Comparison of Additive and Multiplicative Bayesian Models for
Longitudinal Count Data With Overdispersion Parameters: A Simulation Study
Peer-reviewed author version

Comparison of Additive and Multiplicative Bayesian Models for Longitudinal Count Data With Overdispersion Parameters: A Simulation Study

Mehreteab Aregay[1]   Ziv Shkedy[2]   Geert Molenberghs[1,2]

[1] *I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium*

[2] *I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium*

Abstract

In applied statistical data analysis, overdispersion is a common feature. It can be addressed using both multiplicative and additive random effects. A multiplicative model for count data incorporates a gamma random effect as a multiplicative factor into the mean, whereas an additive model assume a normally distributed random effect, entered into the linear predictor. Using Bayesian principles, these ideas are applied to longitudinal count data, based on the work of Molenberghs, Verbeke, and Demétrio (2007). The performance of the additive and multiplicative approaches is compared using a simulation study.

Some Keywords: Additive model; Multiplicative model; Deviance information criteria; Overdispersion.

1   Introduction

Overdispersion refers to the presence of excess variability in a data set, relative to the model-based variance. It is common in practice and often results from unexplained heterogeneity in the study population. An ubiquitous example is extra-Poisson variation in count data. Hinde (1982) mentions inappropriate independence assumptions and omitted explanatory variables as important reasons for overdispersion.

One solution is the use of a so-called multiplicative model (Brillinger 1986; Manton *et al.* 1981), in the sense of including a random effect as a factor in the mean model. The most common example is the negative-binomial model, where a gamma random effect is used in the Poisson mean (Greenwood and Yule 1920; Ehrenberg 1959), that is, $Y|\mu \sim \text{Poisson}(\mu), \mu \sim \Gamma(\alpha_1, \alpha_2)$. Customarily, α_1 and α_2 are estimated using maximum likelihood. However, this precludes the use of prior information. To allow for this, several researchers proposed a Bayesian approach in which prior information is used for model

parameters α_1 and α_2 (Deely and Smith 1998; Schluter *et al.* 1997). Nevertheless, none of these authors has obtained a closed-form solution. Bradlow *et al.* (2000) has obtained closed forms by approximating the ratio of two gamma functions with a polynomial expansion. Computation is facilitated because the gamma distribution is conjugate for the Poisson. This implies, at the same time, that for other outcome types, specific conjugate distributions need to be considered.

Alternatively, additive models introduce random effects into the linear predictor (Aitkin 1996), a natural approach when overdispersion is thought to arise from the omission of key cofactors. Such random effects, in count-data models or, more generally, throughout the exponential family, are typically assumed to be normally distributed.

Both of these ideas were brought together in a so-called combined model by Booth *et al.* (2003) and Molenberghs *et al.* (2007). Moreover, Aregay *et al.* (2012) adopted a Bayesian approach based on Markov Chain Monte Carlo for longitudinal overdispersed data. However, less attention has been devoted to evaluation of the relative performance of the additive and multiplicative models. In this paper, we extend the additive model and compare it to the multiplicative model, from a Bayesian perspective. Simulations are used to this effect.

The paper is structured as follows. Section 2 is devoted to the description of a motivating set of data. Section 3 formalizes the additive and multiplicative models. The data are analyzed in Section 4 and the simulation study is described and results reported in Section 5.

2   Anti-epileptic Drug Data

The data were obtained from 89 epileptic patients in which 44 were assigned to the new anti-epileptic drug (AED) and the rest to the placebo group. The patients were measured weekly, followed during 16 weeks, with some of them extended to 27 weeks. The outcome is the number of seizures experienced during the most recent week and the research question is the reduction in the number of seizures by the new therapy. Ample details are given elsewhere, including in Aregay *et al.* (2012) and references therein.

3 Overdispersion Models

In turn, we introduce an additive and multiplicative overdispersion model for the data in Section 2. The approach simultaneously accounts for data hierarchies and extra-Poisson dispersion.

3.1 An Additive Overdispersion Model

To account for overdispersion, McLachlan (1997) among others extends the generalized linear model framework (GLM) by adding an additive random effect to the linear predictor. This is appealing whenever overdispersion is believed to arise from un-modeled heterogeneity, e.g., through covariate omission. We here extend their work to longitudinal count data by considering two separate random effects; one for clustering and the other for overdispersion not already accommodated.

Let Y_{ij} be the number of epileptic seizures for patient i in week j, $(i = 1, 2, \ldots, 89; \; j = 1, 2, \ldots, n_i)$, where n_i is the number of repeated measurements for patient i. Assume $Y_{ij}|b_i, \theta_{ij} \sim \text{Poisson}(\lambda_{ij})$, with parameter given by:

$$\ln(\lambda_{ij}) = \beta_{00} \cdot \text{treat}_i + \beta_{01} \cdot (1 - \text{treat}_i) + \beta_{10} \cdot \text{treat}_i \cdot t_{ij} + \beta_{11} \cdot (1 - \text{treat}_i) \cdot t_{ij} + b_i + \theta_{ij}. \quad (1)$$

Here, t_{ij} denotes the time in weeks at which Y_{ij} is measured, treat_i is 1 for a treated (AED) subject and 0 for a placebo subject. We further assume an independent vague normal prior for the "fixed effect" parameters $\boldsymbol{\beta} = (\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})'$; $\theta_{ij} \sim N(0, \sigma_\theta^2)$ is the overdispersion parameter; $b_i \sim N(0, \sigma_b^2)$ represents the subject-specific effect. The hyper-parameters σ_θ^{-2} and σ_b^{-2} are assumed to have a flat Gamma prior distribution: $\sigma_\theta^{-2} \sim \Gamma(0.01, 0.01)$ and $\sigma_b^{-2} \sim \Gamma(0.01, 0.01)$ (Gelman 2006).

3.2 A Multiplicative Overdispersion Model

The multiplicative overdispersion model coincides with the combined model proposed by Molenberghs *et al.* (2007). In this model, the overdispersion parameter is entered as a multiple factor into the mean of the Poisson distribution. While in the additive approach all random effects enter the predictor function directly, here an outcome-type-specific random effect is introduced, customarily of a conjugate type. For instance, a gamma distribution is often used for count data or times-to-event, whereas a beta distribution is considered for binomial data. While a convenient choice in view of parameter estimation and inferences,

the lack of generality could be seen as a disadvantage. On the other hand, conjugacy comes with its advantages, as discussed in Molenberghs *et al.* (2010).

We now assume that $Y_{ij}|b_i, \theta_{ij} \sim \text{Poisson}(\theta_{ij}\kappa_{ij})$ with

$$\ln(\kappa_{ij}) = \beta_{00} \cdot \text{treat}_i + \beta_{01} \cdot (1 - \text{treat}_i) + \beta_{10} \cdot \text{treat}_i \cdot t_{ij} + \beta_{11} \cdot (1 - \text{treat}_i) \cdot t_{ij} + b_i,$$

$b_i \sim N(0, \sigma_b^2)$, and $\theta_{ij} \sim \Gamma(\alpha_1, \alpha_2)$. For convenience of interpretation and identifiability, we assume $\alpha_1 = \alpha_2$, and denote the common parameter simply by α. Thus, $E(\theta_{ij}) = 1$ and $\text{var}(\theta_{ij}) = 1/\alpha$. We assumed the same prior distribution for the parameters as in Section 3.1, with the exception that a uniform prior distribution is assumed for $\alpha \sim U(0, 100)$.

4   Analysis of the Anti-epileptic Drug Data

For model fitting, we adopted a Bayesian approach using MCMC through R2WinBUGS (Sturtz *et al.* 2005) package. For each one of the models, we used three chains of 100,000 iterations, a 10,000 burn-in sequence, and 100 thinning. Note that the thinning parameter is large, to remove residual autocorrelation from the sequence. The Deviance Information Criteria (DIC, Spiegelhalter *et al.* 2002; Gelman *et al.* 2004) was used for model selection. As Johnson (2007) points out, the DIC comes with certain disadvantages in the context of hierarchical data. Now, similar but different drawbacks apply to all information criteria, and these drawbacks may vary across settings (Verbeke and Molenberghs 2000). Therefore it is advisable to consider several, perhaps just a few, and check whether the broad conclusions agree across them.

Convergence was checked using trace plots and estimated potential scale reduction factors, $\widehat{R}$ (Gelman and Rubin 1992). The values of $\widehat{R}$ for all parameters were close to one, which implies good convergence (Table 1). In addition, the trace plots considered indicate convergence for all model parameters.

A summary of the model fits is shown in Table 1. The DIC value of the multiplicative overdispersion model is smaller than that for the additive overdispersion model, indicating a better fit for the first. In general, both models provide similar result for the posterior mean estimate, the standard deviations, and the 95% credible intervals. In neither model is there a significant treatment effect. The variance of the

random effect is approximately 1.2 in both models. However, the posterior estimate of the intercept for the additive overdispersion model is slightly smaller than for the multiplicative overdispersion model. The variance of the overdispersion parameter, $\sigma_\theta^2 = 0.4021$, for the additive model indicates the presence of extra-variability in the data, which is in line with the value of the variance of the overdispersion parameter for the multiplicative model: α=2.482 and $\text{var}(\theta_{ij}) = 1/2.482 = 0.4029$. When we compare the result of the variance of the random effect with the variance of the overdispersion parameter, the first one is larger than the second one, which shows that the between-subject variability is larger than the extra-model variability. Note that the MC errors are much smaller than the posterior standard deviation, indicating the efficiency of the posterior sample mean for the corresponding population estimand.

Furthermore, to compare both models, the posterior predicted values for each individual were calculated. Results are shown in Figure 1. Both models returned similar predictions. The correlation between the posterior predictive values for the number of epileptic seizures under both models correlated as highly as 0.999. Additionally, we graphically investigated the relationship between the overdispersion parameters for both models; the result is shown in Figure 2. They are highly correlated, with a small amount of non-linearity visible in the relationship. They correlate as highly as 0.97. The above results provide the motivation for a simulation study to investigate the relative performance of both models.

5   Simulation Study

Using simulations, we studied the performance of the additive and multiplicative overdispersion models in terms of parameter estimation and computation time. We first present the setup, then the estimation strategy, and finally the results. In particular, data will be generated under both the additive and multiplicative model, and then both models will be fitted to such simulated data.

5.1   Setup of the Simulation Study

5.1.1   Definition of the Simulation Scenarios

The general principles of this simulation study are similar to the ones of Aregay *et al.* (2012).

In the first setting, we simulated data according to both models in Section 3, with true values of β_{00}=2,

Table 1: *Epilepsy Data. Posterior summary statistics for the the additive and multiplicative models.*

Par.	Mean		SD		MC error		95% Credible interval		$\widehat{R}$	
	Add	Mult	Add	Mult	Add	Mult	Add	Mult	Add	Mult
β_{00}	0.4696	0.6485	0.1874	0.1854	0.0048	0.0042	(0.0895, 0.8379)	(0.2813, 0.9949)	1.00	1.00
β_{01}	0.736	0.9117	0.1808	0.1835	0.0045	0.0042	(0.3755, 1.094)	(0.5538, 1.269)	1.00	1.00
β_{10}	-0.0133	-0.0117	0.0076	0.0075	1.42E-04	1.52E-04	(-0.0284, 0.0012)	(-0.0262, 0.0036)	1.00	1.01
β_{11}	-0.0273	-0.0249	0.0078	0.0078	1.38E-04	1.53E-04	(-0.0430, -0.0119)	(-0.0398, -0.0091)	1.00	1.01
β_{10}- β_{11}	0.0132	0.0131	0.0109	0.0108	1.99E-04	2.03E-04	(-0.0079, 0.0348)	(-0.0084, 0.0348)	1.00	1.00
β_{10}/β_{11}	0.5384	0.5098	0.9448	0.5744	0.0179	0.0107	(-0.0486, 1.532)	(-0.1411, 1.69)	1.00	1.05
σ_b^2	1.19	1.186	0.2042	0.2084	3.64E-03	3.86E-03	(0.8668, 1.662)	(0.8563, 1.666)	1.00	1.00
σ_θ^2	0.4021	-	0.0364	-	2.36E-04	-	(0.3359, 0.4772)	-	1.00	1.00
α	-	2.482	-	0.2109	-	-	0.0043	(2.093, 2.93)	1.00	1.00
DIC	4868.32	4838.73								

β_{01}=-2, β_{10}=0.05, β_{11}=0.2 and different overdispersion levels. The true values of the dispersion parameter α for the multiplicative overdispersion model were 0.25, 1, and 25, representing high, moderate, and low overdispersion, respectively (Aregay *et al.* 2012). The standard deviation of the overdispersion parameter (σ_θ) for the additive model was varied by specifying different values: 2, 0.8, and 0.2, also here indicating high, moderate, and low overdispersion. We selected these values because when we simulated the data under the multiplicative model with α=0.25, 1, and 25 and fitting the additive model to the simulated data, we obtained estimated values of σ_θ around 2, 0.8, and 0.2, respectively. The values for σ_b were equal to 0.1 and 0.5 (Aregay *et al.* 2012). Covariates for treatment and time were included. The number of time points was equal to 2, 5, 10, and 20. The sample sizes were 30, 60, and 120 subjects; equally divided between the two treatment arms (experimental and placebo). Note that in setting 1, we used also unbalanced time points up to a maximum of 27 similar to the epilepsy data set for 60 subjects. All of these choices taken together led to 84 scenarios. For each scenario, 100 data sets were sampled. Note that the second and third settings are presented in the Supplementary Appendix.
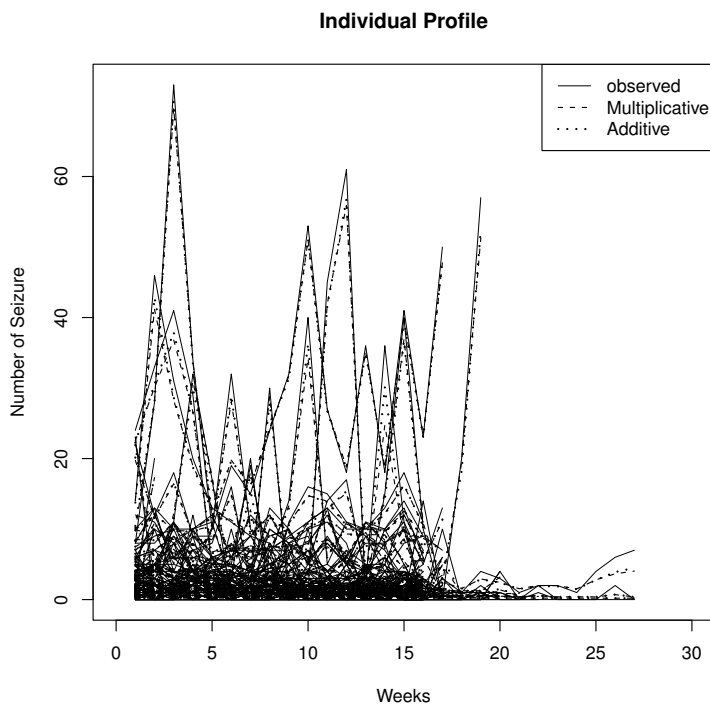
Individual Profile

Figure 1: *Epilepsy Data. Comparison of additive and multiplicative models using fitted individual profiles.*

5.1.2 Fitting Procedure

Both models were fitted for each simulated data set using a Bayesian approach. JAGS (Plummer, 2003) was the Bayesian software used and it was evoked within R via the package R2jags (Su and Yajima, 2011). We used one chain of size 60,000 MCMC iterations with 30,000 iterations burn-in, and 100 thinning. This nicely stabilizes the sequence and avoids auto-correlation. Note that we considered the same prior distribution for the parameters as in Section 3.1 and Section 3.2.

5.2 Simulation Results

The results for the first setting with sample size 60 subjects and cluster size 10 are shown in Table 2. From the table, we can clearly see that both models provide similar bias, relative bias, variance, and MSE for all the parameters when the data are simulated with low overdispersion, regardless of the model. We noticed that there was high bias and high MSE in α for the data simulated from the multiplicative model with low overdispersion. However, the bias and the MSE of σ_θ is small for the data simulated from model
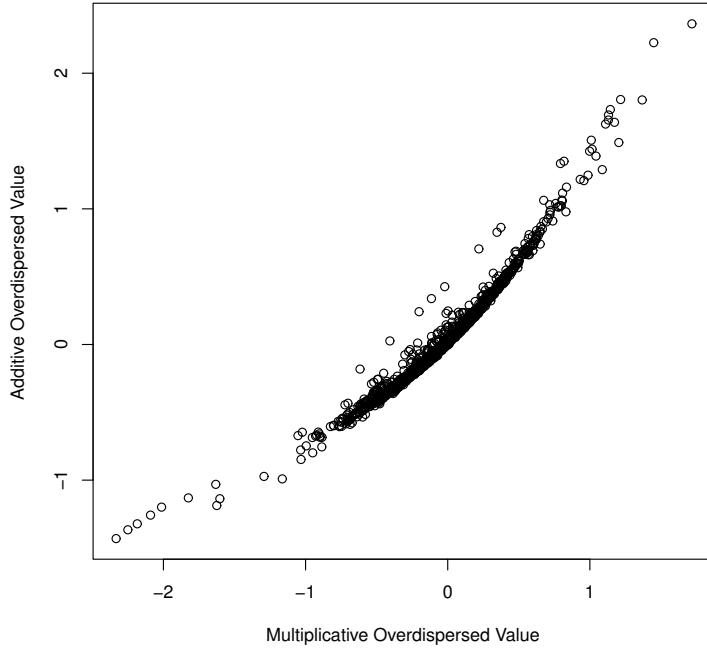
7

Figure 2: *Epilepsy Data. Correlation between the overdispersion parameter obtained from the additive model and the logarithm of the overdispersion parameter obtained from the multiplicative model.*

(1). Hence, caution should be used with the estimate of α when using the multiplicative model for data with low values of overdispersion.

On the other hand, when the data are generated with high and moderate overdispersion levels in the multiplicative setting, the bias and the MSE of the intercept for the additive model (misspecified model) were higher than for the multiplicative model. The reverse is true when the data are generated from model (1). Moreover, for the higher overdispersion scenarios for the data generated under model (1), the bias and MSE of σ_b for the multiplicative model (misspecified model) were higher than for the additive model (Figure 3). However, for the other scenarios, the bias and the MSE of σ_b obtained from both models were similar (Figure 4). Besides, the bias and the MSE of the slopes obtained from both models were similar under all scenarios for moderate and higher overdispersion levels (Table 2; see Supplementary Appendix).

8

The performance of both models was compared in terms of their computation time as well. The results are shown in Table 3. When the data are generated with high and moderate overdispersion levels, regardless of the model, the computation time of the additive model is smaller than that of the multiplicative model. This indicates that the additive model converges quickly compared to the multiplicative model. However, when the data are generated with low overdispersion level, especially with $\sigma_b = 0.1$, the computation time of the additive model does not differ much from the multiplicative model.

In addition to the first setting, two other settings with different true values of the regression coefficients were considered. The results of these settings were similar to the first setting (see Supplementary Appendix). Hence, changing the true values of the regression coefficients does not affect the results.

In all scenarios, in which data are simulated from the multiplicative model with high and moderate overdispersion, fitting the additive model slightly underestimates the intercepts, that is, the bias of the intercepts is negative. In contrast, when the data are simulated under an additive model with high and moderate overdispersion, fitting the multiplicative model slightly overestimates the intercepts (Table 2; see Supplementary Appendix).

Furthermore, we investigated the effect of sample size and cluster size. As expected, the bias and the MSE of the parameters decrease as the cluster and sample sizes increase (Figures 3–8; see Supplementary Appendix). We noticed that, in most of the cases, for data simulated from the multiplicative model with cluster size 2, the bias and MSE of β_{01} and β_{11} obtained from the additive model were much higher than from the multiplicative model. The reverse is true for data simulated from model (1); see Supplementary Appendix. Hence, caution should be used when employing both models for overdispersed data with small cluster sizes.

The results of the data generated with unbalanced time points and 60 subjects, similar in structure to the motivating data set, are shown in Table 4. The results of this setting were similar to the results of the first setting (Table 2). Moreover, we have generated data with sample sizes $n = 30$ and $n = 120$ with unbalanced time points; the results (details not shown) were similar to the results of the balanced time points with the same sample size (see Supplementary Appendix).
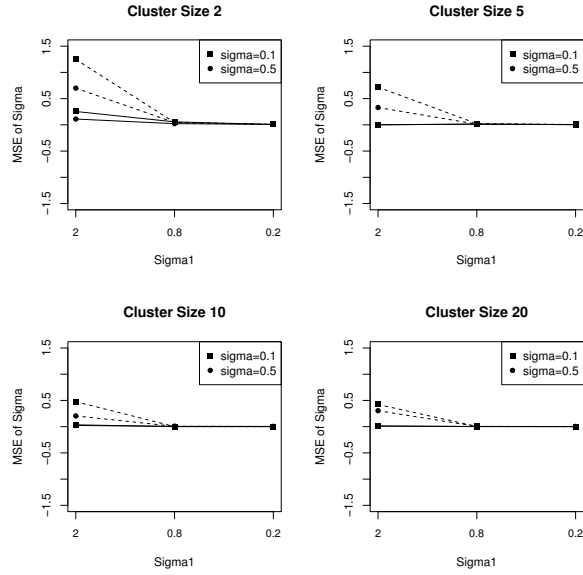
Figure 3: *MSE of σ_b for data generated from additive model and fitting the additive model (solid line) and multiplicative model (dashed line). The x-axis represents the value of σ_θ and the y-axis represents the MSE. $n = 60$.*

In general, both models perform similarly, except that the additive model provides smaller bias and MSE for σ_b for data generated from model (1) with higher overdispersion levels. Note that, when data are generated with $\sigma_b = 0.5$, in most of the scenarios, the bias and MSE of the parameters are slightly smaller than when generated under $\sigma_b = 0.1$.

We observe that, while underdispersion is also possible, it is not possible to generate it directly in a hierarchical fashion. As a consequence, the current study should not be seen as making statements about underdispersion as well.

Table 2: *Summary of the simulation results for $n = 60$, $t = 10$, for data generated and fitted using the additive and multiplicative models, for different overdispersion levels.*

σ_b	Par.			α						σ_θ					
			0.25		1		25		2		0.8		0.2		
			Add	Mult	Add	Mult	Add	Mult	Add	Mult	Add	Mult	Add	Mult
0.1	β_{00}	Bias	-1.789	-0.011	-0.483	-0.012	-0.022	-0.015	-0.056	1.554	0.015	0.305	-0.004	0.015
		Rel.Bias	-0.895	-0.005	-0.241	-0.006	-0.012	-0.008	-0.028	0.777	0.008	0.152	-0.002	0.008
		var	0.099	0.062	0.023	0.0001	0.004	0.003	0.071	0.146	0.012	0.012	0.003	0.003
		MSE	3.300	0.063	0.256	0.016	0.004	0.003	0.074	2.562	0.012	0.104	0.003	0.003
	β_{01}	Bias	-1.912	-0.008	-0.535	-0.022	-0.051	-0.064	0.019	1.673	-0.016	0.301	-0.019	0.0003
		Rel.Bias	0.956	0.004	0.268	0.011	0.025	0.032	-0.009	-0.837	0.008	-0.151	0.009	-0.0002
		var	0.169	0.127	0.085	0.069	0.079	0.059	0.103	0.142	0.080	0.049	0.074	0.074
		MSE	3.825	0.127	0.371	0.069	0.082	0.063	0.104	2.942	0.0801	0.140	0.074	0.074
	β_{10}	Bias	-0.001	-0.003	-0.0001	0.0003	-0.0005	0.001	0.005	0.008	0.0007	<0.0001	0.0002	0.0002
		Rel.Bias	-0.028	-0.058	-0.011	0.007	-0.010	-0.002	0.096	0.160	0.160	<0.0001	0.004	0.004
		var	0.003	0.002	0.0001	0.0004	<0.0001	6.12E-05	0.002	0.003	0.0003	0.0003	6.19E-5	6.17E-5
		MSE	0.003	0.002	0.0001	0.0004	<0.0001	6.33E-05	0.002	0.003	0.0003	0.0003	6.19E-5	6.17E-5
	β_{11}	Bias	0.016	-0.005	0.005	0.001	0.004	0.007	-0.001	-0.004	0.001	-0.0007	0.001	0.001
		Rel.Bias	0.081	-0.027	0.023	0.007	0.020	0.034	-0.005	-0.022	0.005	-0.004	0.006	0.005
		var	0.003	0.003	0.002	0.001	0.001	0.001	0.002	0.003	0.001	0.001	0.001	0.001
		MSE	0.003	0.003	0.002	0.001	0.001	0.001	0.002	0.003	0.001	0.001	0.001	0.001
	σ_b	Bias	0.212	0.170	0.087	0.083	0.015	0.019	0.172	0.679	0.061	0.093	0.021	0.022
		Rel.Bias	2.120	1.702	0.870	0.828	0.154	0.199	1.718	6.786	0.606	0.929	0.212	0.222
		var	0.007	0.006	0.002	0.002	0.0005	0.0005	0.006	0.017	0.001	0.003	0.0004	0.0004
		MSE	0.052	0.035	0.009	0.009	0.0007	0.0009	0.0351	0.477	0.001	0.002	0.0009	0.0009
	α	Bias	-	0.003	-	0.021	-	9.714	-	-	-	-	-	-
		Rel.Bias	-	0.011	-	0.021	-	0.389	-	-	-	-	-	-
		Var	-	0.0005	-	0.014	-	151.08	-	-	-	-	-	-
		MSE	-	0.0005	-	0.015	-	245.44	-	-	-	-	-	-
	σ_θ	Bias	-	-	-	-	-	-	-0.006	-	-0.008	-	-0.003	-
		Rel.Bias	-	-	-	-	-	-	-0.003	-	-0.010	-	-0.017	-
		Var	-	-	-	-	-	-	0.008	-	0.002	-	0.0008	-
		MSE	-	-	-	-	-	-	0.008	-	0.002	-	0.0008	-
0.5	β_{00}	Bias	-1.759	-0.0104	-0.497	0.008	-0.007	0.020	-0.028	1.604	-0.002	0.313	-0.0006	0.018
		Rel.Bias	-0.879	-0.005	-0.249	0.004	-0.003	0.010	-0.014	0.802	-0.0001	0.156	-0.0003	0.009
		var	0.102	0.078	0.023	0.027	0.009	0.009	0.081	0.147	0.018	0.019	0.011	0.011
		MSE	3.196	0.078	0.270	0.027	0.009	0.009	0.082	2.719	0.018	0.117	0.011	0.011
	β_{01}	Bias	-2.029	-0.009	-0.559	0.0006	-0.0367	-0.068	0.039	1.714	0.003	0.287	-0.026	-0.008
		Rel.Bias	1.015	0.005	0.279	-0.0003	0.018	0.034	-0.019	-0.857	-0.001	-0.144	0.013	0.004
		var	0.282	0.161	0.096	0.082	0.064	0.067	0.132	0.154	0.061	0.066	0.071	0.069
		MSE	4.402	0.161	0.409	0.082	0.065	0.071	0.134	3.093	0.061	0.148	0.071	0.070
	β_{10}	Bias	-0.012	0.002	-0.002	-0.004	-0.001	0.001	0.002	0.004	-0.001	-0.001	-0.001	-0.001
		Rel.Bias	-0.241	0.049	-0.030	-0.078	-0.022	0.028	0.046	0.084	-0.020	-0.021	-0.021	-0.021
		var	0.003	0.002	0.0005	0.0005	<0.0001	7.13E-05	0.008	0.002	0.0003	0.0003	5.85E-5	5.89E-5
		MSE	0.003	0.002	0.0005	0.0005	<0.0001	7.33E-05	0.008	0.002	0.0003	0.0003	5.97E-5	6.0E-5
	β_{11}	Bias	0.029	0.001	0.009	-0.002	0.001	0.006	-0.008	-0.014	-0.002	-0.0003	0.002	0.002
		Rel.Bias	0.149	0.005	0.046	-0.008	0.006	0.030	-0.038	-0.068	-0.008	-0.001	0.008	0.009
		var	0.005	0.003	0.002	0.001	0.001	0.0009	0.003	0.004	0.001	0.001	0.001	0.001
		MSE	0.006	0.003	0.002	0.001	0.001	0.0009	0.003	0.004	0.001	0.001	0.001	0.001
	σ_b	Bias	-0.054	-0.004	-0.017	-0.004	-0.0006	0.018	-0.017	0.425	0.0004	0.028	0.018	0.019
		Rel.Bias	-0.108	-0.008	-0.034	-0.008	-0.001	0.036	-0.033	0.849	0.001	0.056	0.036	0.037
		var	0.019	0.024	0.008	0.007	0.003	0.004	0.023	0.025	0.006	0.006	0.004	0.004
		MSE	0.022	0.024	0.008	0.007	0.003	0.004	0.023	0.205	0.006	0.007	0.004	0.004
	α	Bias	-	-0.003	-	0.012	-	8.814	-	-	-	-	-	-
		Rel.Bias	-	-0.010	-	0.012	-	0.353	-	-	-	-	-	-
		Var	-	0.0005	-	0.009	-	169.91	-	-	-	-	-	-
		MSE	-	0.0005	-	0.009	-	247.58	-	-	-	-	-	-
	σ_θ	Bias	-	-	-	-	-	-	0.021	-	0.002	-	-0.003	-
		Rel.Bias	-	-	-	-	-	-	0.011	-	0.002	-	-0.015	-
		Var	-	-	-	-	-	-	0.007	-	0.001	-	0.0008	-
		MSE	-	-	-	-	-	-	0.007	-	0.001	-	0.0008	-

Table 3: *Computational time of the simulation results for $n = 60$, $t = 10$, for data generated and fitted using the additive and multiplicative models, for different overdispersion levels.*

	Data generated from additive model					
			σ_θ			
	2		0.8		0.2	
	Add	Mul	Add	Mul	Add	Mul
$\sigma_b = 0.1$ computational time	14:22:04	21:38:11	15:26:09	21:17:58	15:37:41	16:34:31
	Data generated from multiplicative model					
			α			
	0.25		1		25	
	Add	Mul	Add	Mul	Add	Mul
$\sigma_b = 0.1$ computational time	17:13:14	22:20:25	16:53:57	21:29:50	17:02:43	17:21:31
	Data generated from additive model					
			σ_θ			
	2		0.8		0.2	
	Add	Mul	Add	Mul	Add	Mul
$\sigma_b = 0.5$ computational time	14:25:40	21:37:17	14:11:42	20:31:43	14:26:06	16:28:56
	Data generated from multiplicative model					
			α			
	0.25		1		25	
	Add	Mul	Add	Mul	Add	Mul
$\sigma_b = 0.5$ computational time	14:48:46	23:05:47	14:24:29	21:34:04	14:34:58	17:02:43

6   Concluding Remarks

This paper focused on the comparison of two existing models, with additive and multiplicative approaches to account for overdispersion. A Bayesian view was adopted. First, a comparison of the two models was undertaken using a previously analyzed set of data on patients with epileptic seizures. Second, a simulation study with a total of 108 scenarios was conducted. Furthermore, we extended the additive model to allow for data hierarchies. In the additive model, the overdispersion parameter θ_{ij} is introduced into the linear predictor, whereas in the multiplicative model it is introduced into the mean of the Poisson distribution.

The Deviance Information Criterion was used to select the best model. The multiplicative model performed better than the additive model. However, the results of the posterior estimate of the parameters
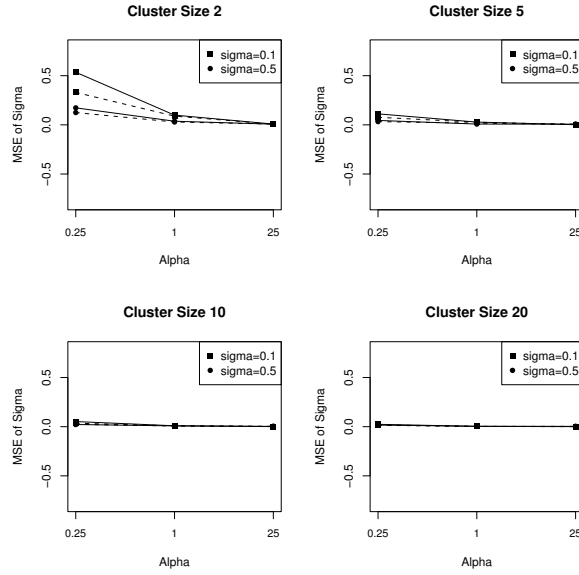
Figure 4: *MSE of σ_b for data generated from multiplicative model and fitting the additive model (solid line) and multiplicative model (dashed line). The x-axis represents the value of α and the y-axis represents the MSE. $n = 60$.*

obtained from the two models were similar. Both models produce non-significant differences between the treatment and placebo groups. Moreover, the estimates of the variance of the random effects were similar in both models. Note that both the posterior mean of the dispersion parameter α of the multiplicative model and the variance of the overdispersion parameter of the additive model, σ_θ^2, indicate that there is excess variability in the data.

To study the relationship between both models, the correlation between the overdispersion parameter θ_{ij} obtained from the additive model and the $\log(\theta_{ij})$ obtained from the multiplicative model was calculated. It was found to be 0.97, which shows that they are highly correlated; this is not surprising, of course. In addition, the individual posterior predictive value of the two models was highly correlated ($\rho = 0.99$), underscoring that both models produce similar predictions over the follow-up period. These individual posterior predictive values were closer to the observed individual profiles, which shows that both models fit the data very well. These results suggest that both model formulations can be used in conjunction perhaps also jointly with the simpler model having no overdispersion parameters. Thus, these taken
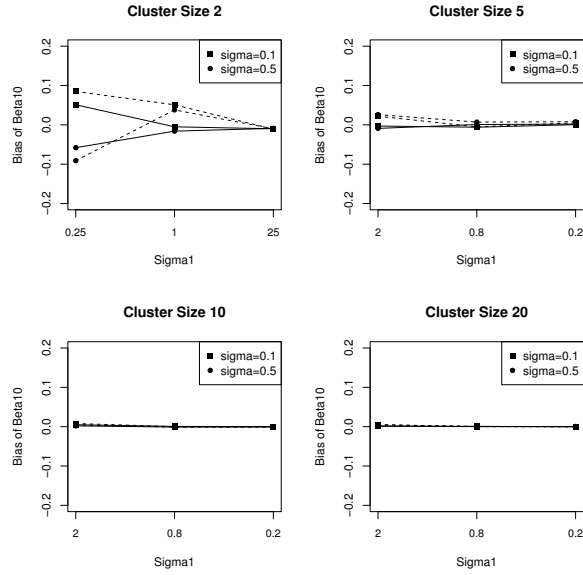
Figure 5: *Bias of β_{10} for data generated from additive model and fitting the additive model (solid line) and multiplicative model (dashed line). The x-axis represents the value of σ_θ and the y-axis represents the bias result.* $n = 60$.

together can be used for sensitivity analysis purposes. Let us give a few examples. If the simpler model differs from the results obtained under the more elaborate models, but these do not differ too much between them, then likely there is overdispersion, neglected by the simpler model but taken up (in various ways) by the more elaborate models. As another example, if both overdispersion models are very different, then either overdispersion is of a very particular shape, picked up by one but not by the other. It is then even possible that there is something else misspecified by them model, and further scrutiny is necessary.

The main findings of our simulation study were as follows. For low overdispersion levels, both models produce similar bias, relative bias, variance, and MSE for all parameters. However, if the data are simulated with moderate to high overdispersion levels, both models produce different result in terms of the intercept, while still maintaining agreement in terms of slopes and variance of the random effect, with the exception that the additive model provides smaller bias and MSE of σ_b^2 than the multiplicative model for data simulated from the additive model with high overdispersion.
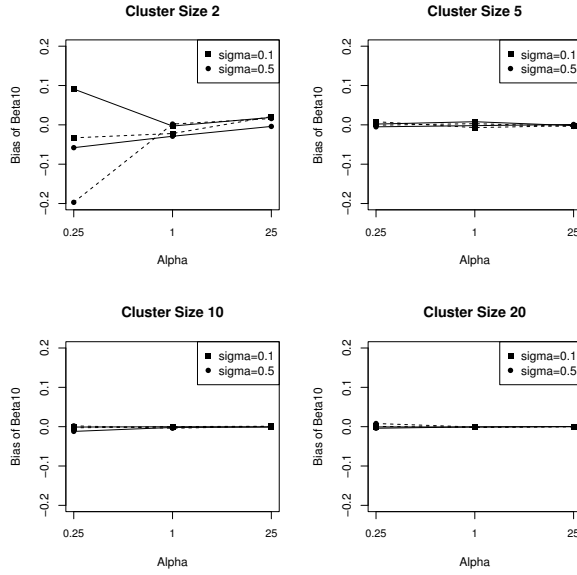
14

Figure 6: *Bias of β_{10} for data generated from multiplicative model and fitting the additive model (solid line) and multiplicative model (dashed line). The x-axis represents the value of α and the y-axis represents the bias result. $n = 60$.*

We also studied the properties of the dispersion parameter α and the variance of the overdispersion parameter, σ_θ. For data generated with high and moderate overdispersion levels, we obtained unbiased and precise estimate of α and σ_θ. However, for low overdispersion level, there is bias, as well as imprecise estimates for α, while we still found unbiased and precise estimates for σ_θ. Hence, we should be careful with the result of α for data with low overdispersion levels.

To assess robustness of the results relative to the choice of true values, three setting were considered. In all cases, the conclusions are in line with expectation. Hence, the results do not depend on the choice of the true values. Besides, the effect of sample size and cluster size was studied. The bias, relative bias, variance, and MSE decrease as the cluster size and sample size increase. Note that both the additive and multiplicative models produce bias and imprecise estimates of β_{01} and β_{10} with cluster size 2. Thus, caution should be exercised when using these models for small cluster sizes.

In summary, both models can be used as useful alternatives for overdispersed data. The additive model is more uniform in terms of using a normal distribution for the overdispersion random effect, whatever the
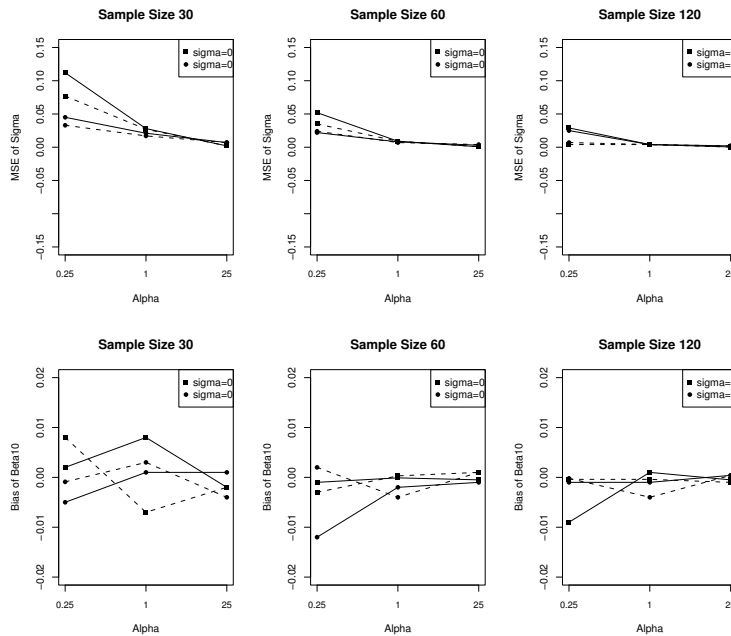
Figure 7: *MSE of σ_b (top figure) and Bias of β_{10} (bottom figure) for data generated from multiplicative model and fitting the additive model (solid line) and multiplicative model (dashed line). Cluster size=10.*

member of the exponential family used would be. On the other hand, the multiplicative model requires specific distributions for the overdispersed parameter for the exponential family under consideration while it is a better choice in terms of parameter estimation and inference. According to the simulation study in this paper, both models perform similarly, except that the additive model produces slightly smaller bias and MSE for σ_b than the multiplicative model for data generated from the additive model with high overdispersion levels. Computationally, the additive model converges faster than the multiplicative model. We recommend to extend the additive model to other exponential family members such as, for example, the binomial model, the Weibull- and exponential-type models, and then compare them with multiplicative model.

Acknowledgments
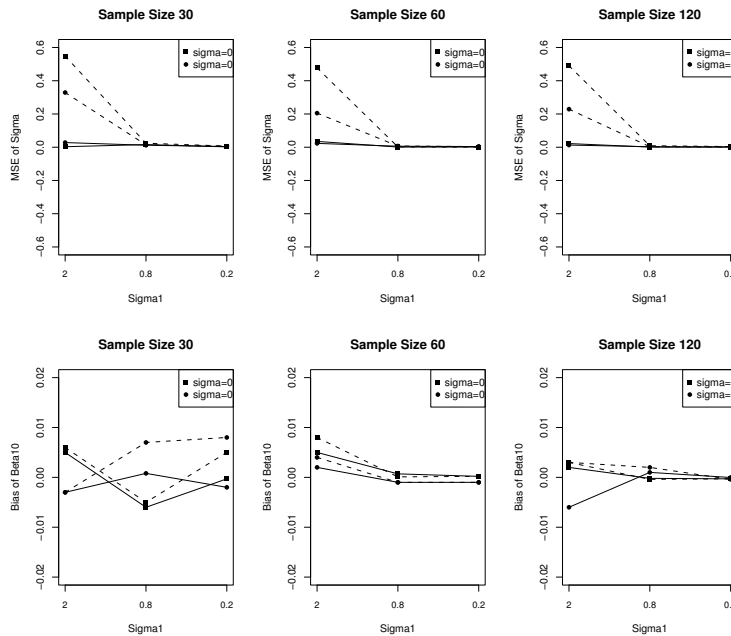
Figure 8: *MSE of σ_b (top figure) and Bias of β_{10} (bottom figure) for data generated from Additive model and fitting the additive model (solid line) and multiplicative model (dashed line). Cluster size=10.*

References

Aitkin, M. (l996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing*, **6**, 251-262.

Aregay, M., Shkedy, Z., and Molenberghs, G. (2012). Hierarchical Bayesian Approach for Analysis of longitudinal count data with Overdispersion Parameters: A Simulation Study. *Submitted for publication.*

Booth, J.G., Casella, G., Friedl, H., and Hobert, J.P. (2003). Negative binomial loglinear mixed models. *Statistical Modelling*, **3**, 179–181.

Brillinger, D.R. (1986). The natural variability of vital rates and associated statistics (with discussion). *Biometrics*, **42**, 693–734.

Deely, J. and Smith, A. (1998). Quantitative refinements for comparisons of institutional performance. *Journal of the Royal Statistical Society, Series A*, **61**, 5–12.

Ehrenberg, A.S.C. (1959). The pattern of consumer purchases. *Applied Statistics*, **8**, 26–41.

Greenwood, M. and Yule, G. (1920). Inquiry into the nature of frequency distributions representative of multiple jappenings with particular reference to the occurrence of multiple attacks of disease or of repeat accidents. *Journal of the Royal Statistical Society*, **83**, 255–279.

Gelman, A. (2006). Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis*, **3**, 515-533.

Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, **7**, 457-511.

Gelman, A., Carlin, J.B., Stern, H.S, and Rubin, D.B. (2004). *Bayesian Data Analysis*. New York: Chapman and Hall.

Hinde, J.P. (1982). *Compound Poisson regression models*. New York: Springer.

Johnson, V.E. (2007). Bayesian model assessment using pivotal quantities. *Bayesian Analysis*, **2**, 719–734.

McLachlan, G.J. (1997). On the EM algorithm for overdispersed count data. *Statistical Methods in Medical Research* , **6**, 76–98.

Manton, K.G., Woodbury, M.A., and Stallard, E. (1981). A variance components approach to categorical data models with heterogeneous cell populations: analysis of spatial gradients in lung cancer mortality rates in North Carolina counties. *Biometrics*, **37**, 259–69.

Molenberghs, G. and Verbeke, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Molenberghs, G., Verbeke, G., and Demétrio, C.G. (2007). An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, **13**, 513–31.

Molenberghs, G., Verbeke, G., Demétrio, C.G.B., and Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325–347.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

Schluter, P. J., Deely, J. and Nicholson, A. (1997). Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model. *The Statistician*, **46**, 293–316.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of Royal Statistical Society, Series B*, **64**, 583-616.

Sturtz, S., Ligges, U. and Gelman, A. (2005). R2winbugs: A package for running WinBUGS from R. *Journal of Statistical Software*, **12**, 1-16.

Su, Y.S. and Yajima, M. (2011). *R2jags: A package for running JAGS from R*. R Foundation for Statistical and Computing. R package version 0.02-14.

Table 4: *Summary of the simulation result for $n = 60$ with unbalanced time points for data generated and fitted using the additive and multiplicative models, for different overdispersion levels.*

σ_b	Par.		α						σ_θ					
			0.25		1		25		2		0.8		0.2	
			Add	Mult	Add	Mult	Add	Mult	Add	Mult	Add	Mult	Add	Mult
0.1	β_{00}	Bias	-1.813	-0.038	-0.503	-0.027	-0.034	-0.002	0.017	1.660	0.005	0.319	-0.0009	0.015
		RelBias	-0.903	-0.019	-0.252	-0.013	-0.017	-0.0009	0.008	0.830	0.003	0.159	-0.0005	0.008
		var	0.057	0.035	0.039	0.008	0.027	0.002	0.041	0.085	0.008	0.008	0.001	0.002
		MSE	3.342	0.036	0.316	0.009	0.027	0.002	0.041	2.842	0.008	0.110	0.001	0.002
	β_{01}	Bias	-1.894	-0.032	-0.526	-0.036	-0.041	0.007	-0.012	1.650	-0.010	0.304	0.002	0.019
		RelBias	0.945	0.016	0.263	0.018	0.021	-0.003	0.006	-0.825	0.005	-0.152	-0.0008	-0.009
		var	0.116	0.054	0.039	0.036	0.038	0.016	0.047	0.094	0.023	0.022	0.013	0.013
		MSE	3.701	0.055	0.316	0.038	0.039	0.016	0.047	2.817	0.023	0.115	0.013	0.013
	β_{10}	Bias	-0.005	0.0009	-0.0005	0.001	-0.0005	-5.65E-05	-0.001	0.003	-0.0009	-0.0008	0.0002	0.0001
		RelBias	-0.104	0.019	-0.009	0.020	-0.009	-0.001	-0.020	0.068	-0.018	-0.0160	0.004	0.002
		var	0.0003	0.0003	8.95E-05	6.51E-05	1.11E-05	1.07E-05	0.0003	0.0006	5.64E-05	6.22E-05	8.27E-6	9.54E-6
		MSE	0.0003	0.0003	8.97E-05	6.61E-05	1.13E-05	1.07E-05	0.0003	0.0006	5.71E-05	6.29E-05	8.31E-6	9.56E-6
	β_{11}	Bias	0.005	-0.0008	0.002	0.002	0.0004	-0.004	0.0005	0.001	-0.0003	-0.0003	-0.0003	-0.0003
		RelBias	0.026	-0.004	0.009	0.009	0.002	-0.002	0.002	0.007	-0.001	-0.001	-0.002	-0.001
		var	0.0007	0.0004	0.0001	0.0001	4.96E-05	5.4E-05	0.0004	0.0008	0.0001	0.0001	5.04E-05	4.44E-5
		MSE	0.0007	0.0004	0.0002	0.0001	4.98E-05	5.41E-05	0.0004	0.0008	0.0001	0.0001	5.05E-05	4.45E-5
	σ_b	Bias	0.163	0.127	0.065	0.055	0.014	0.014	0.133	0.615	0.045	0.075	0.016	0.014
		RelBias	1.629	1.272	0.650	0.545	0.138	0.142	1.329	6.154	0.454	0.747	0.163	0.143
		var	0.006	0.004	0.002	0.0009	0.0004	0.0005	0.005	0.020	0.0009	0.002	0.0005	0.0003
		MSE	0.032	0.020	0.006	0.004	0.0006	0.0007	0.0231	0.399	0.003	0.007	0.0008	0.0005
	α	Bias	-	0.015	-	0.021	-	5.297	-	-	-	-	-	-
		RelBias	-	0.060	-	0.021	-	0.212	-	-	-	-	-	-
		Var	-	0.007	-	0.007	-	79.789	-	-	-	-	-	-
		MSE	-	0.008	-	0.007	-	107.85	-	-	-	-	-	-
	σ_θ	Bias	-	-	-	-	-	-	-0.002	-	0.0009	-	-0.006	-
		RelBias	-	-	-	-	-	-	-0.001	-	0.001	-	-0.028	-
		Var	-	-	-	-	-	-	0.004	-	0.0008	-	0.0004	-
		MSE	-	-	-	-	-	-	0.004	-	0.0008	-	0.0004	-
0.5	β_{00}	Bias	-1.830	-0.023	-0.458	-0.025	-0.022	-0.007	0.014	1.629	0.005	0.318	0.0005	0.019
		RelBias	-0.915	-0.011	-0.229	-0.012	-0.011	-0.004	0.007	0.815	0.003	0.159	0.0003	0.009
		var	0.074	0.044	0.021	0.015	0.012	0.013	0.055	0.129	0.018	0.017	0.011	0.012
		MSE	3.424	0.045	0.231	0.016	0.013	0.013	0.055	2.725	0.018	0.119	0.011	0.012
	β_{01}	Bias	-1.854	-0.024	-0.548	0.006	0.0008	0.019	-0.032	1.652	8.0E-05	0.319	0.004	0.026
		RelBias	0.927	0.012	0.274	-0.003	-0.0004	-0.009	0.016	-0.826	-4E-05	-0.159	-0.002	-0.013
		var	0.135	0.059	0.039	0.041	0.025	0.025	0.077	0.109	0.029	0.029	0.026	0.026
		MSE	3.572	0.059	0.340	0.041	0.025	0.025	0.078	2.837	0.029	0.132	0.026	0.027
	β_{10}	Bias	-0.001	0.0004	-0.004	0.0007	-6.03E-05	0.0002	-0.002	0.002	-0.002	-0.001	-0.0004	-0.0004
		RelBias	-0.026	0.008	-0.072	0.014	-0.001	0.004	-0.043	0.034	-0.032	-0.029	-0.008	-0.008
		var	0.0004	0.0003	0.0001	9.58E-05	1.27E-05	1.84E-05	0.0004	0.0008	6.64E-05	7.33E-05	1.00E-5	1.07E-5
		MSE	0.0004	0.0003	0.0001	9.63E-05	1.27E-05	1.85E-05	0.00008	0.002	6.88E-05	7.54E-05	1.02E-5	1.09E-5
	β_{11}	Bias	0.003	0.0009	0.004	-3.15E-05	-0.0008	-0.0007	0.002	0.002	0.0003	-0.0001	1.00E-05	4.3E-06
		RelBias	0.013	0.005	0.022	-0.0002	-0.004	-0.004	0.009	0.011	0.001	-0.0006	5.0E-05	2.17E-05
		var	0.0007	0.0004	0.0002	0.0002	6.45E-05	6.39E-05	0.0004	0.0009	0.0001	0.0001	6.82E-05	7.4E-05
		MSE	0.0007	0.0004	0.0002	0.0002	6.51E-05	6.44E-05	0.0004	0.0009	0.0001	0.0001	6.82E-05	7.4E-05
	σ_b	Bias	-0.065	-0.009	-0.005	0.016	0.009	0.009	-0.008	0.379	0.017	0.030	0.013	0.013
		RelBias	-0.131	-0.018	-0.011	0.032	0.017	0.018	-0.017	0.758	0.034	0.060	0.025	0.026
		var	0.016	0.018	0.006	0.007	0.004	0.004	0.018	0.032	0.006	0.006	0.004	0.004
		MSE	0.019	0.018	0.006	0.007	0.004	0.004	0.018	0.176	0.006	0.007	0.004	0.004
	α	Bias	-	0.012	-	0.009	-	4.573	-	-	-	-	-	-
		RelBias	-	0.046	-	0.009	-	0.183	-	-	-	-	-	-
		Var	-	0.007	-	0.006	-	78.11	-	-	-	-	-	-
		MSE	-	0.007	-	0.006	-	99.02	-	-	-	-	-	-
	σ_θ	Bias	-	-	-	-	-	-	0.011	-	0.005	-	-0.005	-
		RelBias	-	-	-	-	-	-	0.006	-	0.006	-	-0.024	-
		Var	-	-	-	-	-	-	0.004	-	0.0009	-	0.0003	-
		MSE	-	-	-	-	-	-	0.004	-	0.0009	-	0.0003	-