa)^{z+1/\kappa}}, & z = 0, 1, 2, \ldots \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

Given these true counts z_i with $i = 1, \ldots, I$, we simulate repeated measurements from a CMP or GP distribution. For the CMP distribution we take a fixed dispersion parameter $\nu_{sim} \in \{8, 20\}$, which is according to the first order approximations of the moments roughly described by the ratio of the conditional variance and conditional mean ($\nu \approx (\mathbb{E}[Y|Z] + 0.5)/\text{Var}[Y|Z] \approx (\mathbb{E}[Y|Z])/\text{Var}[Y|Z]$). Note that especially for high true counts, the difference between $\mathbb{E}[Y|Z]+0.5$ and $\mathbb{E}[Y|Z]$ gets small.

We now take the simulation parameter $\zeta_{sim,ijk} = z_i \cdot a_j \cdot b_k + (\nu_{sim,ijk} - 1)/2\nu_{sim,ijk}$, where a_j is the effect of tool j and b_k the effect of operator k. In case of the GP distribution, we

have a fixed dispersion parameter that directly follows from the ratio of the conditional variance and conditional mean, namely $\varphi_{sim} = \sqrt{var/mean} - 1$. Moreover, the simulation parameter $\tau_{sim,ijk} = z_i \cdot a_j \cdot b_k$. Having defined the parameter settings for each part i, tool j and operator k combination, we now draw three samples for each combination from either the $CMP(\zeta_{sim,ijk}, \nu_{sim})$ or $GP(\tau_{sim,ijk}, \varphi_{sim})$ distribution, obtaining the observations Y_{ijk1}, Y_{ijk2} and Y_{ijk3}.

The full simulation study is performed in the statistical software [R], version 3.4.2. To generate random values from a CMP distribution we could use the *rCMP* function of the *'COMPoissonReg'* Package in [R] (Sellers et al., 2017). However, this function falls short when using too large values for $\lambda(= \zeta^\nu)$ due to a too low precision that is used in the calculation of the normalization constant. As a result of this, all probability masses become 0 and therefore the quantiles can't be determined. To solve this precision issue, the *rCMP-Paus* function is applied. This function, which is shown in Appendix B.1, is an adapted version of the *rCMP* function of the *'COMPoissonReg'* Package in [R], where the internal precision is increased using the *'Rmpfr'* package (Maechler, 2016).

For generating random samples from the GP distribution, we use the *GP-Paus* function, which is described in Appendix B.2. This new function uses the probability masses of the GP distribution obtained by the *dgenpois* function in the *'VGAM'* package in [R] (Yee, 2017). Since the cumulative distribution function (CDF) does not always sum to 1 in case of under-dispersion, due to the truncation, we normalize the densities with the cumulative distribution function of the GP. In the *GP-Paus* function, a random number is drawn from the uniform distribution $U[0, 1]$, after which the corresponding quantile is determined using the normalized probability masses, such that $q(u) = \inf\{x : F(x) \geq u\} = F^{-1}(u)$. In addition, we relax the constraint lower limit of the dispersion parameter $\lambda \geq -1$, since this is overly conservative and not allowing for more severe dispersion when the mean count is high. To assure that the probability mass sums to 1, we structurally normalize the probability masses with the CDF.

5.3 Analysis simulated experiments

Analysis methods

The simulation scenarios from Table 5.1 are analyzed with various methods to determine the estimation and testing performance. An overview of all analysis methods and the corresponding settings can be found in Table 5.2. In the analysis, we make a distinction based on the latent variable (L.V.) distribution used to simulate the data. The analysis methods with a Poisson L.V. distribution are only applied on the simulated datasets (scenario 1-12 with a Poisson L.V.) and the negative binomial L.V. analysis method only to the simulated datasets from the same distribution (scenario 13-20). Moreover, two general models are considered for all simulated datasets, namely analysis method 15 and 16. Notice that the truncation of the latent variable used for analysis method (1-6 and 8-13) are flexible, in a sense that the truncation term is determined by the relative contribution criterion, as described in (3.5), and the inserted parameter values of that iteration step. Analysis method number 1 is considered the reference likelihood (*Ref.LLH*) as it is closest to the true likelihood of the probability model. Due to its double summation, the number of calculation steps grows quadratically, making this procedure extremely inefficient for large counts. Due to the high computation time of the reference likelihood for the datasets from simulation scenarios 13-20 (average of 21 to 43 hours per dataset), we only apply this analysis method for 100 out of the 500 simulated datasets. For the methods that use a truncation for either the normalization constant or the latent variable distribution, a relative contribution criteria of 10^{-5} is applied.

Implementation in software

The maximization of the (pseudo) likelihood functions is actually a constrained optimization problem. To find the (pseudo) maximum likelihood estimates, the *optim* function of the [R] package *'stats'* was used (R Core Team, 2017). In order to reduce the computation time of the calculation of

Table 5.2: Analysis methods that are used to analyze datasets from the different simulation scenarios in Table 5.1, using different approximation methods for the normalization constant (NC) and latent variable (L.V.) distributions representing the count on parts.

Nr.	Name method [1]	Method	Distribution		Convergence criteria	
			L.V.	Repeats	L.V. integral (z)	Approximation NC
Simulated datasets with Poisson L.V. (Scenario 1-12)						
1	Ref.LLH.poisson	LLH	Poisson	$CMP(\zeta,\nu)$	Truncation	Truncation
2	Exp.O1.poisson	LLH	Poisson	$CMP(\zeta,\nu)$	Truncation	Expansion (order 1)
3	Exp.O2.poisson	LLH	Poisson	$CMP(\zeta,\nu)$	Truncation	Expansion (order 2)
4	Exp.O3.poisson	LLH	Poisson	$CMP(\zeta,\nu)$	Truncation	Expansion (order 3)
5	Exp.O4.poisson	LLH	Poisson	$CMP(\zeta,\nu)$	Truncation	Expansion (order 4)
6	Pois.nu1.poisson	LLH	Poisson	$CMP(\zeta,1)$	Truncation	Closed form
7	MOM.poisson	MoM	Poisson	$CMP(\zeta,\nu)$	NA	Expansion (order 1)
Simulated datasets with negative binomial L.V. (Scenario 13-20)						
8	Ref.LLH.nbinom	LLH	NB	$CMP(\zeta,\nu)$	Truncation	Truncation
9	Exp.O1.nbinom	LLH	NB	$CMP(\zeta,\nu)$	Truncation	Expansion (order 1)
10	Exp.O2.nbinom	LLH	NB	$CMP(\zeta,\nu)$	Truncation	Expansion (order 2)
11	Exp.O3.nbinom	LLH	NB	$CMP(\zeta,\nu)$	Truncation	Expansion (order 3)
12	Exp.O4.nbinom	LLH	NB	$CMP(\zeta,\nu)$	Truncation	Expansion (order 4)
13	Pois.nu1.nbinom	LLH	NB	$CMP(\zeta,1)$	Truncation	Closed form
14	MOM.nbinom	MoM	NB	$CMP(\zeta,\nu)$	NA	Expansion (order 1)
Simulated datasets with Poisson or negative binomial L.V. (Scenario 1-20)						
15	Gamma	LLH	Gamma	$CMP(\zeta,\nu)$	Closed form	Expansion (order 1)
16	QP.Norm	LLH	Normal	$Pois(\theta,\delta)$	GHQ	NA

the maximum likelihood estimates, selecting proper starting values in the optimization is key. For the method of moments approach, no starting values are required. Recall that a Poisson GLMM was fitted with the *glmer* function in [R] and the deviance was scaled with the degrees of freedom to obtain the dispersion parameter. Here, the function's default procedure for starting values is used. For the analysis methods $1-6$, $8-13$ and 15, starting values should be provided. A fixed effects model could be used, however, the *glm.cmp* function of the [R] package *'COMPoissonReg'* (Sellers et al., 2017) has numerical problems for large values $\lambda = \zeta^\nu$ as stated in the previous paragraph. We therefore use a fixed effects generalized model with a quasi-Poisson distribution to determine starting values. Here, the estimates for tools and operators (β_j) will be directly used. Moreover, the reciprocal of the observed dispersion of the quasi-Poisson distribution, that describes the ratio of the variance and the mean, is used as starting value for ν_0. The starting values for the parameters of the latent variable distribution are obtained by fitting the prior distribution corresponding to the analysis method on the fixed effects estimates of the individual parts. Moreover, to approximate the integral of the likelihood of the quasi-Poisson model with normal random effects, the Gauss-Hermite quadrature (GHQ) technique is applied, as this is considered the most reliable method among the implemented approximation procedures.

5.4 Performance measures

In order to assess the quality of the estimated models, we define a set of performance measures. The obtained estimates are evaluated in terms of bias and mean square error (MSE). Firstly, the bias describes the difference between expected value and the estimated value of the parameter and is denoted by $b(\hat{\theta}) = \mathbb{E}_{\hat{\theta}}[\hat{\theta} - \theta]$. Secondly, the mean square error of an estimator is defined by $MSE(\hat{\theta}) = \mathbb{E}_{\hat{\theta}}[(\hat{\theta} - \theta)^2]$. In addition, the coverage probability of the estimates is determined. This denotes the proportion of the times the true value for the parameter of interest is contained in the 95%-confidence interval.

Although the main focus of the simulation is on the quality of the estimators and hypothesis tests, the deviation in terms of likelihood value for the asymptotic expansions compared to the reference likelihood (which is the truncation approach). This will provide insight in the required order of the expansion, to get a good approximation of the 'true likelihood'.

Based on the method of moments, we established an hypothesis test for the measurement system's tool and operator effect analogue to the F-test for ANOVA, see Paragraph 3.2. The robustness of the F-test to the CMP latent variable model is assessed by a Kolmogorov-Smirnov test. In this goodness of fit, the empirical cumulative distribution function (ECDF) of the observed F-statistic F_0 and the theoretical cumulative distribution function (CDF) of the (non-)central F-distribution are compared. Note that the KS-test is actually a test for continuous distributions, whereas ties could occur in the mean squares due to the discrete nature of the data. Nevertheless, this test will give a good approximation of the goodness of fit. Finally, the Type I error given and the observed power based of the test.

Results

In this chapter we will first elaborate on the performance of the asymptotic expansion based approximation of the normalization constant. Secondly, the results of the simulation study will be presented. Here, we visualize the performance measures for the twenty simulation scenarios that represent realistic experiments. After this, the performance on some large sized experiments are presented to get a brief idea of the asymptotic behavior of all methods.

6.1 Performance asymptotic expansion of the normalization constant

In this section we present a small study on the percentage error of the asymptotic expansion of the normalization constant defined in (2.8) with the true normalization constant $S(\zeta, \nu)$ in (2.5) in case of under-dispersion. In Table 6.1, the percentage errors are provided for settings of the CMP parameters ζ and ν for five different orders of the asymptotic expansion. Note that the constants for the asymptotic expansion c_n were defined in (D.5) and that c_0 denotes a first order asymptotic expansion, $c_0 - c_1$ a second order and so on.

Observe in Figure 6.1 the relation between the ratio of the CMP parameters ν and ζ and the percentage error that is obtained. Here, the actual value of ζ is basically irrelevant in comparison to the ratio, which can be derived from the fact that the percentage errors for a certain combination of order n and ν/ζ ratio do overlap. When the ratio $\nu/\zeta \leq 1$ the relative difference seems more or less negligible for a second order asymptotic expansion or higher. A small difference becomes visible for $1 < \nu/\zeta \leq 5$ (in particular for the first order asymptotic expansion) and the difference becomes large when the ratio $\nu/\zeta > 5$, in particular for the higher order asymptotic expansions.
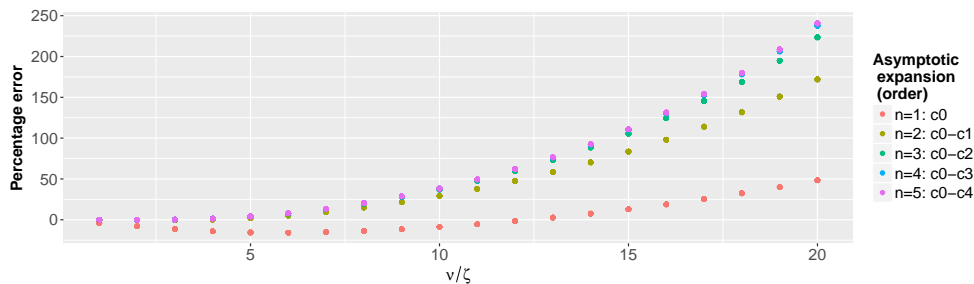
Figure 6.1: Percentage errors for the order n asymptotic expansion of the normalization constant. Here, the x-axis describes the ratio of the CMP parameters ν and ζ. Note that for each ratio ν/ζ and asymptotic expansion order, various values for ζ are used, namely $\zeta = \{2, 5, 10, 20, 25, 50, 100, 200)\}$.

Table 6.1: The percentage error of the asymptotic expansion approximation of the CMP normalization constant $S(\zeta, \nu)$ from (2.5) with the true normalization constant.
Note that the true normalization constant $S(\zeta, \nu)$ is determined by summing until k such that the relative improvement of adding the $(k + 1)^{th}$ term is lower than 10^{-15}. Since the purpose is to determine the difference with the asymptotic expansion method.

		ν					
ζ		2	5	8	20	50	100
	c_0	-1.312	-3.995	-6.422	-14.196	-8.911	48.052
	$c_0 - c_1$	-0.078	-0.155	-0.281	0.068	29.028	171.417
5	$c_0 - c_2$	-0.009	-0.002	-0.003	1.326	37.005	222.937
	$c_0 - c_3$	-0.001	0.001	0.000	1.383	38.077	237.130
	$c_0 - c_4$	0.000	0.000	-0.001	1.382	38.173	239.991
	c_0	-0.639	-1.999	-3.247	-7.986	-15.567	-8.835
	$c_0 - c_1$	-0.018	-0.039	-0.072	-0.338	2.016	29.147
10	$c_0 - c_2$	-0.001	0.000	0.000	-0.001	3.865	37.078
	$c_0 - c_3$	0.000	0.000	0.000	0.007	3.989	38.171
	$c_0 - c_4$	0.000	0.000	0.000	0.007	3.994	38.281
	c_0	-0.316	-1.000	-1.632	-4.076	-9.827	-15.552
	$c_0 - c_1$	-0.004	-0.010	-0.018	-0.089	-0.438	2.039
20	$c_0 - c_2$	0.000	0.000	0.000	-0.001	0.055	3.876
	$c_0 - c_3$	0.000	0.000	0.000	0.000	0.072	4.003
	$c_0 - c_4$	0.000	0.000	0.000	0.000	0.072	4.009
	c_0	-0.125	-0.400	-0.655	-1.650	-4.080	-7.986
	$c_0 - c_1$	0.000	-0.001	-0.003	-0.014	-0.085	-0.319
50	$c_0 - c_2$	0.000	0.000	0.000	0.000	-0.001	0.001
	$c_0 - c_3$	0.000	0.000	0.000	0.000	0.000	0.010
	$c_0 - c_4$	0.000	0.000	0.000	0.000	0.000	0.010
	c_0	-0.062	-0.200	-0.328	-0.828	-2.061	-4.081
	$c_0 - c_1$	0.001	0.000	-0.001	-0.004	-0.022	-0.085
100	$c_0 - c_2$	0.001	0.000	0.000	0.000	0.000	-0.001
	$c_0 - c_3$	0.001	0.000	0.000	0.000	0.000	0.000
	$c_0 - c_4$	0.001	0.000	0.000	0.000	0.000	0.000

6.2 Simulation study

Firstly, we present an overview of the number of completed simulations in Table 6.2 and the corresponding average computation time for the likelihood based methods in Table 6.3. A more extensive overview of the optimization times can be found in Appendix C.1. An exponential increase in the computation time when increasing the average count was observed for the analysis methods that have no closed form expression (*Exp.O1* to *Exp.O4*, *Ref.LLH* and *Pois.nu1*).

Subsequently, the bias, mean squared error and coverage probabilities of the estimators are presented. Since the scale of the parameters differs, most parameters are presented individually, however, the estimating performance of the parameters for the tools and operators are combined. Note that the conclusions that can be directly drawn from the figures are contained in the caption of that particular figure. More specific, the x-axis will contain the analysis method and the y-axis the performance measure.

Table 6.2: Number of completed observations out of the 500 performed simulations for all methods (exception: only 100 simulations for the *Ref.LLH* analysis of datasets 13-20).
Apart from the Gamma and QP.norm method, the suffix of the latent variable is suppressed. For dataset 1-12 this is Poisson and for 13-20 negative binomial.

| Sim.sc	Asymptotic expansion				Gamma	Pois.nu1	QP.norm	Ref.LLH	MOM
	O1	O2	O3	O4					
Simulated datasets with Poisson prior									
1	500	500	500	500	500	500	500	500	500
2	500	500	500	500	500	500	500	500	500
3	500	500	500	500	500	500	500	500	500
4	500	500	500	500	500	500	500	500	500
5	497	498	500	500	500	500	500	500	500
6	500	500	500	500	500	500	500	500	500
7	500	500	500	500	500	500	500	500	500
8	500	500	500	500	500	500	500	500	500
9	499	500	500	499	500	500	500	500	500
10	500	500	500	500	500	500	500	500	500
11	500	500	500	500	500	500	500	500	500
12	500	500	500	500	500	500	500	500	500
Simulated datasets with negative binomial prior									
13	491	419	418	416	500	499	500	96	500
14	480	405	407	410	499	500	500	98	500
15	492	437	440	441	500	500	500	100	500
16	472	424	419	419	500	498	500	99	500
17	486	320	329	324	500	497	500	100	500
18	450	303	316	321	499	497	500	98	500
19	488	351	340	356	500	498	500	100	500
20	463	342	351	351	499	500	500	96	500

Table 6.3: Average computation times the simulated datasets 1-12 with a mean count of 10 and the simulated datasets 13-20 with a mean count of 50 for the likelihood based analysis methods.

|	Exp.O1	Exp.O2	Exp.O3	Exp.O4	Gamma	Pois.nu1	QP.norm	Ref.LLH
Scenario 1-12	5.21	5.88	6.27	6.61	0.30	1.20	0.02	27.66
Scenario 13-20	26.01	25.54	28.73	31.18	0.25	13.98	0.02	1736.64

Ratio with the reference likelihood

In Figure 6.2, the ratios of the likelihood values of all likelihood based methods with the reference likelihood *(Ref.LLH)* are visualized. The difference of the *Ref.LLH* with the analysis methods *Pois.nu1* and *QP.norm* is relatively large. For a better comparison between the expansion methods with the reference likelihood, see Figure C.2 in Appendix C.1. By comparing the simulation scenarios with a Poisson latent variable with the negative binomial prior, it can be seen that if the average count is higher, the relative differences compared to the reference likelihood are lower. Moreover, for the asymptotic expansion of a second order (*Exp.O2*), the improvement of increasing the order is negligible as this method already closely describes the reference likelihood. This can be explained by means of the applied ratios of ν/ζ from the previous section, since a count on a part near zero will imply a high ratio of ν/ζ. Since in our case we have $\nu_{ij}/\zeta_{ij}(Z_i)$, a $Pois(10)$ will more likely have low counts, compared to the $NBinom(50, 1/2)$. In addition, the first order expansion (*Exp.O1*) shows the same pattern as the one with the gamma prior (*Gamma*), although more variation is observed for the *Gamma*.
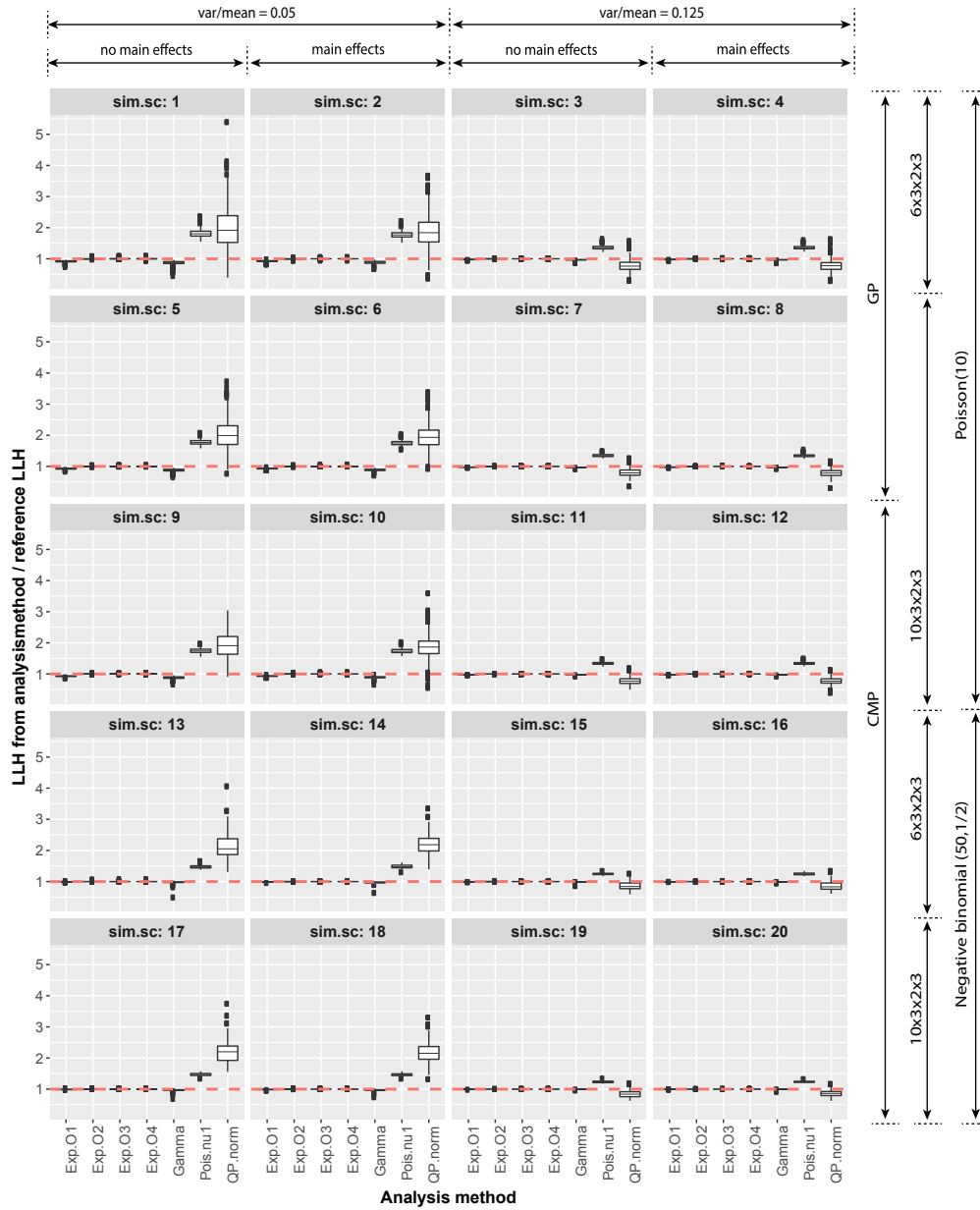
Figure 6.2: Ratio of the (pseudo)-log-likelihoods of all analysis methods with the reference likelihood setting (dotted red line) for simulation scenarios 1 to 20. *Number of simulations = 100.*
1) Var/mean 0.05 vs. 0.125: More spread in the LLH value is observed when the under-dispersion is more severe.
2) Experiment size ($6 \times 3 \times 2 \times 3$ vs. $10 \times 3 \times 2 \times 3$): The larger the experiment, the smaller the variation becomes relative to the Ref.LLH

Performance measures - Main effects (tools and operators)

The bias and mean squared error of the point estimates for the tool and operator effects are presented in Figure 6.3 and Figure 6.4, respectively. Moreover, the coverage probabilities of the tool and operator effects for a 95%-confidence interval are presented in Figure 6.5.
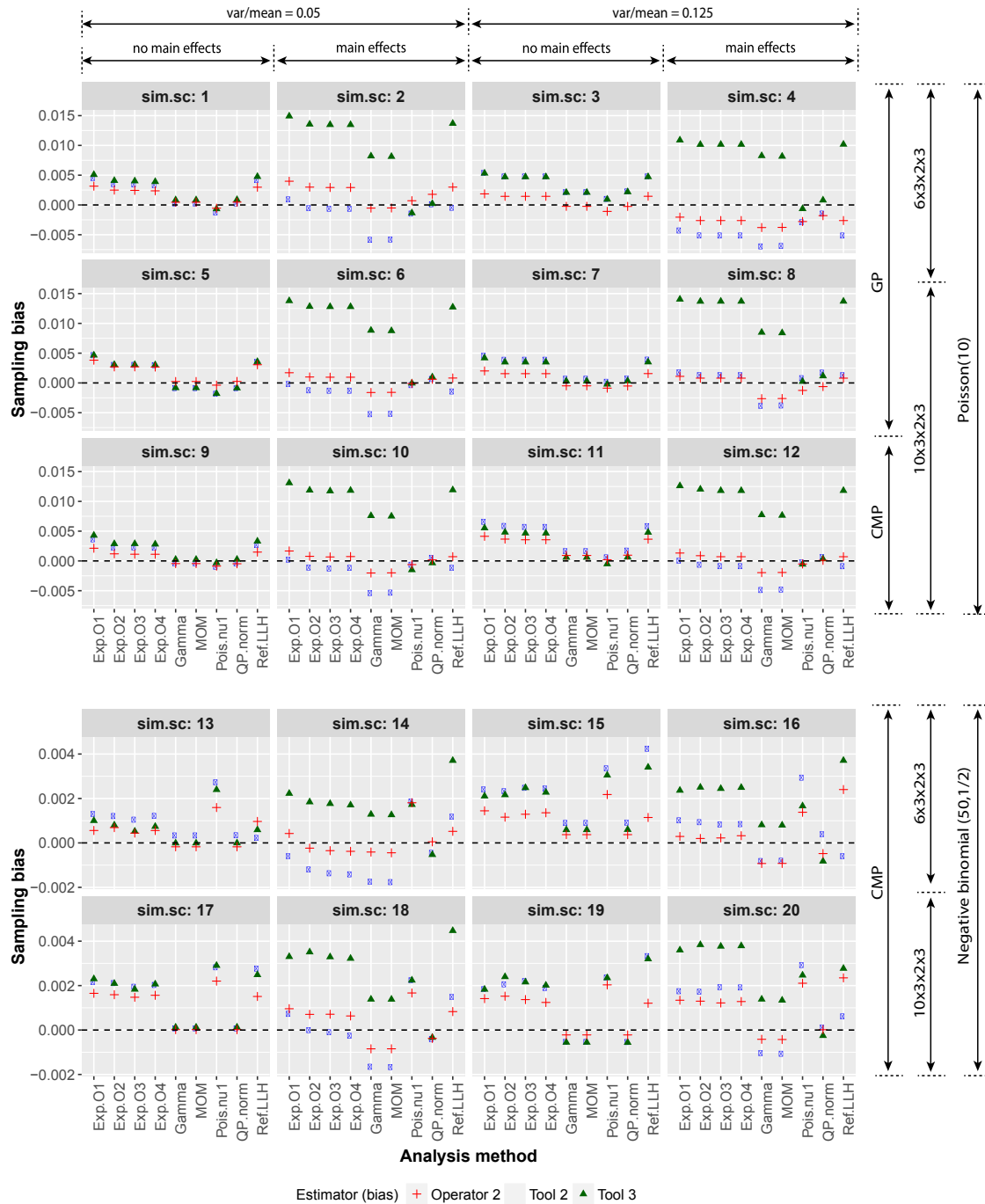
Figure 6.3: Sampling bias of the estimates for the tool and operator effects of simulation scenario 1 to 20. All methods except the method of moments (MOM) present a bias of the covariates that are in the log-link function. For the MOM the bias of a_j and b_k is presented (see (3.23)).
Number of simulations = 500 (only 100 for scenario 13-20 with Ref.LLH).
1) In general, a higher bias was obtained for the models with main effects. In particular, the expansion methods (Exp.O1 to Exp.O4) and the reference likelihood have a relatively large bias compared to the other methods, although small in absolute terms (max. 0.015). Where the Gamma, MOM, Pois.nu1 seem unbiased in case of no main effects, the QP.norm seems always unbiased.
2) Considering the scale of the y-axis, a lower bias was found for the scenarios simulated from a negative binomial distribution with a higher average count compared to the scenarios from a Poisson distribution with a lower count.

41
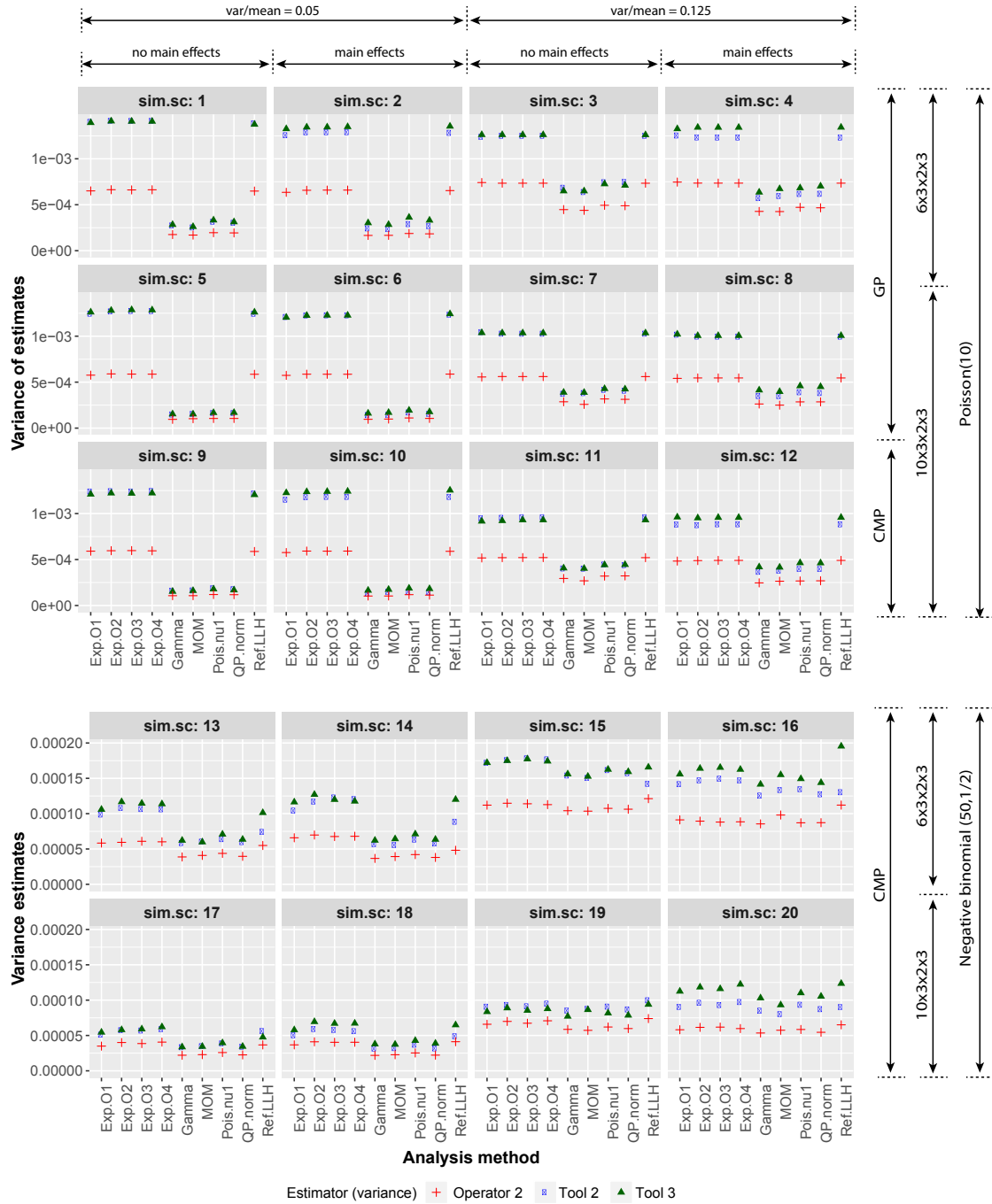
Figure 6.4: Mean squared error of the estimates for the tool and operator effects of simulation scenario 1 to 20. *Number of simulations = 500 (only 100 for scenario 13-20 with Ref.LLH).*
1) L.V. distribution: the general level of the MS_E is lower for the scenarios 13-20 with a negative binomial prior compared to a Poisson prior in scenario 1-12. This implies that if the count is higher, the effects can estimated in a better way.
2) Effect dispersion and average count (Sc. 13, 14, 17, 18 vs. 15, 16, 19, 20): For high average count the point estimates are relatively more precise (smaller variance) for severe under-dispersion (var/mean=0.05) than less severe under-dispersion (var/mean=0.125).
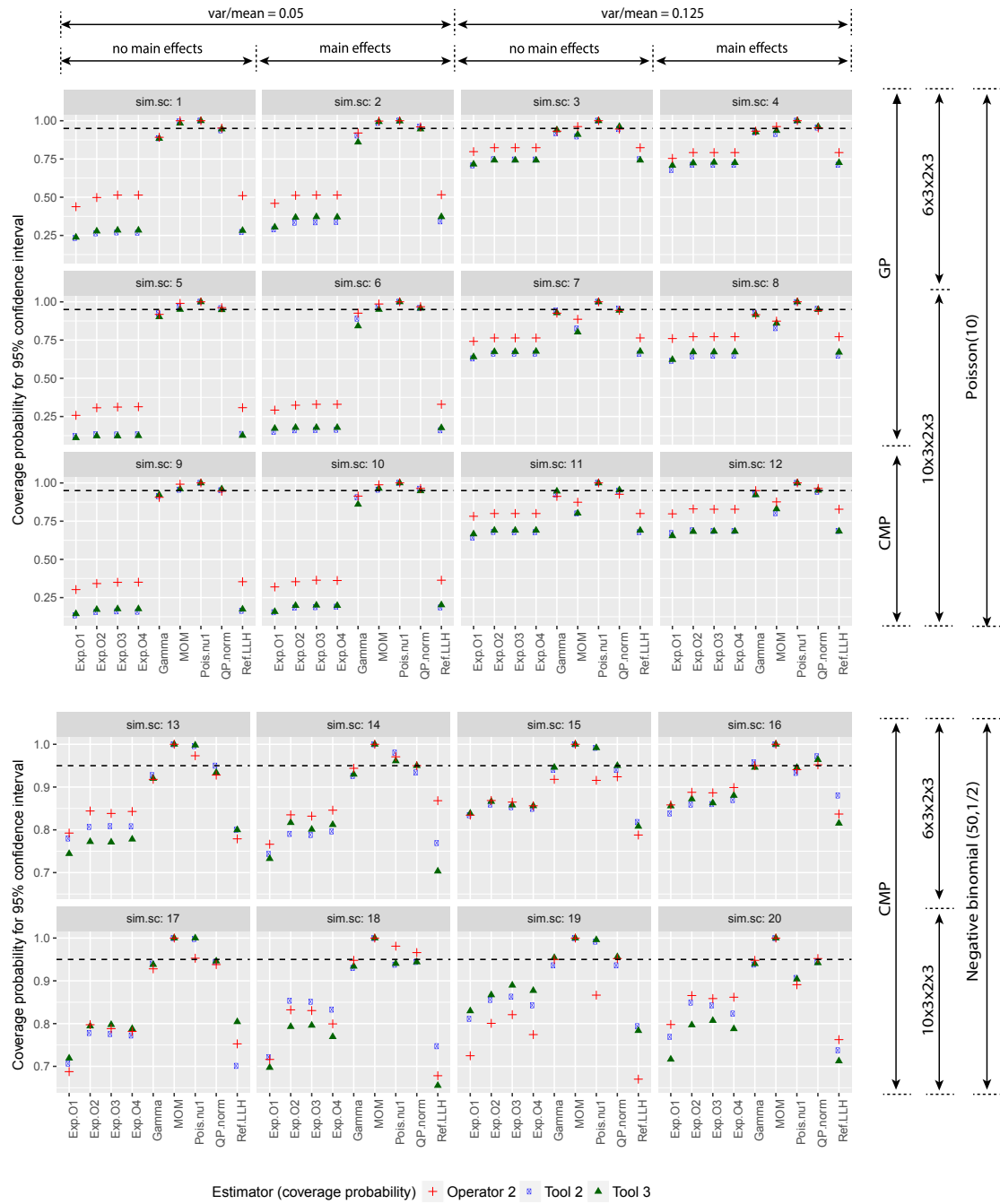
Figure 6.5: Coverage probabilities for a 95 % confidence interval of the estimates for the tool and operator effects of simulation scenario 1 to 20.
Number of simulations = 500 (only 100 for scenario 13-20 with Ref.LLH).
1) The Gamma and QP.norm basically have coverage probabilities of around 95% for all scenarios. The Pois.nu1 is overly conservative for scenarios 1-12. 2) The expansion methods Exp.O1 to Exp.O4 and the reference likelihood Ref.LLH are anti-conservative for all simulation scenarios. However, the coverage probabilities are substantially higher for the scenarios with a higher count and negative binomial prior compared to the Poisson scenarios with a lower count

Performance measures - Dispersion and latent variable parameters

For the dispersion and latent variable parameters, the coverage probabilities are presented in Figure 6.6 and Figure 6.8, respectively. Note that the *QP.norm* approach is not included for the dispersion, since standard errors were derived for the dispersion parameter of the quasi-Poisson. The bias and variances of the point estimate for the dispersion are shown in Appendix C.1. Here, a significant bias for the *Gamma* approach was found, which explains the low coverage probability for the scenarios with a Poisson distribution, combined with severe under-dispersion. Moreover, the sampling bias of the mean of the latent variable is visualized in Figure 6.7. Finally note that the coverage probability of the mean of the latent variable is not provided for the *Gamma* approach, however, given the low observed bias, it is expected that this probability is near 0.95.
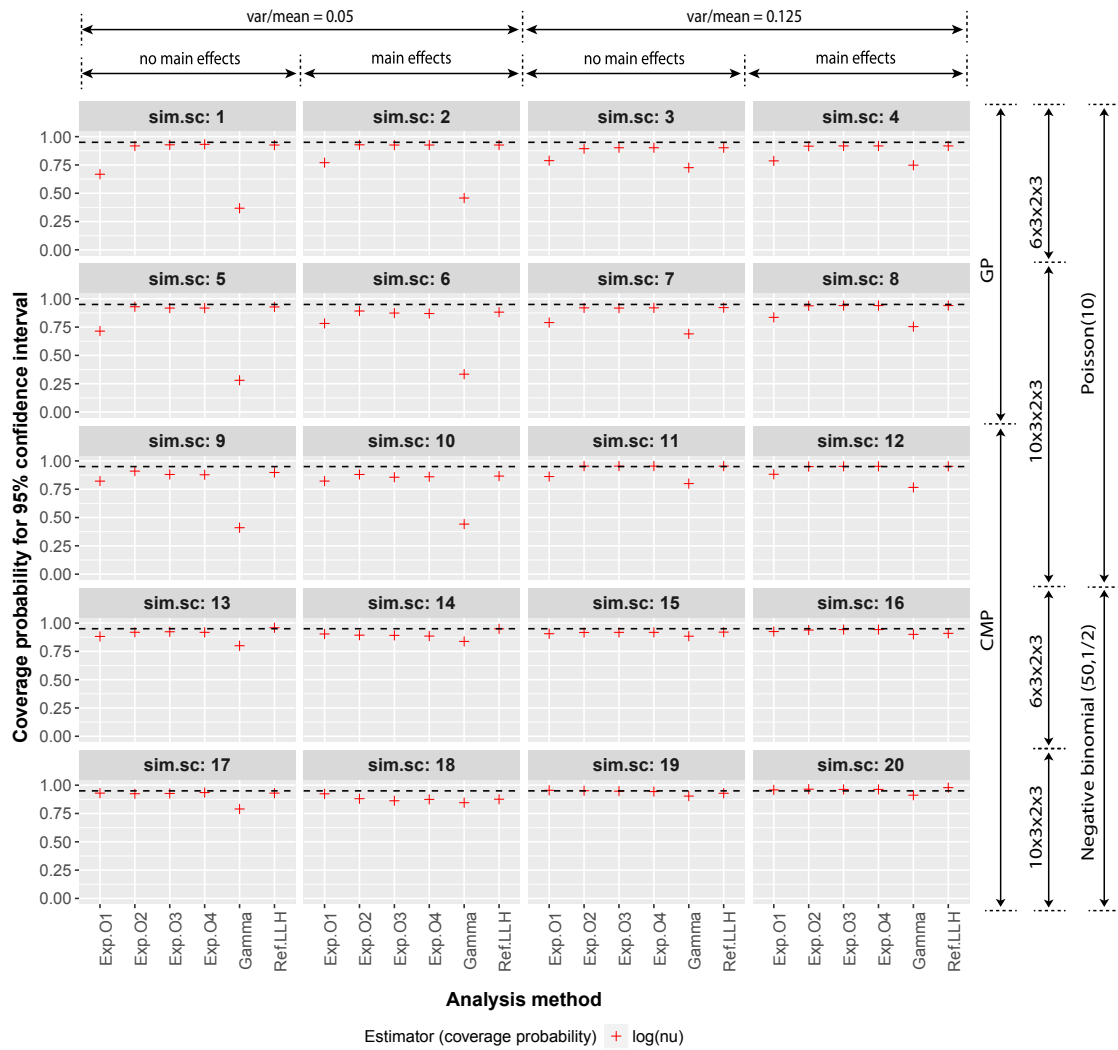
Figure 6.6: Coverage probabilities for a 95 % confidence interval of the dispersion parameter $\gamma_0 = \log(\nu_0)$ of simulation scenario 1 to 20. *Number of simulations = 500 (only 100 for scenario 13-20 with Ref.LLH).*

Note that the MOM obtained a coverage probability of zero for all methods.

1) A low coverage probability for the Gamma is observed for low counts with severe under-dispersion. Also the first order approximation Exp.O1 has a lower coverage probability.

2) The coverage probability seems to be closer to 0.95 in case of a higher count.
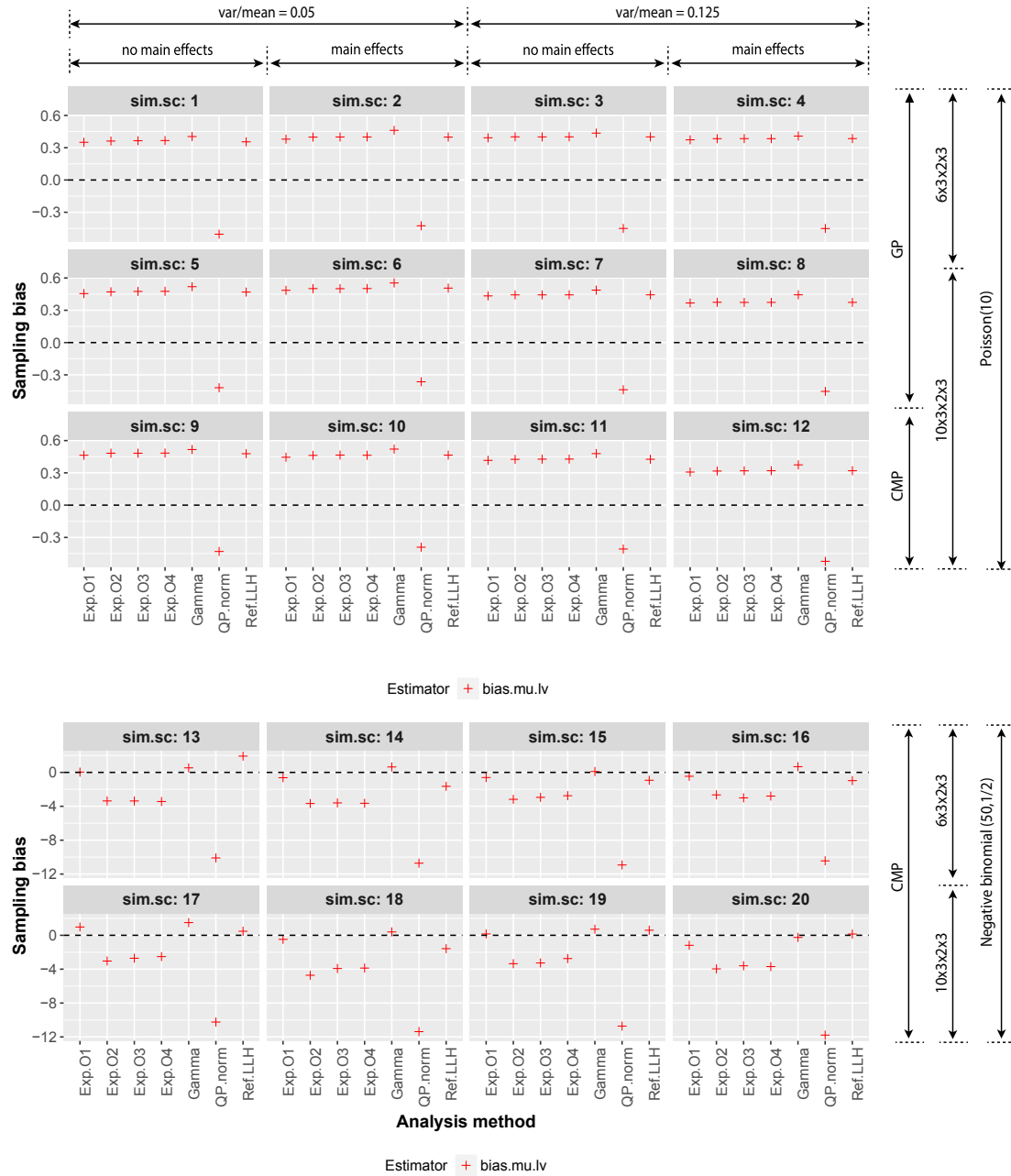
Figure 6.7: Sampling bias for the expected value of the latent variable.
Number of simulations = 500 (only 100 for scenario 13-20 with Ref.LLH).
1) Whereas for the scenarios with a Poisson distribution seem approximately equal for all methods, the QP.norm method obtains a large bias for negative binomial distribution, which is more skewed as a result of the over-dispersion. Here the Gamma is more or less unbiased.
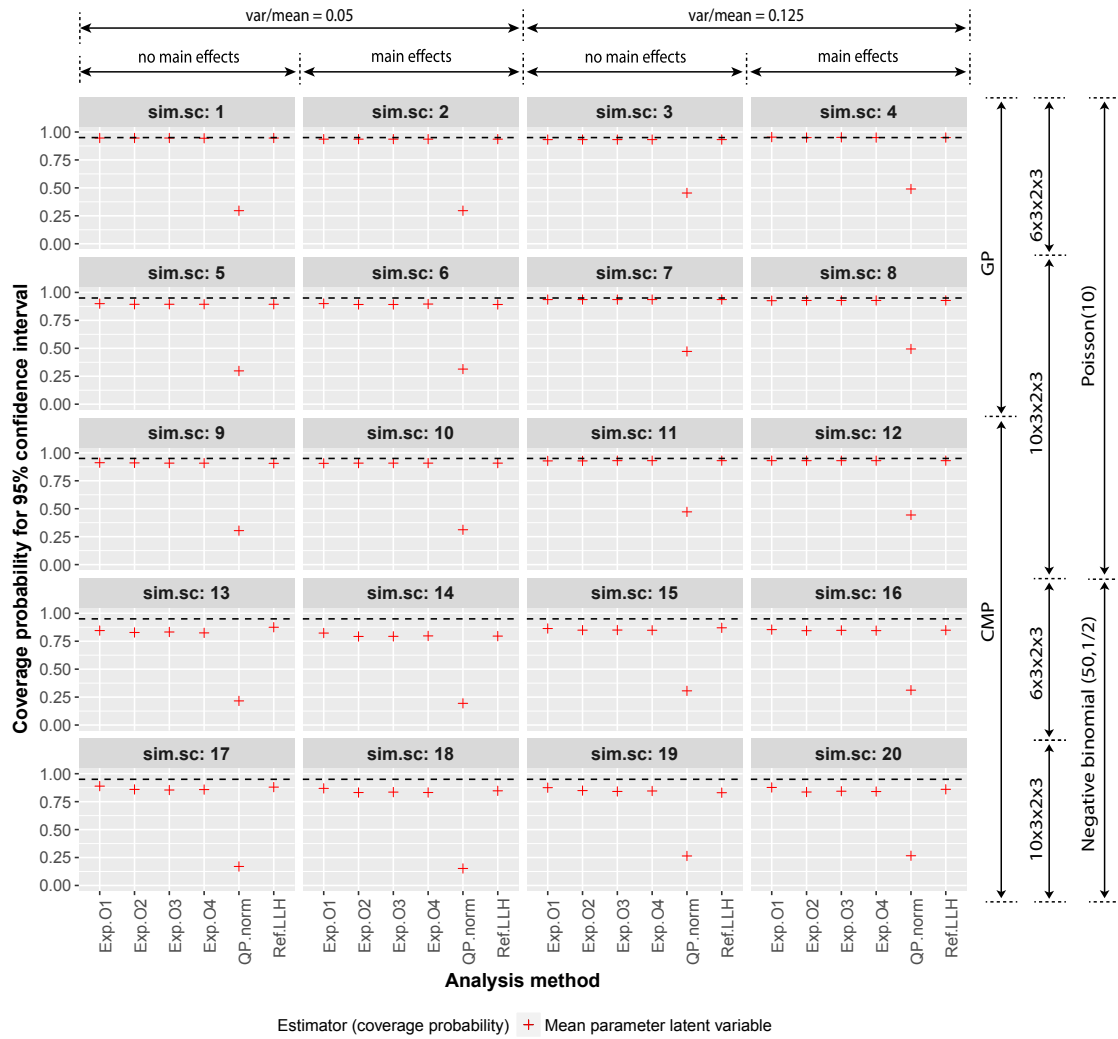
Figure 6.8: Coverage probabilities for a 95 % confidence interval of the mean parameter of the latent variable, for simulation scenario 1 to 20.
Number of simulations = 500 (only 100 for scenario 13-20 with Ref.LLH).
1) A low coverage probability was observed for the *QP.norm* approach, whereas the other methods performed well.

Asymptotic behavior of tool and operator estimates

In this paragraph, the results of some smaller simulations are presented to study the asymptotic behavior when increasing the number of parts and/or repeats. The focus is here on the estimates for the tool and operator effects, and not on the dispersion and latent variable parameter. In Figure 6.9 and Figure 6.10, all methods are analyzed for the situation where only the number of parts is increased. Given the fact that for almost all previous simulations the bias for *Exp.O2* was similar to higher order asymptotic expansions *Exp.O3*, *Exp.O4* and the reference likelihood *Ref.LLH*, we will only use *Exp.O2* in the comparison with the *Gamma, QP.norm, MOM and Pois.nu1* approaches in Figure 6.11 and 6.12, respectively.

For the latter two analyses, where both the number of parts and repeats are increased, a difference was found between using a relative contribution criterion of $\epsilon_1 = 10^{-5}$ and $\epsilon_1 = 10^{-20}$. This difference can be explained by the term $R(a)$ in $k_t := \arg\min_{a > \bar{Y}_{..}} \left((R^{(a)} - R^{(a-1)})/R^{(a-1)} < \epsilon_1 \right)$, as defined in (3.4) for the *Ref.LLH* and in (3.16) for the asymptotic expansions *Exp.O1* to *Exp.O4* and *Pois.nu1*. Whereas the term $R(a)$ scales linearly for an increasing number of parts due to the summation, it scales non-linear for increasing the repeats as a result of the product of all repeats. Therefore, in case of an increasing number of repeats, it is important to scale the relative contribution criterion.
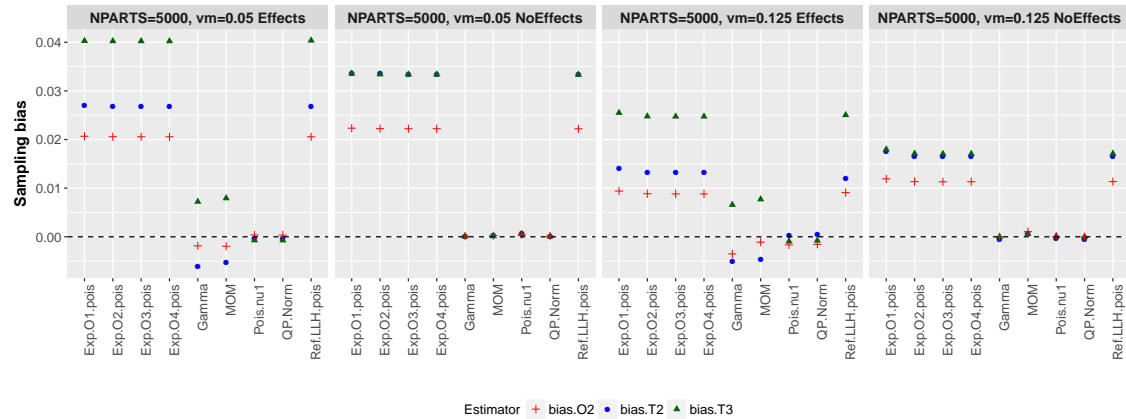
Figure 6.9: Asymptotic bias for datasets simulated from a $Pois(10)$ distribution with 5000 parts, 3 tools, 2 operators and 3 repeats. Note that vm denotes variance/mean.
*1) Whereas a small bias is observed for the Gamma and MOM approach in the presence of main effects, the bias is negligible when there are no main effects.*
2) Observe an asymptotic bias for the expansion methods (Exp.O1-Exp.O4) and the reference likelihood (Ref.LLH).
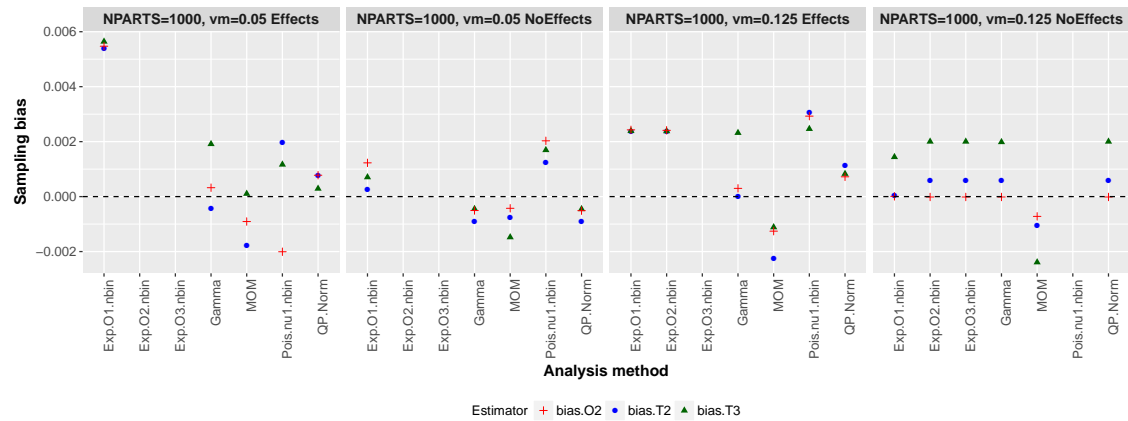
Figure 6.10: Asymptotic bias for datasets simulated from a $NBinom(50, 1/2)$ distribution with 1000 parts, 3 tools, 2 operators and 3 repeats. Note that vm denotes variance/mean.
*1) For various settings, the second order asymptotic expansion (Exp.O2) and higher (Exp.O3 and Exp.O4) did not converge.*
The observed bias seemed small for all methods and although the Gamma and QP.norm seem slightly better compared to the other methods, no clear superior method could be defined.
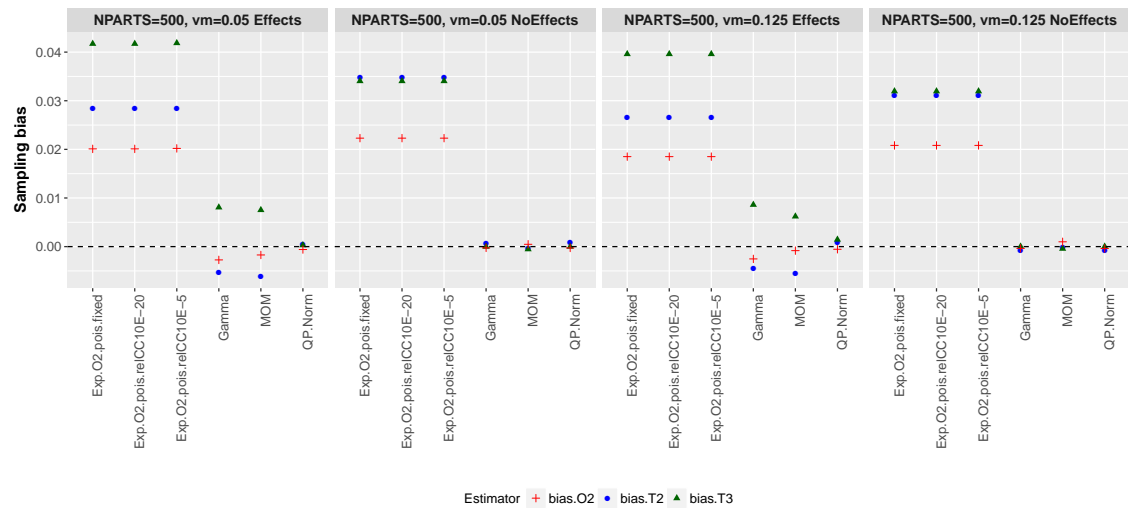
Figure 6.11: Asymptotic bias for datasets simulated from a $Pois(10)$ distribution with *500 parts*, 3 tools, 2 operators and *50 repeats* with a relative contribution criteria $\epsilon_1 = 10^{-20}$ for the summation of the latent variable. In addition for the asymptotic expansion, also $\epsilon_1 = 10^{-5}$ and a fixed truncation of the summation of the latent variable at $z = 100$ was simulated. Note that vm denotes variance/mean.
1) For Exp.O2, the relative contribution criteria resulted into a truncation of the latent variable between $z=21$ and $z=23$. This is far below the fixed truncation threshold of 100. Nevertheless, no difference in bias was observed between the asymptotic expansion (Exp.O2) for the different truncation approaches.
2) A large asymptotic bias (between 0.025 and 0.04) was observed for the Exp.O2 method.
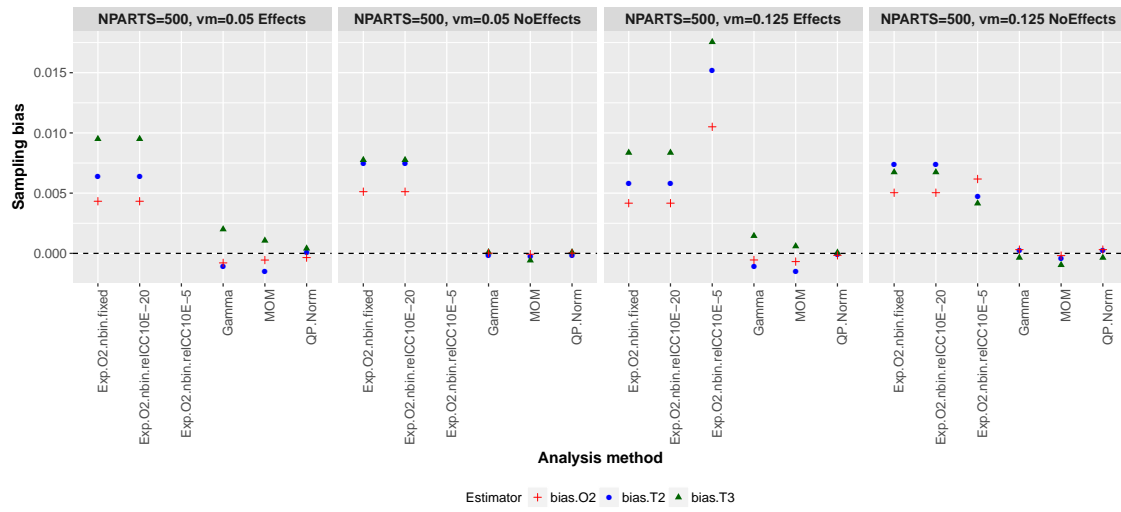
Figure 6.12: Asymptotic bias for datasets simulated from a $NBinom(50, 1/2)$ distribution with 500 parts, 3 tools, 2 operators and 50 repeats with a relative contribution criteria $\epsilon_1 = 10^{-20}$. Note that vm denotes variance/mean.

1) For the Exp.O2 approach, also a relative contribution criteria of $\epsilon_1 = 10^{-5}$ and a fixed truncation of the sum of the latent variable at $z = 500$ are applied. There is no visible difference between using a relative contribution criteria of 10^{-20} and using a fixed truncation. Moreover, in this setting a clear difference is visible between taking $\epsilon_1 = 10^{-5}$ and $\epsilon_1 = 10^{-20}$.

2) The two scenarios with a variance/mean of 0.05 did not converge for the $\epsilon_1 = 10^{-5}$.

3) A small positive bias remains present in the for the asymptotic expansion methods without a closed form.

Method of moments - F-test

In this paragraph, the performance of the F-test for assessing the tool and operator effect in case of a CMP latent variable model is presented. Since the method of moments is a computationally fast approach, we extended the number of simulations to 10,000 datasets for each simulation scenario for the purpose of this analysis. Due to the difference in degrees of freedom, the different experiment sizes and the two (F-)statistics MS_T/MS_E and MS_O/MS_E are individually evaluated.

Firstly, the empirical cumulative distribution functions (ECDF) of the F-statistics of the (odd-numbered) simulation scenarios without a tool and operator effect ($a_j = b_k = 1, \forall j, k$) are shown in Figure 6.13. These are compared with the corresponding theoretical F-distribution. In Table 6.4, the Type I errors for a test with level $\epsilon = 0.05$ and the p-values of the Kolmogorov-Smirnov (KS) goodness of fit test are shown. For the F-test on the operator mean square and error mean square ratio, the p-values of the KS test seem to contradict with the impression of a nearly perfect fit in Figure 6.13. This can be explained by high proportion of $MS_O = 0$, due to the low number of tools in combination with the response being a count.

Secondly, the EDCFs of the F-statistics of the evenly numbered simulation scenarios, that incorporate the presence of both a tool and operator effect, are presented in Figure 6.14. Here, the observed Type II error β for a level α in the simulated scenarios is the area under the ECDF curves left of the 95th quantile of the theoretical non-central F-distribution (red dotted vertical line). The observed power $(1 - \beta)$ of the hypothesis test that all tool (or operator) effects are zero is shown in Table 6.5. Moreover, the KS goodness of fit test provided p-values lower than $2 \cdot 10^{-16}$ for all simulation scenarios, indicating that under the alternative hypothesis, the observed mean square ratios do not follow a non-central F-distribution with the corresponding degrees of freedom and noncentrality parameter.

Table 6.4: The p-value of the Kolmogorov-Smirnov (KS) goodness of fit test. Observed Type I error when performing an F-test with a level of $\alpha = 0.05$. Number of simulations = 10,000.

Sim.Sc	Type I error		KS Test		Proportion	
	MS_T/MS_E	MS_O/MS_E	MS_T/MS_E	MS_O/MS_E	$MS_T = 0$	$MS_O = 0$
1	0.0541	0.0508	0.478	<0.001	0.0042	0.0561
3	0.0493	0.0551	0.411	<0.001	0.0025	0.0334
5	0.0495	0.048	0.080	<0.001	0.0026	0.0435
7	0.0466	0.0515	0.723	<0.001	0.0007	0.0268
9	0.0527	0.0493	0.048	<0.001	0.0034	0.0395
11	0.0505	0.0486	0.777	<0.001	0.0014	0.0276
13	0.0569	0.0567	0.346	<0.001	0.0015	0.027
15	0.0557	0.055	0.300	0.006	0.0004	0.0154
17	0.0494	0.0467	0.464	0.0008	0.0005	0.0198
19	0.0485	0.0494	0.593	0.112	0.0003	0.012

Table 6.5: Power of the F-test for simulation scenarios with both tool and operator effect with level $\alpha = 0.05$. *Number of simulations = 10,000.*

	Observed power	
Sim.Sc	MS_T/MS_E	MS_O/MS_E
2	1	0.9524
4	1	0.6078
6	1	0.9958
8	1	0.835
10	1	0.9941
12	1	0.8167
14	1	1
16	1	0.9929
18	1	1
20	1	0.9998
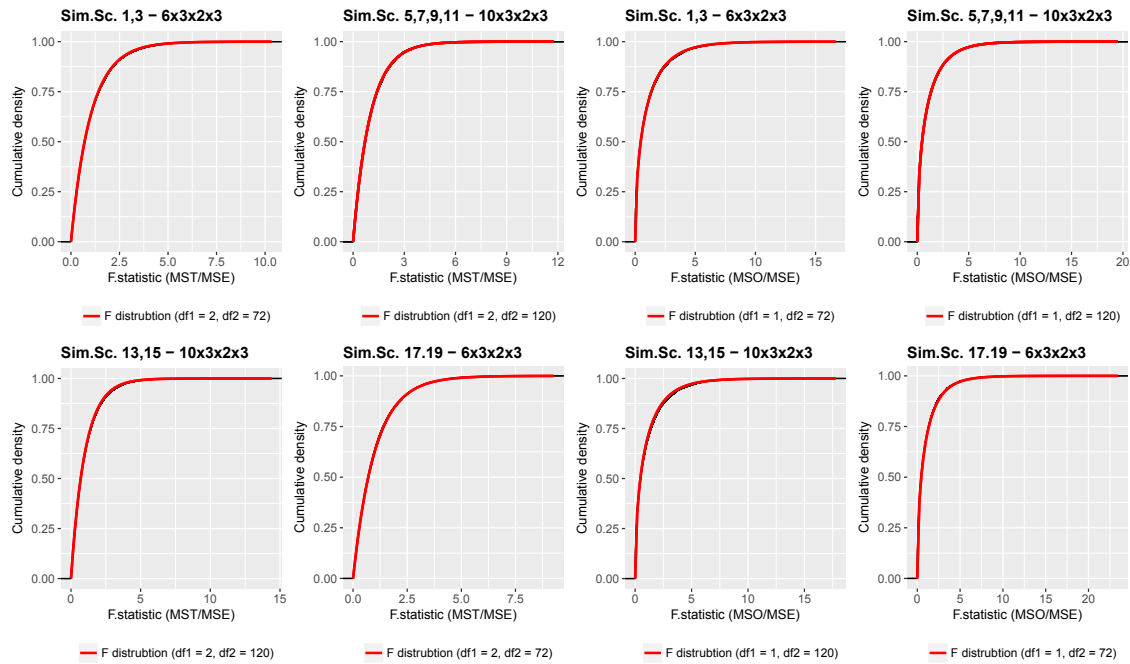
Figure 6.13: ECDF of F-statistic for simulation scenarios *without* tool and operator effects compared with the true cumulative distribution function of the F-distribution.
Number of simulations = 10,000.

Note that the simulations scenarios that have the same degrees of freedom, and therefore the same central F-distribution, are combined in a single plot. The four plots on the left describe the MS ratio MS_T/MSE and the four on the right the MS ratio MS_O/MS_E.

1) Observe that under H_0, the MS ratio MS_T/MS_E and MS_O/MS_T of the CMP latent variable model follows an F-distribution.
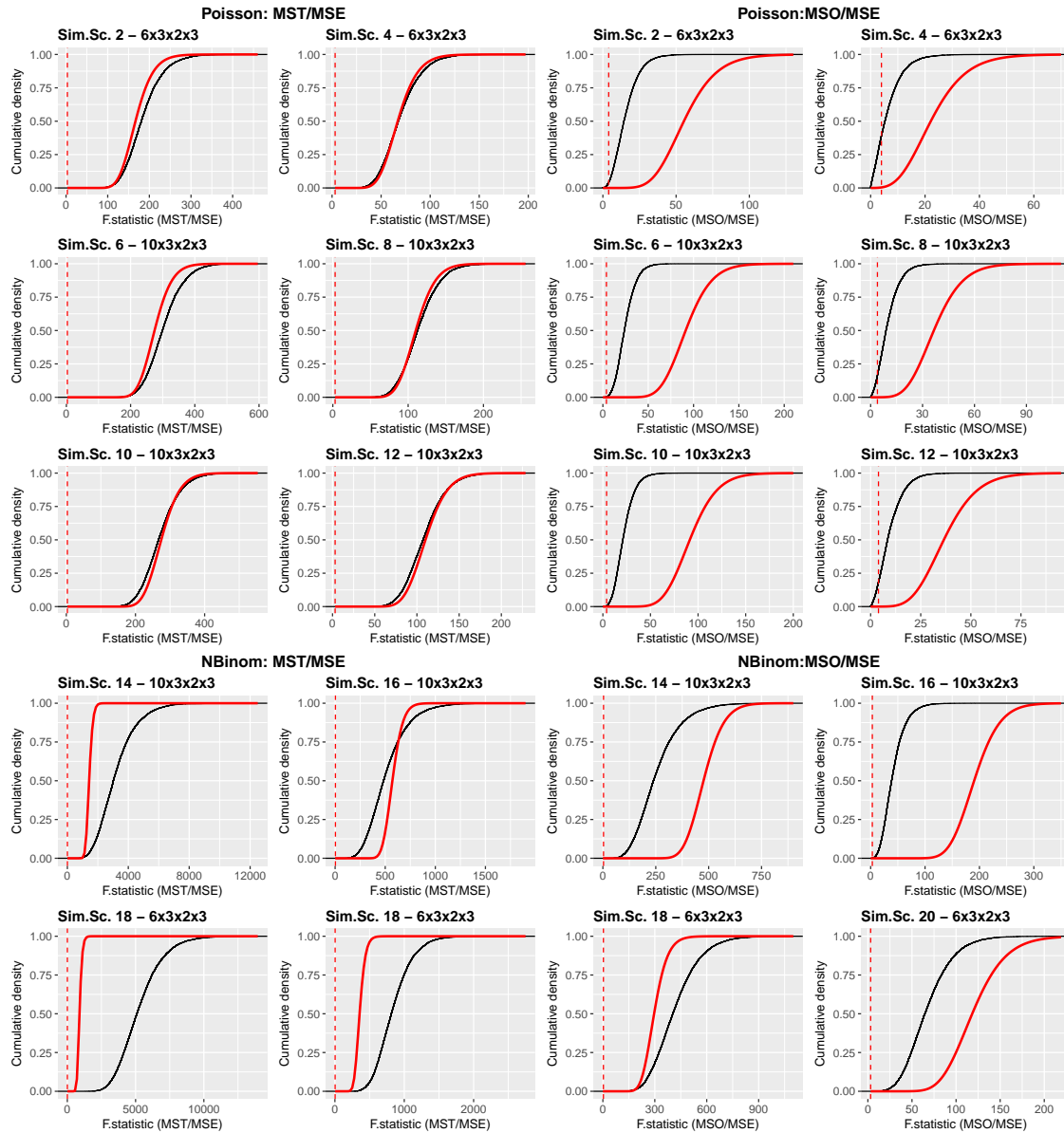
Figure 6.14: ECDF of F-statistic for simulation scenarios with both tool and operator effects compared with the true cumulative distribution function of the F-distribution
Number of simulations = 10,000.
Note that the dotted red line represents the critical value of the F-statistic is rejected.
1) Observe that under H_1 and a Poisson distribution, the MS ratio MS_T/MS_E seems to be well described by the theoretical non-central F-distribution (in red). For the ratio MS_O/MS_E this does not hold. 2) Under H_1 with a negative binomial distribution, both mean square ratios deviate significantly from the theoretical non-central F-distribution (in red).

<div align="right">

CHAPTER 7

</div>

Conclusions & Discussion

In this research we developed statistical techniques for measurement system analysis for under-dispersed count data. In practice, this situation appears in a repeated measurement setting where the measured count by some measurement system is determined by some unobserved true count of a part or process. Here, the unobserved true count is considered a random variable.

Accordingly, we focused on the parameter estimation of the Conway-Maxwell-Poisson (CMP) latent variable model. As a consequence of the presence of the latent variable in the normalization constant of the CMP, that itself has no closed form expression, no closed form was obtained for the likelihood. This results into a non-trivial estimation problem, that cannot be evaluated analytically. In order to obtain the maximum likelihood estimates of the CMP latent variable model, various estimation methods are presented. Two types of techniques can be distinguished, namely the likelihood based methods and an approach based on the method of moments. Among the pseudo-likelihood methods, we provided various algorithmic solutions, involving truncations of both the latent variable and normalization constant based on some relative contribution criteria and order n asymptotic expansions for the normalization constant. Here, the method with a double truncation is considered the reference likelihood (*Ref.LLH*), since this methods is the closest approximation of the true likelihood of the CMP latent variable model. Although, by definition of a count it would be straightforward to have a discrete latent variable, also a Gamma prior distribution is applied in combination with a first order asymptotic expansion (denoted by *Gamma)*. A closed form expression of the likelihood of the CMP latent variable model was obtained, making this method more computationally attractive compared to approaches involving a truncation. Closed form expressions of the likelihood were not found for discrete distributions. To examine whether this additional effort and complexity of analyzing the CMP latent variable model is necessary, we also considered a quasi-Poisson generalized linear mixed model (GLMM), involving normally distributed random effects.

In Section 6.1, we started with an assessment of the performance of the order n asymptotic expansion for the normalization constant, where $n = 1, \ldots, 5$. Here, it is shown that the percentage error of this asymptotic expansion of order n, compared with the true normalization constant is almost completely determined by the ratio of the CMP parameters ν/ζ, which are the dispersion and location parameter, respectively. Whereas in case of $\nu/\zeta < 5$, the relative error is small for second to fifth order expansion, the percentage error grows exponentially when this ratio increases, and grows even larger when increasing the order as well. For large ratios, a first order expansion would be more suitable.

Through an extensive simulation study, consisting of twenty simulation scenarios with varying settings that are based on a case study about counting particles on surfaces, the estimating performance of all considered estimation methods is demonstrated. Next to the reference likelihood

(*Ref.LLH*), we applied the first to fourth order asymptotic expansion or the normalization constant (*Exp.O1* to *Exp.O4*) and the Gamma prior distribution (*Gamma*). Methods that are also considered, are the quasi-Poisson GLMM with normal random effects (*QP.norm*) and a Poisson latent variable model that is obtained by fixing the dispersion of the CMP model at $\nu_{ij} = 1$.

Firstly, it was shown that for data that is less severe under-dispersed or has a higher average count (and therefore a lower ratio ν/ζ), the relative difference of the reference likelihood compared to the asymptotic expansion methods is lower. In particular, the analysis methods using second, third or fourth order asymptotic expansions, appeared to be good approximations of the more computational intensive reference likelihood. Surprisingly, these methods were inferior to the methods *Gamma, MOM, QP.norm* and even the *Pois.nu1*. With regard to the number of finished analyses of the simulated datasets, the asymptotic expansions of order two to four perform poorly (e.g. for simulation scenario 18, only 60.6% was completed for a second order asymptotic expansion). This dissatisfactory behavior can be addressed as a convergence issue, which was observed in literature before when fitting a CMP GLM model with both the full first moments as well as the first order approximations (Jowaheer & Khan, 2009).

This convergence problem could be assigned to the selected relative contribution criterion, that is used to determine where to truncate the summation of the latent variable value z, however, this would imply a bad performance of the *Pois.nu1* method as well. We investigated the effect of truncating the summation of the latent variable in the log-likelihood at an iteration-specific term k_t, where k_t is based on some relative contribution criterion ϵ_1. In case a too stringent relative contribution criterion is applied, it is expected that the same value of the truncation term is selected in consequent iterations of the likelihood optimization procedure. On the other hand, when using a too loose criterion, it is expected that the path of subsequently selected values for the truncation term is subject to more variation. From Figure C.3, it was observed that for higher counts, this path of truncation terms for successive iterations is more volatile than for the scenarios from a Poisson distribution with a lower count. In Paragraph 6.2, this effect was ascribed to the measure $(R(a) - R(a - 1))/R(a)$, which is used when assessing the relative contribution criterion. This measure scales linearly for the number of parts, but non-linear with the number of repeats. We therefore recommend a more sophisticated alignment of this relative contribution criteria based on the size of the experiment, the number of digits of the likelihood and required precision for the estimates.

In existence of convergence problems and local optima for the asymptotic expansion methods and the reference likelihood, as shown in Figure C.4, one could employ multiple starting values to increase the probability of ending up in the global instead of local optimum. For example, by using the function *multistart* of the [R] Package *'optimr'* (Nash, n.d.). Moreover, to avoid selection of boundary parameters, for example a rate of zero for a Poisson, a logarithmic barrier penalty could be implemented.

Regarding the performance measures, we considered the mean squared error, the bias and the coverage probabilities for a 95% confidence interval. For the asymptotic expansion methods without a closed form and the reference likelihood, a coverage probability of around 0.30 was observed for the tool and operator estimates in case of the scenarios with a Poisson(10) prior and var/mean ratio of 0.05. In contrast, a coverage probability of around 0.95 is observed for the *Gamma, QP.norm* method, for all simulation scenarios. For low counts the *Pois.nu1* was a bit conservative, where for higher counts it was between 0.9 and 1.

A possible explanation for the low coverage probabilities could be a somewhat poor derivation of the standard errors for the asymptotic expansion methods without closed form and the reference likelihood. Note that the standard errors of the maximum likelihood estimates are based on the observed Fisher Information, which is obtained by the Hessian of negative log-likelihood. This derivation with the Hessian is more volatile than the theoretical Fisher Information derived from a probability model. Nevertheless, the standard errors are unlikely to be the root cause of the low coverage probabilities, since the *Gamma* and *Pois.nu1* both use the same theory, but both have high coverage probabilities. For the coverage probabilities of the moment estimators, it should be noted that the standard errors were defined in a univariate setting by means of a Taylor expansion

and therefore did not include any covariances between the moments.

With respect to estimation of the dispersion parameter, a low coverage probability for the *Gamma* method was a result of a significant bias. This occurred in the setting with low counts and severe under-dispersion. In absence of the standard errors for dispersion parameter of the quasi-Poisson, no statements have been made about the coverage probability of this method. For the mean parameter of the latent variable and the simulation scenario with a negative binomial latent variable, a large bias was observed for the *QP.norm*. The other methods, and the *Gamma* in particular, perform really well under a more skewed latent variable.

Furthermore, an hypothesis test on the effect of tools and operators is constructed based on the ratio of the tool (or operator) mean squares with the mean square error, and evaluated according to an F-distribution. Under the null-hypothesis, it was shown that the MS_T/MS_E and $MS_O/_M S_E$ followed an F-distribution with the degrees of freedom corresponding to the mean squares. Under the alternative hypothesis, the ratio MS_T/MS_E was only well described by a non-central F-distribution for the main effects scenarios with a Poisson distribution. In contrast, a poor fit was found in scenarios with a negative binomial latent variable, but also for the ratio MS_O/MS_E for all datasets. Since the assumption of an F-test is normality of the data, the Poisson distribution is expected to have a better fit than the negative binomial distribution, since for increasing rate, the Poisson can be better approximated with a normal distribution.

Summarizing, the quasi-Poisson generalized linear mixed model and the CMP with a Gamma distributed latent variable are considered the superior methods for estimating the parameters of the CMP latent variable model for severely under-dispersed count data.

Further research

This thesis mainly focused on the Conway-Maxwell-Poisson as conditional distribution to model repeated measurements. Moreover, the Generalized Poisson distribution is used for the simulation of datasets. A distribution that was not considered is the hyper Poisson (hP) distribution (Bardwell & Crow, 1964). In contrast to the CMP distribution, the expectation of the hP distribution can be directly expressed in terms of its parameters. Note that the hP distribution involves the Kummer confluent hypergeometric function (KCHF) (Georgiev & Georgieva-Grosse, 2005). However, creating a model in which the KCHF contains the latent variable Z_i will result in an equally complex estimation problem as for the CMP, where the normalization constant involves the latent variable. Additional research should be performed on the effectiveness of the hP distribution. An analogue methodology could be applied, where an asymptotic expansion of the KCHF is used (Georgiev & Georgieva-Grosse, 2005) to derive a pseudo-likelihood.

The current mathematical model, the effect of tools and operators are considered to be fixed. Further research should be performed to extend this developed framework to the random effects settings, which often appears in practice. Moreover, to keep the size of the simulation study manageable, we considered two settings for the effect of the dispersion, distribution of the latent variable, effect size and experimental size. More in depth simulation studies could be performed focusing on the effect of number of repeats, more severe under-dispersion levels, as well as inclusion of interaction effects. In addition, whereas we used a relative contribution criterion for the truncation of both the normalization constant and the summation of the latent variable in the likelihood, also a fixed truncation could be applied. For this situation, it should be investigated at which term this truncation should occur in order to obtain stable estimates.

Although we simulated some large experiments next to the main simulation study, to get a sense of the asymptotic behavior, we mainly focused on the finite setting. Therefore, a thorough theoretical analysis of the asymptotic behavior of the likelihoods and moment estimators is needed, when increasing the number of parts and repeats per unique measurement setting. Here, it remains to be verified whether the moment estimators of the CMP latent variable model using a first order approximation of the normalization constant are asymptotically normal.

Recommendations for ASML

With regard to the measurement system analysis on microscopes that are used for counting the defects on a surface, some key recommendations will be presented. First of all, in case of some underlying distribution of the true counts on a part or in process, and a repeated measurement setting that severely is under-dispersed, the use of a quasi-Poisson GLMM with normally distributed random effects or a Conway-Maxwell-Poisson repeated measurement distribution with a Gamma latent variable is advised. Although, in case the distribution of the latent variable is somewhat skewed, the CMP model with a Gamma prior is more appropriate. A practical advantage of the quasi-Poisson GLMM with normal random effects is that this method easily follows from the Poisson GLMM, which is a standard implementation in the statistical software [R], as described in Paragraph 3.1.

In addition, this study showed that inference based on a lower counts will result into a larger bias and lower coverage probability. Therefore, in the setting of measuring defects of a certain diameter size, it is recommended to use a smaller lower limit of the diameter size, to increase the number of defects. In this case, the effects of tool and operators can be better discriminated. If this is not possible the experiment size could be increased.

Moreover, the robustness of the F-test for a CMP latent variable model was assessed. In particular, under the null hypothesis that there are no tool (or operator effects) it seems appropriate to evaluate the ratio of mean square tool and the mean square error with a central F-distribution. The power of this test is defined by the effect size and the sample size. In this analysis it was found that for a low number of degrees of freedom, it was harder to detect variation in the mean squares. Therefore, reducing the number of repeats to two, as a measure to scale down the experiment size, is discouraged in case of count data.

Bibliography

(AIAG), A. I. A. G. (2010). *Measurement systems analysis reference manual.* AIAG Southfield (MI). 1, 3

Bardwell, G. E. & Crow, E. L. (1964). A two-parameter family of hyper-poisson distributions. *Journal of the American Statistical Association*, *59*(305), 133–141. 55

Basem, E.-H. (2008). *Design for six sigma: A roadmap for product development.* McGraw-Hill Publishing. 1

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01 18

Consul, P. & Shoukri, M. (1985). The generalized poisson distribution when the sample mean is larger than the sample variance. *Communications in Statistics-Simulation and Computation*, *14*(3), 667–681. 8, 9

Consul, P. C. & Famoye, F. (2006). *Lagrangian probability distributions.* Springer Science & Business Media. 8

Consul, P. C. & Jain, G. C. (1973). A generalization of the poisson distribution. *Technometrics*, *15*(4), 791–799. 5, 8

Conway, R. W. & Maxwell, W. L. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, *12*(2), 132–136. 5

Czado, C., Erhardt, V., Min, A. & Wagner, S. (2007). Zero-inflated generalized poisson models with regression effects on the mean, dispersion and zero-inflation level applied to patent outsourcing rates. *Statistical Modelling*, *7*(2), 125–153. 5

Dobson, A. J. & Barnett, A. (2008). *An introduction to generalized linear models.* CRC press. 19

Durrett, R. (2005). *Probability: theory and examples.* Cambridge university press. 67

Famoye, F. (1993). Restricted generalized poisson regression model. *Communications in Statistics-Theory and Methods*, *22*(5), 1335–1354. 5

Gaunt, R. E., Iyengar, S., Daalhuis, A. B. O. & Simsek, B. (2016). An asymptotic expansion for the normalizing constant of the conway-maxwell-poisson distribution. *arXiv preprint arXiv:1612.06618*. 7, 8, 85, 86

Georgiev, G. N. & Georgieva-Grosse, M. N. (2005). The kummer confluent hypergeometric function and some of its applications in the theory of azimuthally magnetized circular ferrite waveguides. *Journal of telecommunications and information technology*, 112–128. 55, 67

Gillispie, S. B. & Green, C. G. (2015). Approximating the conway–maxwell–poisson distribution normalization constant. *Statistics*, *49*(5), 1062–1073. 7

Guikema, S. D. & Goffelt, J. P. (2008). A flexible count data regression model for risk analysis. *Risk analysis*, *28*(1), 213–223. 6

Huang, A. (2016). Mean-parametrized conway–maxwell–poisson regression models for dispersed counts. *Statistical Modelling*, 1471082X17697749. 5, 11

Jowaheer, V. & Khan, N. M. (2009). Estimating regression effects in com-poisson generalized linear model. *International Journal of Computational and Mathematical Sciences*, *1*(2), 59–63. 6, 54

Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 619–678. 3

Li, M.-H. C. & Al-Refaie, A. (2008). Improving wooden parts' quality by adopting dmaic procedure. *Quality and Reliability Engineering International*, *24*(3), 351–360. 1

Mächler, M. (n.d.). *Arbitrarily accurate computation with r: The rmpfr package.* 75

Maechler, M. (2016). Rmpfr: R mpfr - multiple precision floating-point reliable [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=Rmpfr` (R package version 0.6-1) 33

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models, no. 37 in monograph on statistics and applied probability.* Chapman & Hall,. 3

Minka, T. P., Shmueli, G., Kadane, J. B., Borle, S. & Boatwright, P. (2003). Computing with the com-poisson distribution. *CMU Statistics Department.* 7, 85

Nash, J. C. (n.d.). optimr: A replacement and extension of the 'optim' function [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=optimr` 54

Osma, A. (2011). An assessment of the robustness of gauge repeatability and reproducibility analysis in automotive components. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, *225*(7), 895–912. 3

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/` 33

Sellers, K. F., Borle, S. & Shmueli, G. (2012). The com-poisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry*, *28*(2), 104–116. 5

Sellers, K. F., Lotze, T. & Raim, A. (2017). Compoissonreg: Conway-maxwell poisson (compoisson) regression [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=COMPoissonReg` (R package version 0.4.1) 33, 34

Sellers, K. F. & Raim, A. (2016). A flexible zero-inflated model to address data dispersion. *Computational Statistics & Data Analysis*, *99*, 68–80. 6

Sellers, K. F. & Shmueli, G. (2010). A flexible regression model for count data. *The Annals of Applied Statistics*, 943–961. 6

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S. & Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the conway–maxwell–poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(1), 127–142. 5, 6, 7, 85

Wald, A. & Wolfowitz, J. (1939). Confidence limits for continuous distribution functions. *The Annals of Mathematical Statistics*, *10*(2), 105–118. 24

Yee, T. W. (2017). VGAM: Vector generalized linear and additive models [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=VGAM` (R package version 1.0-4) 33

Zamani, H. & Ismail, N. (2012). Functional form for the generalized poisson regression model. *Communications in Statistics-Theory and Methods*, *41*(20), 3666–3675. 5, 9

Notation index

$\text{Pois}(\theta)$	the Poisson distribution with rate θ
$\text{Pois}(\theta, \delta)$	the quasi-Poisson distribution with rate θ and dispersion parameter δ
$\text{CMP}(\zeta, \nu)$	the Poisson distribution with rate θ
$\text{GP}(\tau, \varphi)$	the generalized Poisson distribution with parameters τ and φ
$\text{NBinom}(\theta, \kappa)$	the Negative binomial distribution with rate θ and size κ
$\mathcal{N}(\mu, \sigma^2)$	the Normal distribution with mean μ and variance σ^2
$\text{Gamma}(\kappa, \lambda)$	the Gamma distribution with shape κ and scale λ
$\text{Unif}(a, b)$	the continuous uniform distribution on the interval $[a, b]$
MS_E	the mean square error
MS_T	the tool mean squares
MS_O	the operator mean squares
MS_P	the part mean squares
a.u.	arbitrary unit
p.m.f.	probability mass function
p.d.f.	probability density function
c.d.f.	cumulative density function
i.i.d.	independently and identically distributed
L.V.	latent variable

List of Figures

List of Tables

Proofs of theorems & lemmas

A.1   Maximum Likelihood Estimation

Proof of Lemma 3.1.1. Assume an order n expansion of the normalization constant (3.8) and take any distribution for the latent variable Z_i, then the corresponding pseudo-log-likelihood (3.1) changes to

$$
\begin{aligned}
l^* &= \sum_{i=1}^{I} \log \left(\int_{z=0}^{\infty} \prod_{j=1}^{J} \frac{1}{\widetilde{S}^{(n)}(\zeta_{ij}(z), \nu_{ij})} \left(\frac{\zeta_{ij}(z)^{y_{ij}}}{y_{ij}!} \right)^{\nu_{ij}} dF_{Z_i}(z)dz \right) \\
&= -A + B + \sum_{i=1}^{I} \log \left(\int_{z=0}^{\infty} z^{C_i} \prod_{j=1}^{J} \frac{1}{\widetilde{S}^{(n)}(\zeta_{ij}(z), \nu_{ij})} dF_{Z_i}(z)dz \right) \\
&= -A + B + \sum_{i=1}^{I} \log \left(\int_{z=0}^{\infty} z^{C_i} \prod_{j=1}^{J} \left(\frac{(2\pi)^{(\nu_{ij}-1)/2} \sqrt{\nu_{ij}} \cdot e^{-\nu_{ij}\zeta_{ij}(z)}}{\zeta_{ij}(z)^{(1-\nu_{ij})/2} \cdot \sum_{k=0}^{n} c_k(\nu_{ij}\zeta_{ij}(z))^{-k}} \right) dF_{Z_i}(z)dz \right) \\
&= -A + B + D + \sum_{i=1}^{I} \log \left(\int_{z=0}^{\infty} z^{C_i} \prod_{j=1}^{J} \left(\frac{e^{-\nu_{ij}\zeta_{ij}(z)} \cdot \zeta_{ij}(z)^{(\nu_{ij}-1)/2}}{\sum_{k=0}^{n} c_k(\nu_{ij}\zeta_{ij}(z))^{-k}} \right) dF_{Z_i}(z)dz \right) \\
&= -A + B + D + E + \sum_{i=1}^{I} \log \left(\int_{z=0}^{\infty} z^{C_i} \prod_{j=1}^{J} \left(\frac{e^{-\nu_{ij}ze^{\boldsymbol{x}_{ij}^T\boldsymbol{\beta}}} z^{(\nu_{ij}-1)/2}}{\sum_{k=0}^{n} c_k(\nu_{ij}\zeta_{ij}(z))^{-k}} \right) dF_{Z_i}(z)dz \right) \\
&= -A + B + D + E + \int_{i=1}^{I} \log \left(\sum_{z=0}^{\infty} z^{C_i+F_i} e^{-zG_i} \cdot \prod_{j=1}^{J} \left(\frac{1}{\sum_{k=0}^{n} c_k(\nu_{ij}\zeta_{ij}(z))^{-k}} \right) dF_{Z_i}(z)dz \right)
\end{aligned}
$$
$$\tag{A.1}$$

where

$$
A = \sum_{i=1}^{I} \sum_{j=1}^{J} \nu_{ij} \log(y_{ij}!), \qquad B = \sum_{i=1}^{I} \sum_{j=1}^{J} \boldsymbol{x}_{ij}^T \boldsymbol{\beta} y_{ij} \nu_{ij}, \qquad C_i = \sum_{j=1}^{J} y_{ij} \nu_{ij},
$$

$$
D = \sum_{i=1}^{I} \sum_{j=1}^{J} \log((2\pi)^{(\nu_{ij}-1)/2} \sqrt{\nu_{ij}}), \qquad E = \sum_{i=1}^{I} \sum_{j=1}^{J} \boldsymbol{x}_{ij}^T \boldsymbol{\beta} \frac{(\nu_{ij}-1)}{2}, \qquad F_i = \sum_{j=1}^{J} \frac{\nu_{ij}-1}{2}, \quad \text{(A.2)}
$$

$$
G_i = \sum_{j=1}^{J} \nu_{ij} e^{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}}.
$$

This form can be rewritten by merging the terms A, B, D, E and C_i, F_i such that

$$l^* = A^\dagger + \sum_{i=1}^{I} \log \left(\int_{z=0}^{\infty} z^{B_i^\dagger} e^{-zC_i^\dagger} \cdot \prod_{j=1}^{J} \left(\frac{1}{\sum_{k=0}^{n} c_k (\nu_{ij} \zeta_{ij}(z))^{-k}} \right) dF_{Z_i}(z) dz \right) \tag{A.3}$$

where

$$A^\dagger = \sum_{i=1}^{I} \sum_{j=1}^{J} -\nu_{ij} \log(y_{ij}!) + \boldsymbol{x}_{ij}^T \boldsymbol{\beta} y_{ij} \nu_{ij} + \frac{\nu_{ij} - 1}{2} \log(2\pi) + \frac{1}{2} \log(\nu_{ij}) + \boldsymbol{x}_{ij}^T \boldsymbol{\beta} \frac{(\nu_{ij} - 1)}{2}$$

$$B_i^\dagger = \sum_{j=1}^{J} y_{ij} \nu_{ij} + \frac{\nu_{ij} - 1}{2} \qquad \text{and} \qquad C_i^\dagger = \sum_{j=1}^{J} \nu_{ij} e^{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}}. \tag{A.4}$$

$\square$

Convergence proofs

Proof of Proposition 3.1.1. Claim: The normalization constant $\sum_{s=0}^{\infty} \left(\frac{z^s}{s!} \right)^n$, which is an infinite series, is a convergent when $1 \le \nu < \infty$ and $0 \le z < \infty$.

Let $1 \le \nu < \infty$ and $z < \infty$. Then by Stirling's approximation we have the bound $s! \ge \sqrt{2\pi s} s^s \exp\{-s\}$. Therefore, we can derive the upper bound for normalization constant

$$\sum_{s=0}^{\infty} \left(\frac{z^s}{s!} \right)^{\nu} \le 1 + \sum_{s=1}^{\infty} (2\pi)^{-\frac{\nu}{2}} \left(\frac{\left(\frac{ez}{s} \right)^s}{\sqrt{s}} \right)^{\nu}$$

$$\le 1 + \sum_{s=1}^{\infty} (2\pi)^{-\frac{\nu}{2}} \left(\left(\frac{ez}{s} \right)^s \right)^{\nu} \tag{A.5}$$

$$\le 1 + \sum_{s=1}^{\infty} \left(\frac{(ez)^{\nu}}{s} \right)^s$$

Let $k = \arg\min_a \{(ez)^{\nu} < a)\}$, we have that

$$1 + \sum_{s=1}^{\infty} \left(\frac{(ez)^{\nu}}{s} \right)^s = 1 + \sum_{s=1}^{k} \left(\left(\frac{(ez)^{\nu}}{s} \right)^s \right) + \sum_{s=k+1}^{\infty} \left(\left(\frac{(ez)^{\nu}}{s} \right)^s \right)$$

$$\le 1 + \sum_{s=1}^{k} \left(\frac{(ez)^{\nu}}{s} \right)^s + \sum_{s=k+1}^{\infty} \left(\frac{(ez)^{\nu}}{k+1} \right)^s \tag{A.6}$$

$$\le 1 + \sum_{s=1}^{k} \left(\frac{(ez)^{\nu}}{s} \right)^s + \frac{1}{1 - \frac{k+1}{(ez)^{\nu}}}$$

Since $1 \le \nu < \infty$ and $0 \le z < \infty$, we have that $k < \infty$, thus

$$\sum_{s=0}^{\infty} \left(\frac{z^s}{s!} \right)^{\nu} \le \sum_{s=0}^{k} \left(\frac{(ez)^{\nu}}{s} \right)^s + \frac{1}{1 - \frac{k+1}{ez}} \le \infty. \tag{A.7}$$

Since $\sum_{s=0}^{\infty} \left(\frac{z^s}{s!} \right)^{\nu}$ is bounded and monotonic increasing, it is a convergent series. $\square$

Proof of Proposition 3.1.2. Let us first rewrite the likelihood (3.9) as

$$l^* = A + \sum_{i=1}^{I} \log \left(\int_{z=0}^{\infty} Q(z)dz \right) \tag{A.8}$$

with

$$Q(z) = z^{B_i} e^{-zC_i} \frac{1}{\prod_{j=1}^{J} \left(\sum_{k=0}^{n} c_k (\nu_{ij} \zeta_{ij}(z))^{-k} \right)} dF_{Z_i}(z). \tag{A.9}$$

Moreover, the terms c_k are determined by (2.9) and A, B_i, C_i are defined by

$$A = \sum_{i=1}^{I} \sum_{j=1}^{J} -\nu_{ij} \log(y_{ij}!) + \boldsymbol{x}_{ij}^T \boldsymbol{\beta} y_{ij} \nu_{ij} + \frac{\nu_{ij} - 1}{2} \log(2\pi) + \frac{1}{2} \log(\nu_{ij}) + \boldsymbol{x}_{ij}^T \boldsymbol{\beta} \frac{(\nu_{ij} - 1)}{2}$$

$$B_i = \sum_{j=1}^{J} y_{ij} \nu_{ij} + \frac{\nu_{ij} - 1}{2} \qquad \text{and} \qquad C_i = \sum_{j=1}^{J} \nu_{ij} e^{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}}. \tag{A.10}$$

Note that in case of (finite) under-dispersion ($1 \leq \nu < \infty$) that A, B_i and C_i are non-negative and finite for all i. Simultaneously, we have that $dF_{Z_i}(z)$ is a probability density function and therefore bounded between 0 and 1. For a first order expansion of the normalization constant ($n = 0$), we now have that

$$z^{B_i} e^{-zC_i} dF_{Z_i}(z) \searrow 0, \qquad \text{if } z \to \infty. \tag{A.11}$$

since the rate of convergence of the exponent e^{-zC_i} to 0 is larger than the rate of divergence of z^{B_i}. Now, except for the situation that dataset only contains zeros or that $dF_{Z_i}(z)$ is a point mass in zero, we have that the log-likelihood l^* is finite and exists.

For the situation that $n > 0$ we obtain as well that

$$Q(z) \searrow 0, \qquad \text{if } z \to \infty \tag{A.12}$$

since again the rate of convergence of the exponent e^{-zC_i} to 0 is larger than the rate of divergence of z^{B_i} and the fact that $\sum_{k=0}^{n} c_k (\nu_{ij} \zeta_{ij}(z))^{-k}$ is a finite constant for fixed k and for all $j = 1, \ldots, J$. The latter follows from the fact that the constants c_k are finite for $1 \leq \nu < \infty$, and therefore $1/(\nu_{ij} \zeta_{ij}(z))^k \to 0$ if $\zeta_{ij}(z) \to \infty$. If $z \to \infty$, we have that $\zeta_{ij}(z) = z \exp\{\boldsymbol{x_{ij}^T \beta}\} \to \infty$, since $\boldsymbol{\beta}$ is not allowed to be $-\infty$.

Moreover, it can be concluded that $Q(0) = 0$ and $0 < Q(a) < \infty$. For every probability distribution with a (sub-)domain of $z \geq 0$, (except $dF_{Z_i}(z = 0) = 1$), we have that l^* is bounded and therefore absolute convergent. $\square$

Derivation closed forms pseudo-likelihoods

Proof of Proposition 3.1.3 (Gamma prior). Given a first order approximation of the normalization constant and a Gamma distributed prior Z_i, with $Gamma(\kappa, \lambda)$, we have the pseudo-log-

likelihood (3.9) from Lemma 3.1.1 obtains the form

$$
\begin{aligned}
l^* &= A + \sum_{i=1}^{I} \log\left(\int_0^\infty z^{B_i} \exp\{-zC_i\} \frac{\left(\lambda^{-\kappa} z^{\kappa-1} \exp\{-\frac{z}{\lambda}\}\right)}{\Gamma(\kappa)}\, dz \right) \\
&= A + \sum_{i=1}^{I} \log\left(\frac{\Gamma(\kappa+B_i)\lambda^{-\kappa}}{\Gamma(\kappa)} \int_0^\infty z^{-1+(\kappa+B_i)} \exp\{-z\left(C_i+\frac{1}{\lambda}\right)\} \frac{1}{\Gamma(\kappa+B_i)}\, dz \right) \\
&= A + \sum_{i=1}^{I} \log\left(\frac{\Gamma(\kappa+B_i)\lambda^{-\kappa+B_i}\left(\frac{\lambda}{C_i\lambda+1}\right)^{\kappa}}{\Gamma(\kappa)} \int_0^\infty z^{-1+(\kappa+B_i)} \exp\{\frac{-z}{\left(\frac{\lambda}{C_i\lambda+1}\right)}\} \frac{\left(\frac{\lambda}{C_i\lambda+1}\right)^{-(\kappa+B_i)}}{\Gamma(\kappa+B_i)}\, dz \right) \\
&= A + \sum_{i=1}^{I} \log\left(\frac{\Gamma(\kappa+B_i)\lambda^{B_i}\left(C_i\lambda+1\right)^{-\kappa}}{\Gamma(\kappa)} \int_0^\infty z^{-1+(\kappa+B_i)} \exp\{\frac{-z}{\left(\frac{\lambda}{C_i\lambda+1}\right)}\} \frac{\left(\frac{\lambda}{C_i\lambda+1}\right)^{-\kappa}}{\Gamma(\kappa+B_i)}\, dz \right) \\
&= A + \sum_{i=1}^{I} \log\left(\frac{\Gamma(\kappa+B_i)\lambda^{B_i}\left(C_i\lambda+1\right)^{-(\kappa+B_i)}}{\Gamma(\kappa)} \right) \\
&= A + \sum_{i=1}^{I} B_i \log(\lambda) - (B_i+\kappa)\log(1+C_i\lambda) + \log(\Gamma(B_i+\kappa)) - \log(\Gamma(\kappa)),
\end{aligned}
$$

where A, B_i, C_i are defined by

$$
A = \sum_{i=1}^{I}\sum_{j=1}^{J} -\nu_{ij}\log(y_{ij}!) + \boldsymbol{x}_{ij}^T\boldsymbol{\beta} y_{ij}\nu_{ij} + \frac{\nu_{ij}-1}{2}\log(2\pi) + \frac{1}{2}\log(\nu_{ij}) + \boldsymbol{x}_{ij}^T\boldsymbol{\beta}\frac{(\nu_{ij}-1)}{2}
$$

$$
B_i = \sum_{j=1}^{J} y_{ij}\nu_{ij} + \frac{\nu_{ij}-1}{2} \qquad \text{and} \qquad C_i = \sum_{j=1}^{J}\nu_{ij}e^{\boldsymbol{x}_{ij}^T\boldsymbol{\beta}}.
$$

(A.13)

$\square$

Closed form expressions of the likelihood with different prior distributions

Other continuous prior distribution that resulted into closed form expressions, but were not included in the simulation study are the Inverse Gamma and the Normal distribution, see Proposition A.1.1 and A.1.2. The results were obtained using Wolfram Mathematica, however, a validating proof of this result remains to be provided.

Remark A.1.1 (Inverse Gamma prior). *Let us assume the CMP latent variable model (2.12), with a first order approximation of the normalization constant and an inverse Gamma latent variable, $Z_i \sim InvGamma(\kappa, \lambda)$. Then using Wolfram Mathematica, the pseudo-log-likelihood (3.9) from Lemma 3.1.1 can be simplified to*

$$
\begin{aligned}
l^* &= A + \sum_{i=1}^{I} \log\left(\int_0^\infty z^{B_i} \exp\{-zC_i\} \frac{\left(\frac{\beta}{z}\right)^{\kappa}\exp\{-\frac{\beta}{z}\}}{z\Gamma(\kappa)}\, dz \right) \\
l^* &= A + \sum_{i=1}^{I} \log\left(\int_0^\infty z^{B_i} \exp\{-zC_i - \frac{\beta}{z}\} \frac{\left(\frac{\beta}{z}\right)^{\kappa}}{z\Gamma(\kappa)}\, dz \right) \\
l^* &= A + \sum_{i=1}^{I} \log\left(\frac{BesselK\left(-B_i+\kappa, 2\sqrt{C_i\beta}\right)}{2B^{\frac{1}{2}(B_i-\kappa)}\beta^{\frac{1}{2}(B_i+\kappa)}\Gamma(\kappa)} \right),
\end{aligned}
$$

with A, B_i and C_i defined in (A.2) and BesselK the modified Bessel function of the second kind.

Remark A.1.2 (Form for normally distributed prior)**.** *Let us assume the CMP latent variable model* (2.12), *with a first order approximation of the normalization constant and a normally distributed latent variable,* $\mathcal{N}(0, \sigma^2)$. *Now, for* $\nu_{ij} \geq 1$, *we obtained by using Wolfram Mathematica that the pseudo-log-likelihood* (3.12) *changes to*

$$l^* = A + \sum_{i=1}^{I} \log\left(\frac{2^{\frac{B_i}{2}-1}\sigma^{B_i}e^{-\frac{\mu^2}{2\sigma^2}}}{\sqrt{\pi}} \left(\Gamma\left(\frac{A+1}{2}\right) {}_1F_1\left(\frac{B_i+1}{2}; \frac{1}{2}; \frac{\left(\mu - C_i\sigma^2\right)^2}{2\sigma^2}\right) + \right. \right.$$

$$\left. \left. \frac{\sqrt{2}}{\sigma}\Gamma\left(\frac{B_i}{2}+1\right)\left(\mu - C_i\sigma^2\right) {}_1F_1\left(\frac{B_i+2}{2}; \frac{3}{2}; \frac{\left(\mu - C_i\sigma^2\right)^2}{2\sigma^2}\right) \right) \right)$$

$$= A + \sum_{i=1}^{I} -\frac{\log(\pi)}{2} + (\frac{B_i}{2}-1)\log(2) + B_i\log(\sigma)(-\frac{\mu^2}{2\sigma^2}) + \Gamma\left(\frac{A+1}{2}\right) + \tag{A.14}$$

$$\sum_{i=1}^{I} \log\left({}_1F_1\left(\frac{B_i+1}{2}; \frac{1}{2}; \frac{\left(\mu - C_i\sigma^2\right)^2}{2\sigma^2}\right) + \right.$$

$$\left. \frac{\sqrt{2}}{\sigma}\Gamma\left(\frac{B_i}{2}+1\right)\left(\mu - C_i\sigma^2\right) {}_1F_1\left(\frac{B_i+2}{2}; \frac{3}{2}; \frac{\left(\mu - C_i\sigma^2\right)^2}{2\sigma^2}\right) \right),$$

where

$$A = \sum_{i=1}^{I}\sum_{j=1}^{J} -\nu_{ij}\log(y_{ij}!) + \boldsymbol{x}_{ij}^T\boldsymbol{\beta}y_{ij}\nu_{ij} + \frac{\nu_{ij}-1}{2}\log(2\pi) + \frac{1}{2}\log(\nu_{ij}) + \boldsymbol{x}_{ij}^T\boldsymbol{\beta}\frac{(\nu_{ij}-1)}{2}$$

$$B_i = \sum_{j=1}^{J} y_{ij}\nu_{ij} + \frac{\nu_{ij}-1}{2} \qquad and \qquad C_i = \sum_{j=1}^{J} \nu_{ij}e^{\boldsymbol{x}_{ij}^T\boldsymbol{\beta}}$$

and ${}_1F_1$ *is the confluent hypergeometric series* Georgiev & Georgieva-Grosse (2005). *This function is an absolutely convergent infinite power series and described by*

$${}_1F_1(a; c; x) = \sum_{\nu=0}^{\infty} \frac{\Gamma(a+\nu)\Gamma(c)}{\Gamma(a)\Gamma(c+\nu)} \frac{(a)_\nu}{(c)_\nu} \frac{x^\nu}{\nu!}.$$

A.2   Method of Moments

Variances and expected mean squares

Lemma A.2.1 (Law of total variance Durrett (2005))**.** *Let X and Y be random variables that are defined on the same probability space, with finite variance of Y, then it holds that*

$$Var(Y) = Var(\mathbb{E}[Y|X]) + \mathbb{E}[Var(Y|X)]. \tag{A.15}$$

Proposition A.2.1 (Variances of the moments $\bar{Y}_{...}, \bar{Y}_{i...}, \bar{Y}_{.j..},$ *and* $\bar{Y}_{..k.}$)**.** *Using the Law of Total Variance (see A.2.1), we can express the variances of* $\bar{Y}_{...}, \bar{Y}_{i...}, \bar{Y}_{.j..},$ *and* $\bar{Y}_{..k.}$ *in terms of the expectation of the conditional variance* ($\mathbb{E}[Var(Y_{ijkl}|Z_i)]$ *and variance of conditional expectation* ($Var(\mathbb{E}[Y_{ijkl}|Z_i])$). *It holds that*

$$Var\left(\bar{Y}_{...}\right) = Var\left(\frac{1}{I}\sum_{i=1}^{I}\bar{Y}_{i...}\right) = \frac{1}{I}Var\left(\bar{Y}_{i...}\right),$$

where

$$Var\left(\bar{Y}_{i...}\right) = Var\left(\mathbb{E}\left[\bar{Y}_{i...}|Z_i\right]\right) + \mathbb{E}\left[Var\left(\bar{Y}_{i...}|Z_i\right)\right]$$

$$= Var\left(\mathbb{E}\left[\frac{1}{JKL}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L}Y_{ijkl}|Z_i\right]\right) + \mathbb{E}\left[Var\left(\frac{1}{JKL}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L}Y_{ijkl}|Z_i\right)\right]$$

$$= Var\left(\frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K}\mathbb{E}\left[Y_{ijkl}|Z_i\right]\right) + \frac{1}{(JK)^2L}\sum_{j=1}^{J}\sum_{k=1}^{K}\mathbb{E}\left[Var(Y_{ijkl}|Z_i)\right].$$

69

Moreover, the variance due to tool and operator are

$$
Var\left(\bar{Y}_{\cdot j \cdot \cdot}\right) = Var\left(\frac{1}{I}\sum_{i=1}^{I}\bar{Y}_{ij\cdot\cdot}\right)
$$

$$
= \frac{1}{I}\left(Var\left(\mathbb{E}\left[\bar{Y}_{ij\cdot\cdot}|Z_i\right]\right) + \mathbb{E}\left[Var\left(\bar{Y}_{ij\cdot\cdot}|Z_i\right)\right]\right)
$$

$$
= \frac{1}{I}\left(Var\left(\frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L}\mathbb{E}\left[Y_{ijkl}|Z_i\right]\right) + \mathbb{E}\left[Var\left(\frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L}Y_{ijkl}|Z_i\right)\right]\right)
$$

$$
= \frac{1}{I}Var\left(\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\left[Y_{ijkl}|Z_i\right]\right) + \frac{1}{IK^2L}\sum_{k=1}^{K}\mathbb{E}\left[Var\left(Y_{ijkl}|Z_i\right)\right],
$$

$$
Var\left(\bar{Y}_{\cdot\cdot k\cdot}\right) = Var\left(\frac{1}{I}\sum_{i=1}^{I}\bar{Y}_{i\cdot k\cdot}\right)
$$

$$
= \frac{1}{I}\left(Var\left(\mathbb{E}\left[\bar{Y}_{i\cdot k\cdot}|Z_i\right]\right) + \mathbb{E}\left[Var\left(\bar{Y}_{i\cdot k\cdot}|Z_i\right)\right]\right)
$$

$$
= \frac{1}{I}\left(Var\left(\mathbb{E}\left[\frac{1}{JL}\sum_{j=1}^{J}\sum_{l=1}^{L}Y_{ijkl}|Z_i\right]\right) + \mathbb{E}\left[Var\left(\frac{1}{JL}\sum_{j=1}^{J}\sum_{l=1}^{L}Y_{ijkl}|Z_i\right)\right]\right)
$$

$$
= \frac{1}{I}Var\left(\frac{1}{J}\sum_{j=1}^{J}\mathbb{E}\left[Y_{ijkl}|Z_i\right]\right) + \frac{1}{IJ^2L}\sum_{j=1}^{J}\mathbb{E}\left[Var\left(Y_{ijkl}|Z_i\right)\right].
$$

Proposition A.2.2 (Approximated variances with first order approximation normalization constant)**.** *Let $Y_{ijkl}|Z_i$ be $CMP(\zeta_{ijk}(Z_i), \nu_{ijk})$ and Z_i the latent variable distribution and let $\zeta_{ijk}(Z_i) = Z_i a_j b_k$ and $\nu_{ijk} = \nu_0$. When taking a first order approximation of the normalization constant with model scenario two, the conditional moments are defined by*

$$\mathbb{E}[Y_{ijkl}|Z_i] \approx \zeta_{ijk}(Z_i) - \frac{\nu_{ijk} - 1}{2\nu_{ijk}} \qquad and \qquad Var(Y_{ijkl}|Z_i) \approx \frac{\zeta_{ijk}(Z_i)}{\nu_{ijk}}.$$

Then, for model scenario 2 (which generalizes scenario 1) with the constraints for the fixed effects $\sum_{j=1}^{J} a_j = J$ and $\sum_{k=1}^{K} b_k = K$, we obtain the approximations of the variances in Proposition A.2.1

$$\begin{aligned}
Var\left(\bar{Y}_{i\cdots}\right) &= Var\left(\frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K}(\zeta_{ijk}(Z_i) - \frac{\nu_{ijk} - 1}{2\nu_{ijk}})\right) + \frac{1}{(JK)^2 L}\sum_{j=1}^{J}\sum_{k=1}^{K}\mathbb{E}\left[\frac{\zeta_{ijk}(Z_i)}{\nu_{ijk}}\right] \\
&= Var\left(\frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K}(Z_i a_j b_k - \frac{\nu_{ijk} - 1}{2\nu_{ijk}})\right) + \frac{1}{(JK)^2 L}\sum_{j=1}^{J}\sum_{k=1}^{K}\mathbb{E}\left[\frac{Z_i a_j b_k}{\nu_{ijk}}\right] \\
&= Var(Z_i) + \frac{1}{JKL}\left(\mathbb{E}\left[Z_i\right]\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j}\right),
\end{aligned}$$

$$\begin{aligned}
Var\left(\bar{Y}_{\cdots\cdots}\right) &= \frac{1}{I}Var\left(\bar{Y}_{i\cdots}\right) \\
&= \frac{1}{I}Var(Z_i) + \frac{1}{IJKL}\left(\mathbb{E}\left[Z_i\right]\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j}\right),
\end{aligned}$$

$$\begin{aligned}
Var\left(\bar{Y}_{\cdot j\cdot\cdot}\right) &= \frac{1}{I}Var\left(\frac{1}{K}\sum_{k=1}^{K}(\zeta_{ijk}(Z_i) - \frac{\nu_{ijk} - 1}{2\nu_{ijk}})\right) + \frac{1}{IK^2 L}\sum_{k=1}^{K}\mathbb{E}\left[\frac{\zeta_{ijk}(Z_i)}{\nu_{ijk}}\right] \\
&= \frac{1}{I}Var\left(\frac{1}{K}\sum_{k=1}^{K}(Z_i a_j b_k - \frac{\nu_{ijk} - 1}{2\nu_{ijk}})\right) + \frac{1}{IK^2 L}\sum_{k=1}^{K}\mathbb{E}\left[\frac{Z_i a_j b_k}{\nu_{ijk}}\right] \\
&= \frac{a_j^2}{I}Var(Z_i) + \frac{1}{IKL}\left(\frac{a_j \mathbb{E}\left[Z_i\right]}{\nu_j}\right),
\end{aligned}$$

$$\begin{aligned}
Var\left(\bar{Y}_{\cdot\cdot k\cdot}\right) &= \frac{1}{I}Var\left(\frac{1}{J}\sum_{j=1}^{J}(\zeta_{ijk}(Z_i) - \frac{\nu_{ijk} - 1}{2\nu_{ijk}})\right) + \frac{1}{IJ^2 L}\sum_{j=1}^{J}\mathbb{E}\left[\frac{\zeta_{ijk}(Z_i)}{\nu_{ijk}}\right] \\
&= \frac{1}{I}Var\left(\frac{1}{J}\sum_{j=1}^{J}(Z_i a_j b_k - \frac{\nu_{ijk} - 1}{2\nu_{ijk}})\right) + \frac{1}{IJ^2 L}\sum_{j=1}^{J}\mathbb{E}\left[\frac{Z_i a_j b_k}{\nu_{ijk}}\right] \\
&= \frac{b_k^2}{I}Var(Z_i) + \frac{1}{IJL}\left(b_k \mathbb{E}\left[Z_i\right]\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j}\right).
\end{aligned}$$

Note that for scenario 1, we have that $\nu_{ijk} = \nu_j = \nu_0$ and therefore $\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j} = \frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_o} = \frac{1}{\nu_0}$.

Proposition A.2.3 (Expected sum of squares)**.** *Let Y_{ijkl} be i.i.d. random variables with Z_i being the true count on part i. Then the expectation of the error, part, tool and operator sum of squares sum, denoted by SS_E, SS_P, SS_T and SS_O respectively, are defined by*

$$
\begin{aligned}
\mathbb{E}[SS_E] &= \mathbb{E}\left[\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L}(Y_{ijkl}-\bar{Y}_{ijk\cdot})^2\right] = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\mathbb{E}\left[\left(\sum_{l=1}^{L}Y_{ijkl}^2\right)-L\cdot\bar{Y}_{ijk\cdot}^2\right]\\
&= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\left(\sum_{l=1}^{L}\mathbb{E}\left[Y_{ijkl}^2\right]\right)-L\cdot\mathbb{E}\left[\bar{Y}_{ijk\cdot}^2\right]\right)\\
&= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\left(\sum_{l=1}^{L}\mathbb{E}\left[\mathbb{E}\left[Y_{ijkl}^2|Z_i\right]\right]\right)-L\cdot\mathbb{E}\left[\mathbb{E}\left[\bar{Y}_{ijk\cdot}^2|Z_i\right]\right]\right)\\
&= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\sum_{l=1}^{L}\mathbb{E}\left[Var\left(Y_{ijkl}|Z_i\right)+\mathbb{E}\left[Y_{ijkl}|Z_i\right]^2\right]\right)-L\cdot\mathbb{E}\left[Var\left(\bar{Y}_{ijk\cdot}|Z_i\right)+\mathbb{E}\left[\bar{Y}_{ijk\cdot}|Z_i\right]^2\right]\\
&= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L}\left(1-\frac{1}{L}\right)\mathbb{E}\left[Var\left(Y_{ijkl}|Z_i\right)\right].
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[SS_P] &= \mathbb{E}\left[JKL\sum_{i=1}^{I}(\bar{Y}_{i\cdots}-\bar{Y}_{\cdots\cdot})^2\right] = JKL\cdot\left(\sum_{i=1}^{I}\mathbb{E}\left[\bar{Y}_{i\cdots}^2\right]-I\cdot\mathbb{E}\left[\bar{Y}_{\cdots\cdot}^2\right]\right)\\
&= JKL\cdot\left(\sum_{i=1}^{I}\left(Var\left(\bar{Y}_{i\cdots}\right)+\mathbb{E}\left[\bar{Y}_{i\cdots}\right]^2\right)-I\left(Var\left(\bar{Y}_{\cdots\cdot}\right)+\mathbb{E}\left[\bar{Y}_{\cdots\cdot}\right]^2\right)\right)\\
&= JKL\left(\left(\sum_{i=1}^{I}Var(\bar{Y}_{i\cdots})\right)-I\cdot Var(\bar{Y}_{\cdots\cdot})\right)+JKL\sum_{i=1}^{I}\left(\mathbb{E}[\bar{Y}_{i\cdots}]-\mathbb{E}[\bar{Y}_{\cdots\cdot}]\right)^2\\
&= JKL(I-1)Var(\bar{Y}_{i\cdots})\\
&= JKL(I-1)\left(Var\left(\frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K}\mathbb{E}\left[Y_{ijkl}|Z_i\right]\right)+\frac{1}{(JK)^2L}\sum_{j=1}^{J}\sum_{k=1}^{K}\mathbb{E}\left[Var\left(Y_{ijkl}|Z_i\right)\right]\right).
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[SS_T] &= \mathbb{E}\left[IKL\sum_{j=1}^{J}(\bar{Y}_{\cdot j\cdot\cdot}-\bar{Y}_{\cdots\cdot})^2\right] = IKL\cdot\mathbb{E}\left[\left(\sum_{j=1}^{J}\bar{Y}_{\cdot j\cdot\cdot}^2\right)-J\cdot\bar{Y}_{\cdots\cdot}^2\right]\\
&= IKL\cdot\left(\sum_{j=1}^{J}\mathbb{E}\left[\bar{Y}_{\cdot j\cdot\cdot}^2\right]-J\cdot\mathbb{E}\left[\bar{Y}_{\cdots\cdot}^2\right]\right)\\
&= IKL\cdot\left(\sum_{j=1}^{J}\left(Var\left(\bar{Y}_{\cdot j\cdot\cdot}\right)+\mathbb{E}\left[\bar{Y}_{\cdot j\cdot\cdot}\right]^2\right)-J\left(Var\left(\bar{Y}_{\cdots\cdot}\right)+\mathbb{E}\left[\bar{Y}_{\cdots\cdot}\right]^2\right)\right)\\
&= IKL\left(\left(\sum_{j=1}^{J}Var(\bar{Y}_{\cdot j\cdot\cdot})\right)-J\cdot Var(\bar{Y}_{\cdots\cdot})\right)+IKL\sum_{j=1}^{J}\left(\mathbb{E}[\bar{Y}_{\cdot j\cdot\cdot}]-\mathbb{E}[\bar{Y}_{\cdots\cdot}]\right)^2.
\end{aligned}
$$

And analogue to $\mathbb{E}[SS_T]$, we have

$$
\begin{aligned}
\mathbb{E}[SS_O] &= \mathbb{E}\left[IJL\sum_{k=1}^{K}(\bar{Y}_{\cdot\cdot k\cdot}-\bar{Y}_{\cdots\cdot})^2\right] = IJL\cdot\mathbb{E}\left[\left(\sum_{k=1}^{K}\bar{Y}_{\cdot\cdot k\cdot}^2\right)-K\cdot\bar{Y}_{\cdots\cdot}^2\right]\\
&= IJL\left(\left(\sum_{k=1}^{K}Var(\bar{Y}_{\cdot\cdot k\cdot})\right)-K\cdot Var(\bar{Y}_{\cdots\cdot})\right)+IJL\sum_{k=1}^{K}\left(\mathbb{E}[\bar{Y}_{\cdot\cdot k\cdot}]-\mathbb{E}[\bar{Y}_{\cdots\cdot}]\right)^2.
\end{aligned}
$$

Derivation of the expected mean squares (3.30). Let us assume a first order approximation of normalization constant with model scenario 2. Then by using Proposition A.2.3, we obtain

$$\mathbb{E}[MS_E] = \frac{1}{IJK(L-1)}\mathbb{E}[SS_E]$$

$$= \frac{1}{IJK(L-1)}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L}\left(\frac{L-1}{L}\right)\mathbb{E}\left[\mathrm{Var}\left(Y_{ijkl}|Z_i\right)\right]$$

$$= \frac{1}{IJKL}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L}\frac{\mathbb{E}\left[Z_i\right]a_j b_k}{\nu_{ijk}}$$

$$= \mathbb{E}\left[Z_i\right]\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j}.$$

$$\mathbb{E}[MS_P] = \frac{1}{I-1}\mathbb{E}[SS_P]$$

$$= \frac{JKL(I-1)}{I-1}\mathrm{Var}(\bar{Y}_{i\cdots})$$

$$= JKL\left(\mathrm{Var}\left(Z_i\right) + \frac{1}{JKL}\left(\mathbb{E}\left[Z_i\right]\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j}\right)\right)$$

$$= JKL\cdot\mathrm{Var}\left(Z_i\right) + \mathbb{E}\left[Z_i\right]\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j}$$

$$= \mathbb{E}[MS_E] + JKL\cdot\mathrm{Var}\left(Z_i\right).$$

$$\mathbb{E}[MS_T] = \frac{1}{J-1}\mathbb{E}[SS_T]$$

$$= \frac{IKL}{J-1}\left(\left(\sum_{j=1}^{J}\mathrm{Var}(\bar{Y}_{\cdot j\cdot\cdot})\right) - J\cdot\mathrm{Var}(\bar{Y}_{\cdots})\right) + \frac{IKL}{J-1}\sum_{j=1}^{J}\left(\mathbb{E}[\bar{Y}_{\cdot j\cdot\cdot}] - \mathbb{E}[\bar{Y}_{\cdots}]\right)^2$$

$$= \frac{IKL}{J-1}\left(\sum_{j=1}^{J}\left(\frac{a_j^2}{I}\mathrm{Var}\left(Z_i\right) + \frac{1}{IKL}\left(\frac{a_j\mathbb{E}\left[Z_i\right]}{\nu_j}\right)\right) - J\left(\frac{1}{I}\mathrm{Var}\left(Z_i\right) + \frac{1}{IJKL}\left(\mathbb{E}\left[Z_i\right]\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j}\right)\right)\right)$$

$$+ \frac{IKL}{J-1}\sum_{j=1}^{J}\left(\left(\mathbb{E}[Z_i]a_j - \frac{\nu_j-1}{2\nu_j}\right) - \left(\mathbb{E}[Z_i] - \frac{1}{J}\sum_{j=1}^{J}\frac{\nu_j-1}{2\nu_j}\right)\right)^2$$

$$= \frac{IKL}{J-1}\left(\frac{\sum_{j=1}^{J}(a_j^2-1)}{I}\mathrm{Var}\left(Z_i\right) + \frac{J-1}{IJKL}\mathbb{E}\left[Z_i\right]\sum_{j=1}^{J}\frac{a_j}{\nu_j}\right)$$

$$+ \frac{IKL}{J-1}\sum_{j=1}^{J}\left(\mathbb{E}[Z_i](a_j-1) - \frac{\nu_j-1}{2\nu_j} + \frac{1}{J}\sum_{j=1}^{J}\frac{\nu_j-1}{2\nu_j}\right)^2$$

$$= \frac{IKL}{J-1}\left(\sum_{j=1}^{J}(a_j^2-1)\left(\frac{\mathrm{Var}\left(Z_i\right)}{I}\right) + \sum_{j=1}^{J}\left(\mathbb{E}[Z_i](a_j-1) - \frac{\nu_j-1}{2\nu_j} + \frac{1}{J}\sum_{j=1}^{J}\frac{\nu_j-1}{2\nu_j}\right)^2\right)$$

$$+ \mathbb{E}\left[Z_i\right]\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j}$$

$$= \mathbb{E}[MS_E] + \frac{IKL}{J-1}\left(\sum_{j=1}^{J}(a_j^2-1)\left(\frac{\mathrm{Var}\left(Z_i\right)}{I}\right) + \sum_{j=1}^{J}\left(\mathbb{E}[Z_i](a_j-1) - \frac{\nu_j-1}{2\nu_j} + \frac{1}{J}\sum_{j=1}^{J}\frac{\nu_j-1}{2\nu_j}\right)^2\right).$$

$$
\mathbb{E}[MS_O] = \frac{1}{K-1}\mathbb{E}[SS_O]
$$

$$
= \frac{IJL}{K-1}\left(\left(\sum_{k=1}^{K}\mathrm{Var}(\bar{Y}_{..k.})\right) - K\cdot\mathrm{Var}(\bar{Y}_{....})\right) + \frac{IJL}{K-1}\sum_{k=1}^{K}\left(\mathbb{E}[\bar{Y}_{..k.}] - \mathbb{E}[\bar{Y}_{....}]\right)^2
$$

$$
= \frac{IJL}{K-1}\sum_{k=1}^{K}\left(\frac{b_k^2}{I}\mathrm{Var}\left(Z_i\right) + \frac{1}{IJL}\left(b_k\mathbb{E}\left[Z_i\right]\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j}\right)\right)
$$

$$
- \frac{IJKL}{K-1}\left(\frac{1}{I}\mathrm{Var}\left(Z_i\right) + \frac{1}{IJKL}\left(\mathbb{E}\left[Z_i\right]\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j}\right)\right)
$$

$$
+ \frac{IJL}{K-1}\sum_{k=1}^{K}\left(\left(\mathbb{E}[Z_i]b_k - \frac{1}{J}\sum_{j=1}^{J}\frac{\nu_j-1}{2\nu_j}\right) - \left(\mathbb{E}[Z_i] - \frac{1}{J}\sum_{j=1}^{J}\frac{\nu_j-1}{2\nu_j}\right)\right)^2
$$

$$
= \frac{IJL}{K-1}\left(\frac{\sum_{k=1}^{K}(b_k^2-1)}{I}\mathrm{Var}\left(Z_i\right) + \frac{K-1}{IJL}b_k\mathbb{E}\left[Z_i\right]\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j}\right) + \frac{IJL}{K-1}\mathbb{E}[Z_i]^2\sum_{k=1}^{K}\left(b_k^2-1\right)
$$

$$
= \frac{IJL}{K-1}\sum_{k=1}^{K}(b_k^2-1)\left(\frac{\mathrm{Var}\left(Z_i\right)}{I} + \mathbb{E}[Z_i]^2\right) + \mathbb{E}\left[Z_i\right]\frac{1}{J}\sum_{j=1}^{J}\frac{a_j}{\nu_j}
$$

$$
= \mathbb{E}[MS_E] + \frac{IJL}{K-1}\sum_{k=1}^{K}(b_k^2-1)\left(\frac{\mathrm{Var}\left(Z_i\right)}{I} + \mathbb{E}[Z_i]^2\right).
$$

For model scenario 1, which is the special case of scenario 2 with $\nu_j = \nu_0$, $\forall j$, we obtain

$$
\mathbb{E}[MS_E] = \frac{\mathbb{E}\left[Z_i\right]}{\nu_0},
$$

$$
\mathbb{E}[MS_P] = \mathbb{E}[MS_E] + JKL\cdot\mathrm{Var}\left(Z_i\right),
$$

$$
\mathbb{E}[MS_T] = \mathbb{E}[MS_E] + \frac{IKL}{J-1}\sum_{j=1}^{J}(a_j^2-1)\left(\frac{\mathrm{Var}\left(Z_i\right)}{I} + \mathbb{E}[Z_i]^2\right),
$$

$$
\mathbb{E}[MS_O] = \mathbb{E}[MS_E] + \frac{IJL}{K-1}\sum_{k=1}^{K}(b_k^2-1)\left(\frac{\mathrm{Var}\left(Z_i\right)}{I} + \mathbb{E}[Z_i]^2\right).
$$

$\square$

Moment estimates

Proof of Proposition 3.2.1. Let us assume model scenario 1 with a first order approximation of the normalization constant. Moreover, let the latent variable Z_i follow a negative binomial distribution, $NB(\theta, \kappa)$. Let us first denote $\hat{\theta}$ and $\hat{\kappa}$ as the moment estimates for θ and κ. Then, by replacing the first moments in Proposition 3.2.1 with their sample versions we obtain

$$
\bar{Y}_{....} \approx \hat{\theta} - \frac{\hat{\nu_0}-1}{2\hat{\nu_0}}, \qquad \bar{Y}_{.j..} \approx \hat{\theta}\hat{a}_j - \frac{\hat{\nu_0}-1}{2\hat{\nu_0}}, \qquad \bar{Y}_{..k.} \approx \hat{\theta}\hat{b}_k - \frac{\hat{\nu_0}-1}{2\hat{\nu_0}}.
$$

$$
MS_P \approx MS_E + JKL\cdot\left(\hat{\theta} + \hat{\kappa}\hat{\theta}^2\right) \qquad MS_E \approx \frac{\hat{\theta}}{\hat{\nu_0}}.
$$

It easily follows that

$$
\hat{\nu_0} = \frac{1}{1 + 2\bar{Y}_{....} - 2\hat{\theta}}.
$$

Moreover, substitution of $\hat{\nu_0}$ in the parameter tool and operator effect gives us

$$\hat{a_j} = \frac{\left(\bar{Y}_{\cdot j \cdot \cdot} - \bar{Y}_{\cdots} + \hat{\theta}\right)}{\hat{\theta}}, \qquad \hat{b_k} = \frac{\left(\bar{Y}_{\cdot \cdot k \cdot} - \bar{Y}_{\cdots} + \hat{\theta}\right)}{\hat{\theta}},$$

$$\hat{\kappa} = \frac{MS_P - MS_E - JKL \cdot \hat{\theta}}{JKL \cdot \hat{\theta}^2}.$$

Additionally, by substitution of $\hat{\nu_0}$ we obtain

$$MS_E = \hat{\theta} \cdot \left(1 + 2\bar{Y}_{\cdots} - 2\hat{\theta}\right).$$

Solving this gives

$$\hat{\theta}^2 - 2 \cdot \frac{1 + 2\bar{Y}_{\cdots}}{4} = -\frac{MS_E}{2}$$

$$\left(\hat{\theta} - \frac{1 + 2\bar{Y}_{\cdots}}{4})\right)^2 = \frac{-MS_E}{2} + \frac{\left(1 + 2\bar{Y}_{\cdots}\right)^2}{16}$$

$$\hat{\theta} = \frac{1 + 2\bar{Y}_{\cdots}}{4} \pm \sqrt{\frac{-MS_E}{2} + \frac{\left(1 + 2\bar{Y}_{\cdots}\right)^2}{16}}$$

When simulation $\theta = 10$, $\nu = 10$, we have $\mathbb{E}[MS_E] = 1$ and $\mathbb{E}[Y_{\cdots}] = 9.55$. Let now, $MS_E = 1$ and $Y_{\cdots} = 9.55$. Then the estimates $\hat{\theta}_1 = 0.05$ and $\hat{\theta}_2 = 10$. So it can be concluded that

$$\hat{\theta} = \frac{1 + 2\bar{Y}_{\cdots}}{4} + \sqrt{\frac{-MS_E}{2} + \frac{\left(1 + 2\bar{Y}_{\cdots}\right)^2}{16}}.$$

$\square$

Algorithms & new functions

B.1   rCMP-Paus function

In order to solve the precision issue in the *rcmp* function of the *'COMPoissonReg'* package in statistical software [R], we present the *rCMP-Paus* function. The main difference with the *rcmp* function is that the *rCMP-Paus* function uses increased precision for calculation of the probability masses (with *dcmp*) and the logarithm of the gamma function *lgamma*. This is increased precision is enabled using the $as(\dots, "mpfr")$ function included in the *'Rmpfr'* package. Here, MPFR is acronym for "Multiple Precision Floating-Point Reliably" Mächler (n.d.).

```
rcmp.manual <- function (n, lambda, nu)
  {
    max.rcmp <- round(max(50, lambda^(1/nu)*2))
    if (length(lambda) == 1) {
      lambda <- rep(lambda, n)
    }
    if (length(nu) == 1) {
      nu <- rep(nu, n)
    }
    u <- runif(n)
    x <- numeric(n)
    z <- computez.manual(lambda, nu, max.rcmp)
    for (i in 1:n) {
      px <- dcmp(x[i], as(lambda,"mpfr")[i], nu[i],
          z = as(z,"mpfr")[i], max = max.rcmp)
      while (px < u[i]) {
        x[i] <- x[i] + 1
        px <- px + dcmp(x[i], as(lambda,"mpfr")[i], nu[i],
          z = as(z,"mpfr")[i], max = max.rcmp)
      }
    }
    return(x)
  }
```

where

```
  computez.manual <- function (lambda, nu, max)
  {
```

```
    n <- length(lambda)
    L <- matrix(log(lambda), nrow = n, ncol = max + 1, byrow = FALSE)
    M <- matrix(nu, nrow = n, ncol = max + 1, byrow = FALSE)
    J <- matrix(0:max, nrow = n, ncol = max + 1, byrow = TRUE)
    log.res <- J * L - M * as(lgamma(J + 1), "mpfr")
    rowSums(exp(log.res))
  }.
```

B.2   rGP-Paus

The structure of the *rGP-Paus* function is equal to the *rCMP-Paus* function. This new function uses the probability masses of the GP distribution obtained by the *dgenpois* function in the *'VGAM'* package in [R]. The *dgenpois* function uses $\theta = \tau(1 - \varphi)^{-1}$ and $\lambda = (1 + varphi)^{-1}\varphi$.

Due to the applied truncation for certain levels of under-dispersion, the cumulative density function (c.d.f.) does not always sum to 1. We therefore normalize the densities with the c.d.f. of the GP. In the *GP-Paus* function, a random number is drawn from the uniform distribution $U[0, 1]$, after which the corresponding quantile is determined using the normalized probability masses, such that $q(u) = \inf\{x : F(x) \geq u\} = F^{-1}(u)$.

```
    rgenpois <- function (n, lambda, theta)
    {
      max.gp <- ifelse(lambda <0, (ceiling(-theta/lambda)-1),Inf)
      if(lambda >=0){cdf.total <- 1}

      if(lambda < 0){
        cdf.total <- 0
        rep(0,max(max.gp,10000))
        for(i in 0:min(max.gp,10000)){
          contribution <- dgenpois(i,lambda=lambda,theta=theta)
          cdf.total <- cdf.total + contribution
          if(i > theta/(1-lambda)){if(contribution/cdf.total < 10^-5){break}}
        }
      }

      if (length(lambda) == 1) {
        lambda <- rep(lambda, n)
      }
      if (length(theta) == 1) {
        theta <- rep(theta, n)
      }
      u <- runif(n)
      x <- numeric(n)
      for (i in 1:n) {
        px <- dgenpois(x[i],lambda=lambda[i],theta=theta[i]) / cdf.total
        while (px < u[i] && x[i]< max.gp) {
          x[i] <- x[i] + 1
          px <- px + dgenpois(x[i],lambda=lambda[i],theta=theta[i]) / cdf.total
        }
      }
      return(x)
    }
```

Additional results simulation study

C.1 Maximum likelihood estimation

Table C.1: A full overview of the average computation times (in minutes) of the likelihood based analysis methods per simulation scenario (Sim.sc)

Sim.sc	Exp.O1	Exp.O2	Exp.O3	Exp.O4	Gamma	Pois.nu1	QP.norm	Ref.LLH
Simulated datasets with a mean count of 10 - Pois(10)								
1	5.37	5.86	6.10	6.21	0.30	1.18	0.02	21.65
2	5.57	6.27	6.90	7.14	0.30	1.25	0.02	29.37
3	4.46	5.20	5.50	5.82	0.27	1.15	0.02	23.27
4	4.64	5.30	5.66	5.95	0.28	1.22	0.02	30.44
5	5.60	6.20	6.42	6.74	0.33	1.20	0.02	24.86
6	5.89	6.74	7.09	7.83	0.34	1.31	0.02	33.91
7	4.59	5.33	5.78	5.98	0.30	1.20	0.02	25.13
8	4.68	5.30	5.76	6.10	0.27	1.06	0.02	29.00
9	6.22	7.07	7.31	7.80	0.33	1.23	0.02	25.39
10	6.09	6.76	7.49	8.02	0.33	1.32	0.02	33.47
11	5.01	5.61	6.08	6.41	0.29	1.19	0.02	25.60
12	4.34	4.96	5.09	5.26	0.25	1.08	0.02	29.78
Simulated datasets with a mean count of 50 - NBinom(50, 1/2)								
13	28.26	29.02	31.91	34.01	0.24	13.72	0.02	1296.55
14	25.90	27.58	29.79	35.13	0.27	12.89	0.02	1613.88
15	22.25	23.65	27.01	30.47	0.23	12.19	0.02	1206.79
16	21.74	23.88	26.45	29.22	0.26	11.59	0.02	1665.46
17	32.00	28.47	31.28	35.30	0.26	16.52	0.02	1606.68
18	29.45	28.24	32.75	33.39	0.29	15.80	0.02	1990.57
19	26.58	23.28	26.67	28.55	0.24	15.86	0.02	1913.85
20	21.93	20.24	23.95	23.40	0.24	13.26	0.02	2599.36

Likelihood ratio of the reference likelihood with the methods using an asymptotic expansion of the normalization constant
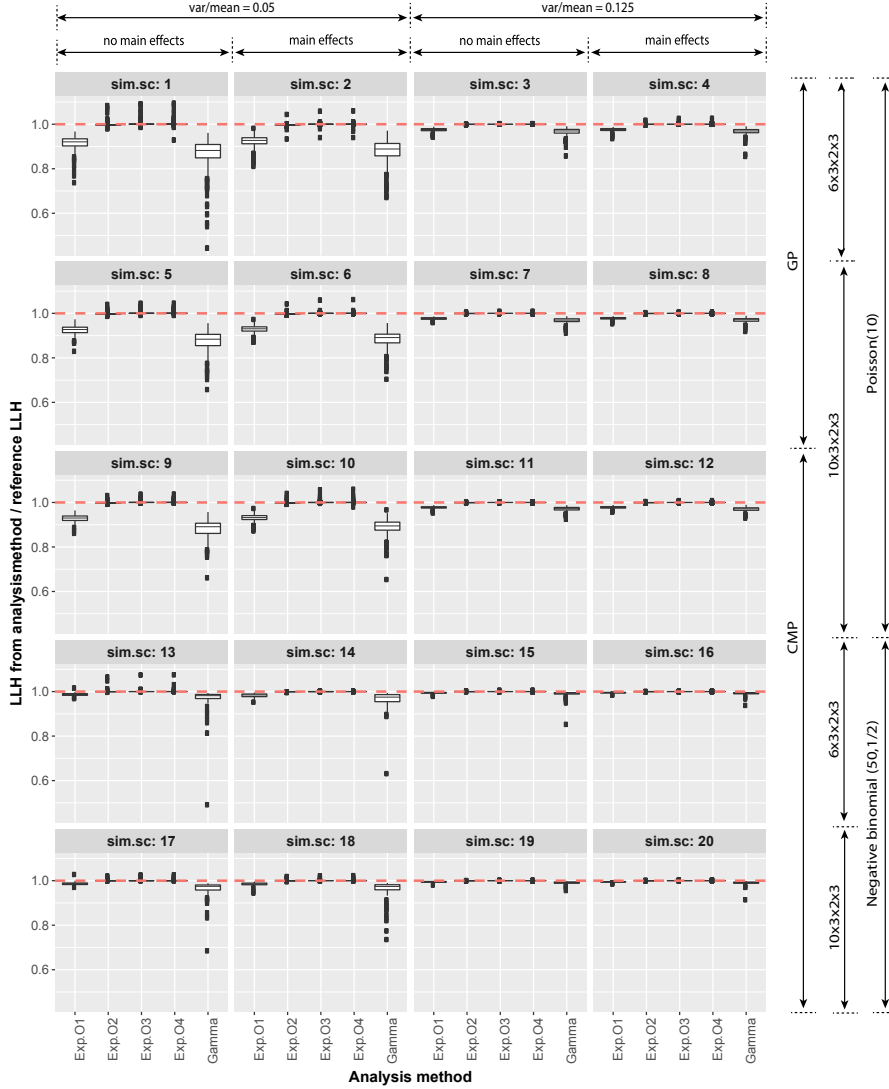

Figure C.1: Ratio of the methods that use an order n asymptotic expansion of the normalization constant with the reference likelihood setting (dotted red line) for simulation scenarios 1 to 20. *Number of simulations = 100.*
*From Section 6.1 it follows that:*
1) Ratio ν/ζ: the lower this ratio, the fewer the relative difference with the reference likelihood. Note that the ratio ν/ζ is higher for $Var/mean = 0.05$ compared to $Var/mean = 0.125$ and higher for $Poisson(10)$ compare to $NBinom(50, 1/2)$. Therefore the dispersion is the highest for the setting $Var/mean = 0.05$ with $Poisson(10)$. 2) For the asymptotic expansion of order 2 (Exp.O2), the improvement of increasing the order is negligible as this method already closely describes the reference likelihood. Moreover, the order 1 expansion (Exp.O1) shows the same pattern as the one with the gamma prior (Gamma). Although more variation is observed for the Gamma method.

Truncation term k_t of the summation of the latent variable Z
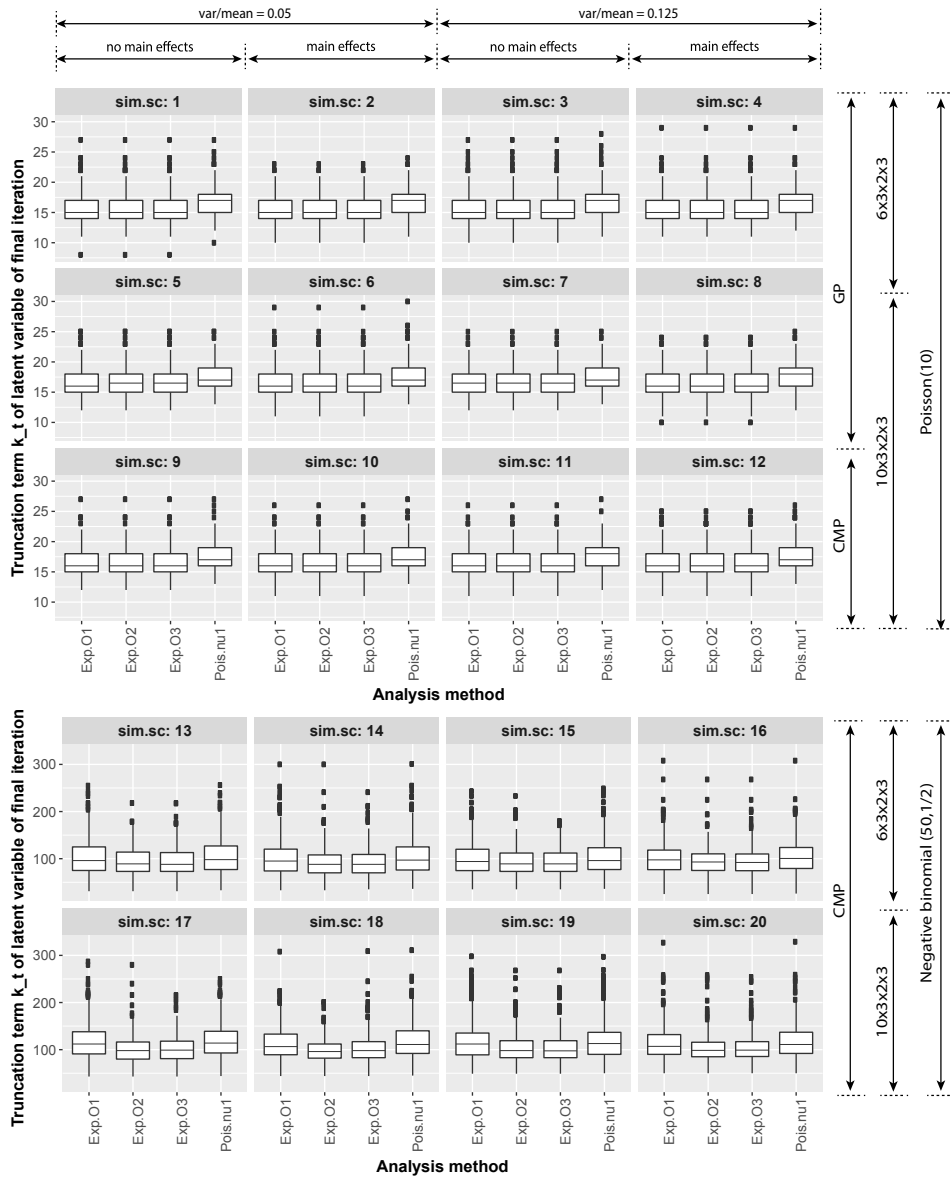
Figure C.2: A boxplot of the values for the truncation term k_t, that was applied for the summation of the latent variable, obtained in the final iteration. Here k_t is based on a relative contribution criterion of $\epsilon_1 = 10^{-5}$ (see Equation 3.5). *Number of simulations = 500 (except for the 'Ref.LLH' in scenario 13 to 20).*

In the final iteration of the maximum likelihood optimization, the summation of the latent variable Z_i was truncated at term k_t. (Note that no statements can be made about the value of the truncation term $k_{t'}$ with $t < t'$.)
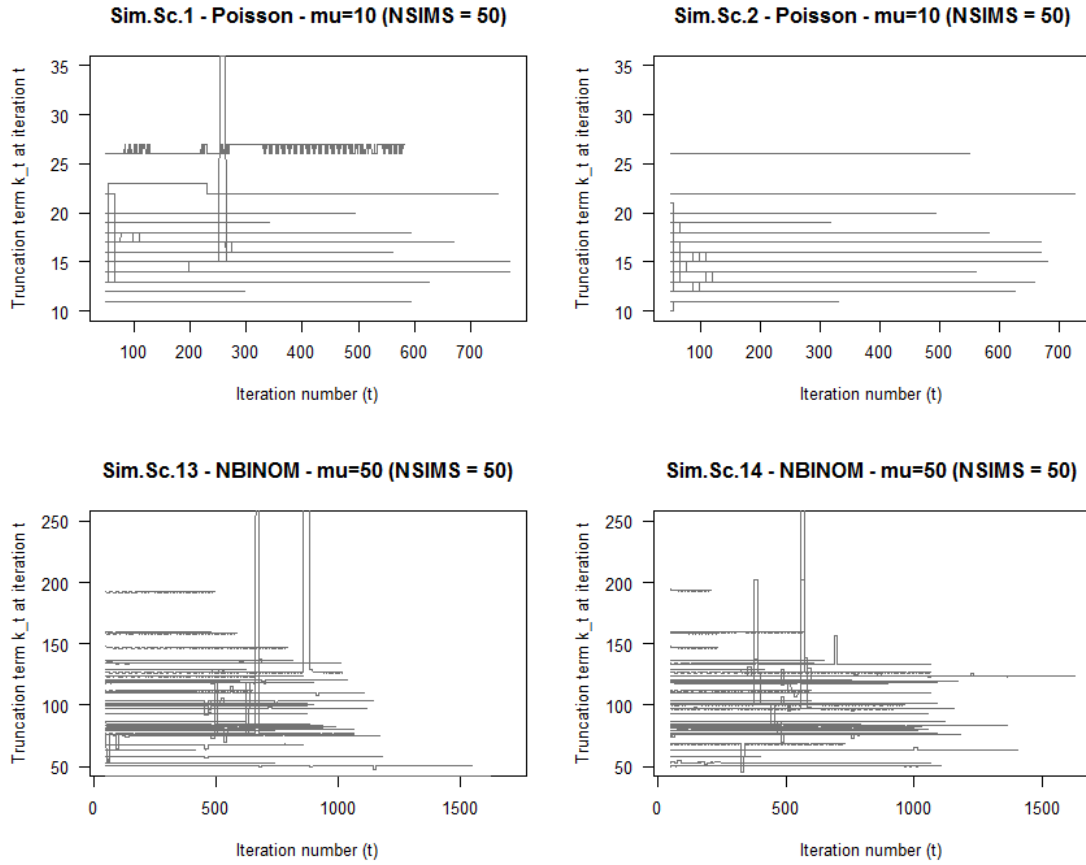
Figure C.3: The value for the truncation term k_t (y-axis) of the summation of the latent variable obtained for each iteration t (x-axis). *Number of simulations = 50.*
Note that the first 50 iterations are omitted in this figure, as they are considered to be the warm-up period of the optimization.
1) Almost no difference in the truncation term k_t was observed (see constant lines) within each maximization procedure for the two simulation scenarios with a Poisson latent variable distribution.
2) For the simulated datasets with a negative binomial, a somewhat larger relative contribution criterion could have been selected to obtain a more constant truncation.

Surface plots likelihood values

(a) *Exp.01* **(b)** *Exp.04* **(c)** *Ref.LLH*

(d) *Pois.nu1* **(e)** *Gamma*
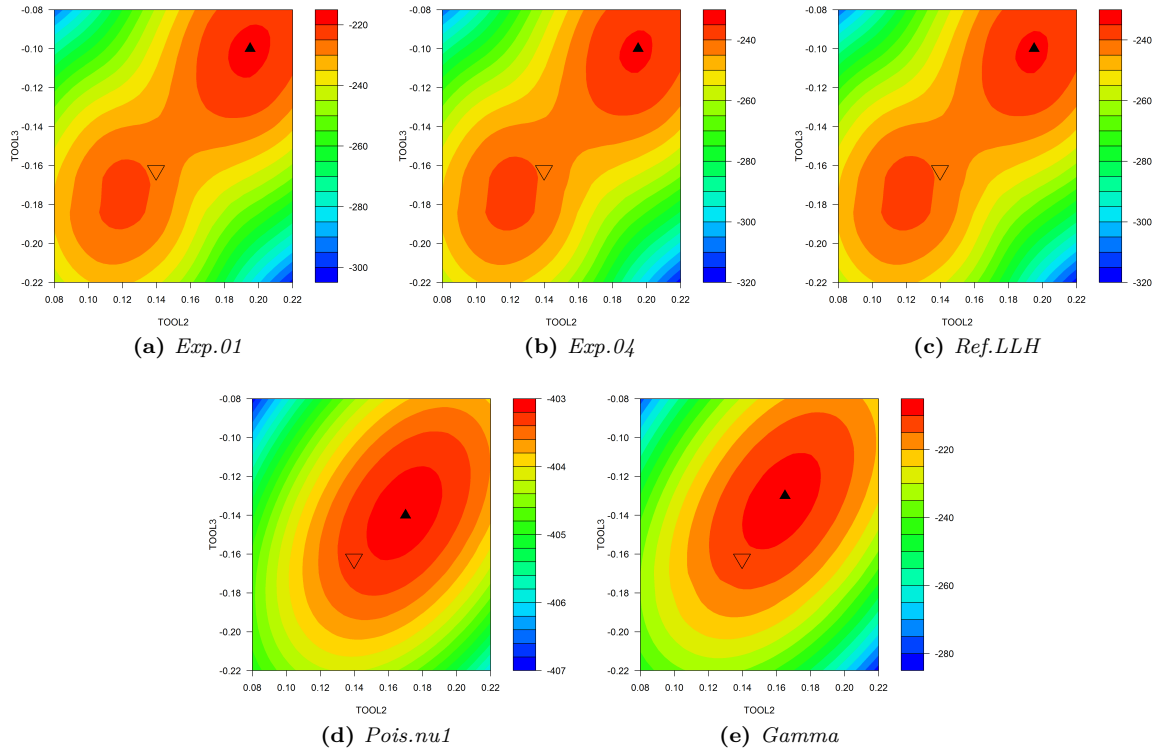
Figure C.4: The 2D-likelihood landscape when fixing all parameters except *TOOL 2* and *TOOL 3*. Simulation setting: *10 parts x 3 tools x 2 operators x 3 repeats*, var/mean=0.05
(x-axis = TOOL 2, y-axis = TOOL 3, filled (upward pointing) triangle is the estimated maximum LLH value, and the not filled (downward pointing) triangle represents the true simulated parameter.)
1) Two local optima are observed for the asymptotic expansion methods (Exp.01 and Exp.O4) and the reference likelihood. This was not observed for the Gamma and Pois.nu1 method.

Performance measures - Dispersion parameter
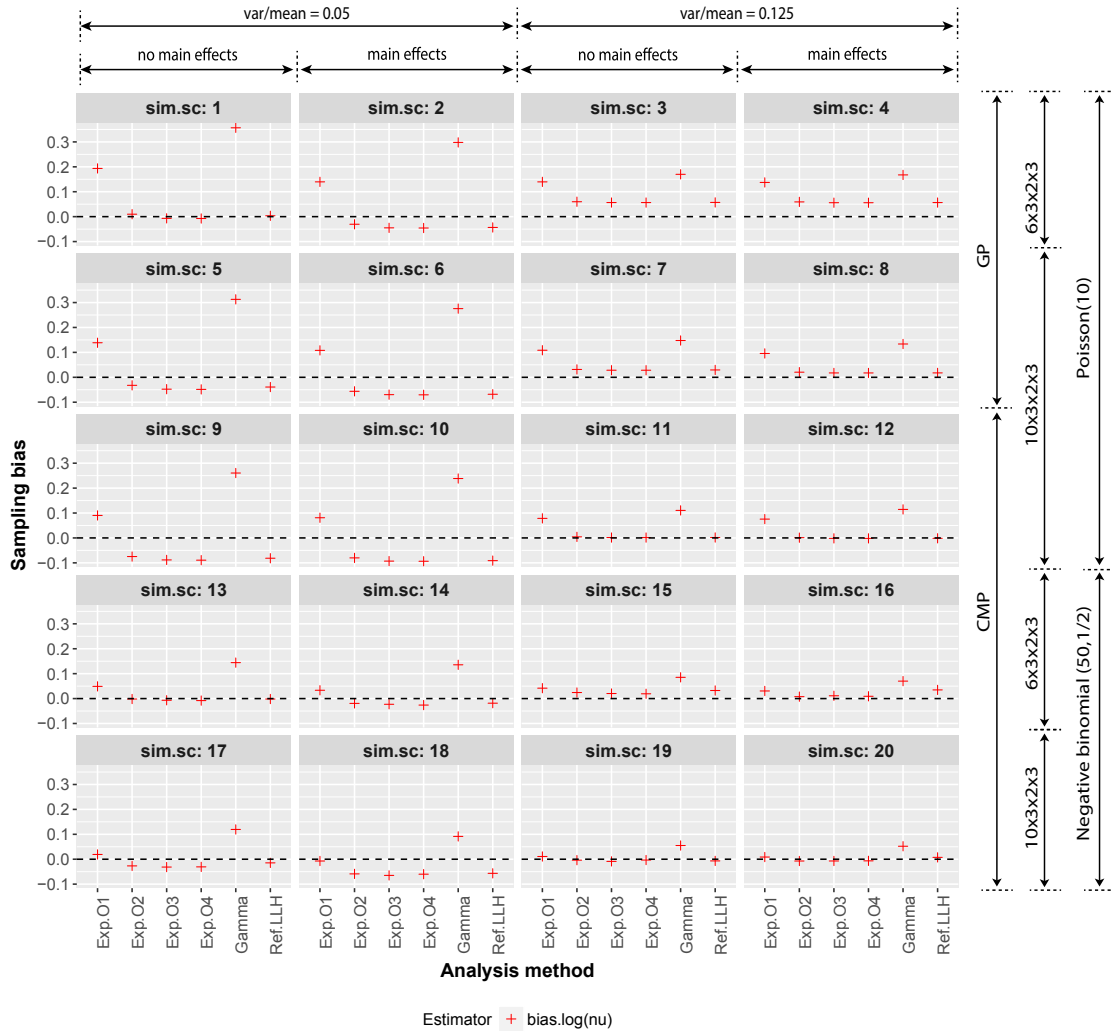
Figure C.5: Sampling bias methods of the dispersion parameter $\gamma_0 = \log(\nu_0)$ based on the CMP latent variable over all simulations for all 20 scenarios. *Number of simulations = 500 (only 100 for Ref.LLH).*

1) Observe a significant bias for the *Gamma* method in the setting of severe under-dispersed data with a Poisson latent variable. The bias decreases for decreasing dispersion levels or increasing average counts.
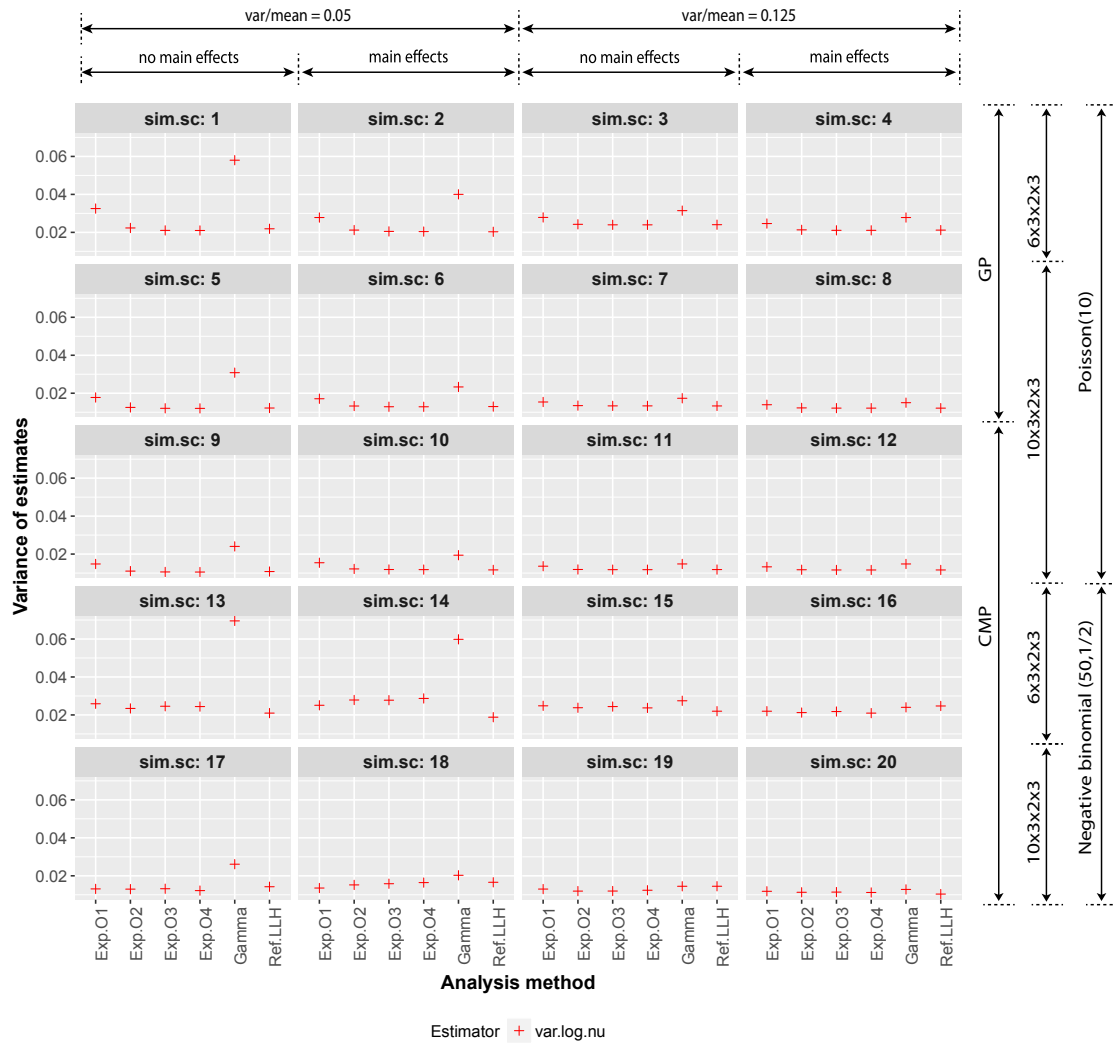
Figure C.6: *Mean square error of the dispersion parameter $\gamma_0 = \log(\nu_0)$ based on the CMP latent variable.*
Note that the mean square error is relatively high for the severe under-dispersed data with a Poisson latent variable.

Approximations for normalization constant $Z(\lambda, \nu)$ (Original form)

Firstly, the truncation of the infinite series at the kth term, that is suggested by Minka et al. (2003), is defined by

$$Z(\lambda, \nu) = \sum_{j=0}^{k} \frac{\lambda^j}{(j!)^\nu} + R_k \qquad (D.1)$$

where $R_k = \sum_{j=k+1}^{\infty} \frac{\lambda^j}{(j!)^\nu}$ is the absolute truncation error. Secondly, an asymptotic approximation for $Z(\lambda, \nu)$ that involves a Laplace approximation of a $(\nu - 1)$-dimensional integral representation was proposed by Shmueli et al. (2005). This representation for $Z(\lambda, \nu)$, that only applies to positive integer values of ν,

$$Z(\lambda, \nu) = \frac{\exp\{\nu \lambda^{1/\nu}\}}{\lambda^{(\nu-1)/2\nu} \cdot (2\pi)^{(\nu-1)/2} \sqrt{\nu}} \{1 + \mathcal{O}(\lambda^{-1\nu})\}. \qquad (D.2)$$

Thirdly, in addition to the first order expansion in (D.2), Gaunt et al. (2016) provided an entire asymptotic expansion for $Z(\lambda, \nu)$ for all real non-negative values of $\nu > 0$. This gives

$$Z(\lambda, \nu) = \frac{\exp\{\nu \lambda^{1/\nu}\}}{\lambda^{(\nu-1)/2\nu} \cdot (2\pi)^{(\nu-1)/2} \sqrt{\nu}} \sum_{k=0}^{\infty} c_k (\nu \lambda^{1/\nu})^{-k}, \qquad \text{as } \lambda \to \infty \qquad (D.3)$$

where the c_j are uniquely determined by the expansion

$$(\Gamma(t+1))^{-\nu} = \frac{\nu^{\nu(t+1)/2}}{(2\pi)^{(\nu-1)/2}} \sum_{j=0}^{\infty} \frac{c_j}{\Gamma(\nu t + (1+\nu)/2 + j)}. \qquad (D.4)$$

The first eight coefficients (from Gaunt et al. (2016)) are

$$c_0 = 1, \quad c_1 = \frac{\nu^2 - 1}{24}, \quad c_2 = \frac{\nu^2 - 1}{1152}(\nu^2 + 23), \quad c_3 = \frac{\nu^2 - 1}{414720}(5\nu^4 - 298\nu^2 + 11237)$$

$$c_4 = \frac{\nu^2 - 1}{39813120}(5\nu^6 - 1887\nu^4 - 241041\nu^2 + 2482411)$$

$$c_5 = \frac{\nu^2 - 1}{6688604160}(7\nu^8 - 7420\nu^6 + 1451274\nu^4 - 220083004\nu^2 + 1363929895)$$

$$c_6 = \frac{\nu^2 - 1}{4815794995200}(35\nu^{10} - 78295\nu^8 + 76299326\nu^6 + 25171388146\nu^4$$
$$- 915974552561\nu^2 + 4175309343349)$$

$$c_7 = \frac{\nu^2 - 1}{115579079884800}(5\nu^{12} - 20190\nu^{10} + 45700491\nu^8 - 19956117988\nu^6$$
$$+ 7134232164555\nu^4 - 142838662997982\nu^2 + 525035501918789)$$

$$(D.5)$$