UNIVERSITY OF TWENTE.

Department of Research Methodology, Measurement and Data Analysis (OMD)

Statistical Analysis of EEG (Trial) Data through Covariance Structure Modelling

Jasmijn L. Bazen

2501791

First supervisor: Jean-Paul Fox

Second supervisor: Stef Baas

Abstract

The analysis of EEG (trial) data is an intensive procedure. Multiple trials will have to be measured to account for random error. The trial data are often averaged, and event related potentials (ERPs) need to be identified. Then, statistical models like the linear mixed effects model (LME) are used for statistical inferences. A good fit of the LME requires support of the factor structure by the data, which can be a challenging task. Furthermore, the random factor variances are restricted to be positive, which limits its suitability for complex structured data. Bayesian Covariance Structure Model (BCSM) bypasses this limitation by modelling directly the correlational structure in the data. A simulation study was conducted for the evaluation of the BCSM compared to the LME model for different sample sizes and covariance values. Then, the BCSM was applied to electroencephalogram (EEG) trial data. The results of the simulation study indicated good performance of the BCSM, for the higher covariance values the models showed comparable performance. For the lower and negative covariances the BCSM performed better. The BCSM showed good recovery of the 95% (posterior) credible intervals and acceptable (posterior) standard deviations. Lastly, the BCSM provided plausible results for the fit on the EEG data. However, as there is no benchmark for quality, any conclusions should be drawn with care. Summarized, this study indicates a promising future for the BCSM, if a way to validate its results can be established.

Keywords: linear mixed effects model, negative correlation, electroencephalogram (EEG), Bayesian Covariance Structure Model (BCSM)

Table of Contents

Introduction

The electroencephalogram (EEG) is a popular means of analysis in multiple fields of research, for example in neuroscience, neurolinguistics, and psychophysiology. This popularity is due to a multitude of reasons, not only is it relatively cost-effective and easily accessible, but also relatively objective due to the absence of introspection – which is the examination of one's mental state – in the measurements. In some cases, it does not need a behavioural response, and the mere presence of a stimulus can be enough for proper measurement of brain activity, making the concept well suited for working with infants who are unable to respond adequately.

Conventional Analysis of Electroencephalogram (EEG) Data

The conventional way of analysis of electroencephalogram (EEG) data is by repeating an event to be researched, this event can be either cognitive, motoric, or sensory, and will result in the brain activity to be analysed. A repetition of the event is referred to as a trial. These raw EEG-signals are then pre-processed and averaged across trials so possible differences can be examined between experimental conditions. This averaging mostly happens in one of two ways (neurotechedu, n.d.). The first is to average all trials of a person in an experimental condition over the random error. This is done through the computing of the grand mean waveform for every time point, making it possible to average over the random error, which enhances the signal-to-noise ratio of the data. This will make it possible to find event-related potentials (ERPs) for this data. An ERP is the to be analysed brain response to a stimulus. The second way is to first find ERPs in the trials, and subsequently average over these ERPs. Finally, the trial ERPs are often used to examine differences between experimental conditions through a regression analysis (Al-Fahoum & Al-Fraihat, 2014; Kretzschmar & Alday, 2020) or an Analysis of Variance (ANOVA) (Kretzschmar & Alday, 2020).

When averaging results across participants (trials), information about differences between participants (trials) will be lost, which leads to a loss of possibly relevant information (Visalli et al., 2023). The method of averaging relies on two assumptions (Frömer et al., 2018). First, it is assumed that averages of all participants are equal in quality and thus have the same weight. In practice, the number of trials can vary per participant. Next to this, the performance accuracy of these trials is variable. These both affect the quality of the averages, and they cannot be expected to be of equal precision. Next to this, not every participant, trial and condition will provide the same number of ERPs, resulting in the same inequality.

Second, by ignoring the heterogeneity within each participant, and only using the mean-averaged ERPs of the participants to analyse between-subject relations, one could gain an incomplete picture, and Simpson's Paradox could occur (Heise et al., 2022). This entails that there can be a positive within-participants effect, but when taking the between-participant effect, it turns out to be negative. This is due to the within-subject variability collapsing. This can result in increased Type-I errors when analysing multiple predictors, or increased Type-II errors with single predictors (Robinson, 1950).

Third, having to measure multiple trials is difficult when working with children, which have difficulty sitting still and keeping attention for a longer time (Heise et al., 2022). Thus, when such an experiment takes place, the researchers will have to perform a case wise deletion of trials based on the quality, this can happen with adult-trials as well. This quality is based on the signal-to-noise ratio; The cut-off-point, however, is arbitrary, meaning it can vary per study. Subsequently, the participants will undergo an arbitrary case wise deletion as well. This entails a high loss of data when conducting mean averaging approaches, which results in large sunk costs and a reduced statistical power. The statistical power of a test is

dependent on sample size, effect size, and variability. Since effect size and variability will stay roughly the same but the sample size will go down, the power will reduce.

The Linear Mixed Effect (LME) Model

One way to analyse EEG data is the linear mixed effect (LME) model. The linear mixed effects (LME) model can be applied on ERPs of participants measured for different conditions and trials. The LME can capture the within-participant and between-participant heterogeneity (e.g., Heise et al., 2022). However, due to the noisy nature of EEG measurements, most EEG studies require some kind of averaging (i.e., repetition of measurements is required) within participants to reduce the random noise and gain statistical power through averaging. It, however, does have several restrictions.

One of the major restrictions of the LME model is it being unable to work with random factor variances close to or below zero (Eldridge et al., 2009). Thus, when modelling variance in ERPs across trials using a random factor, it is necessary that there is substantial variance across trials for each participant. However, in a more complex design, some of the random factors may not be supported by the data which results in the model not fitting properly. Thus, the dependence structure implied by the model is not supported by the data. Next to this, the negative, or small positive correlations in the data are difficult to capture by the LME model. This entails Type-I errors, where one rejects an effect being present when the correlation is there. Because the LME model cannot process these correlations, one would need to know the factor structure of the data beforehand. This gives rise to a paradox, since identifying the most suitable factor structure is usually a goal.

Alongside of this, an LME model requires a certain window of observations per cluster in order to function properly. When this number of observations is too low and there are too many clusters, the random effects (random factor) structure will be overparametrized, meaning there are too many parameters for the amount of data (Matuschek et al., 2017). This

results in Type-I errors and the model not functioning due to unsupported parameters. Furthermore, LME requires a sufficient number of clusters to estimate the random factor variances. These sample size restrictions make it problematic to identify the most suitable random effects structure for the data.

The Bayesian Covariance Structure Model (BCSM)

A flaw in most current methods is the dependency upon the identification of ERPs and the averaging method to analyse these. A model is needed that is applicable to raw EEG data, and thus does not rely upon identifying and averaging ERPs. A new model is proposed to overcome these difficulties; the Bayesian Covariance Structure Model (BCSM). With the BCSM the covariance structure is directly modelled through a structured covariance matrix, instead of modelling covariances through random effects. Thus, while the model still possesses the ability to fit a complex covariance structure, this structure does not need to be supported by the data to be fitted on the data.

Next to this, the BCSM can support the structure of a random effects model, while being able to model positive as well as negative correlations (obeying the positive-definite restriction of the covariance matrix). This ensures that the BCSM can be fitted on data even without support of one or more effects. Furthermore, without using random effects, the sample size restrictions of the BCSM are more lenient than that of the LME.

The BCSM requires a clustered data design to model correlations between and within clusters. Therefore, (cross) serial correlation in the raw EEG data is captured by defining blocks, also known as clusters, and to model (serial) correlation between and within blocks. This results in the direct modelling of the correlation structure of the data. Subsequently, when accounting for the correlations in the data, one could perform a multitude of conventional analyses, among which an Analysis of Variance (ANOVA), which could be used to compare different raw data measurements with each other.

The Aim of this Thesis

In this thesis, the performance of the BCSM is examined through a simulation study. The problematic designs of the LME are considered in which the number of clusters is small and some covariances are close to zero, zero, or negative. This comparison is done through the estimation of both models, which are subsequently compared and analysed using root mean squared errors (RMSEs) for both model parameters, and standard deviations (standard errors/posterior standard deviations) and 95% (posterior) credible intervals for the BCSM only, since the LME model does not provide these statistics for its random factor variables.

The main research question is whether the BCSM will perform better than the LME model in estimating (co)variance parameters. Special attention is given to small samples sizes and/or covariances which are close to zero, zero, or negative. The BCSM can be fitted on data that does not fully support its correlational structure (close to zero and zero). It can handle negative covariance values as well, as long as it does not overstep the positive-definite restriction of the covariance matrix. This lenient property to fit a BCSM will be tested in this study.

Subsequently, the performance of both models will be compared to each other using a simulation study, in which the RMSE is used as a performance statistic. The RMSE will indicate the differences between the true and estimated values of each model parameter. This way, the accuracy of the estimates for both models can be compared.

The BCSM will provide (posterior) standard deviations and 95% (posterior) credible intervals of the estimated covariance parameters. The standard deviation is an indication of the precision of the estimates. The Type-I error rate is examined to check the accuracy of the computed credible intervals. Both statistics will be examined through a simulation study to verify their accuracy.

Finally, the point estimator under the BCSM will be examined, since the posterior distribution of the covariance parameter is expected to be skewed to the right. The posterior mean and in lesser degree the posterior median as a point estimator is expected to be higher than the posterior mode. A simulation study will show the accuracy of the posterior median as a point estimator.

Method Simulation Study

Introducing the Bayesian Covariance Structure Model (BCSM)

Consider a linear mixed effects model (LME) for subject i with fixed effect predictors $\mathbf{X}_i$ and (independent and categorical) random factors $\mathbf{Z}_i$, then the LME model is represented as:

$$Y_i = \sum_{r=1}^{R} X_{ir}\beta_r + \sum_{q=0}^{Q} Z_{iq}\eta_{iq} \qquad (1.1)$$

$$\eta_{iq} \sim N(0, \tau_q)$$

The dependent variable in the model is Y_i. Alongside of these fixed effect variables, there are random factor variables Z_i, where each Z_{iq} is a binary factor variable indicating whether an effect is present or not, and η_{iq} indicates the random effect for subject i for factor variable q. The residual error term is defined for q=0, for all subjects $\mathbf{Z}_0 = 1$, with residual variance τ_0.

The BCSM defines the covariance structure directly without including any random effects. Therefore, the covariance structure implied by the LME is modelled. It follows that:

$$Y_i = \sum_{r=1}^{R} X_{ir}\beta_r + E_i$$

$$E_i \sim N(0, \Sigma_i) \qquad (1.2)$$

$$\Sigma_i = \sum_{q=0}^{Q} \tau_q \, Z_{iq} Z_{iq}^t$$

In the BCSM, the dependent variable has a multivariate normal distribution, and the covariance matrix has a structured design to describe the implied dependences in the data through random effects. Note that the random effect variance parameters τ_q in the LME model are similar to the covariance parameter τ_q when there is a positive covariance among observations. The random effect variance is restricted to be positive in the LME, but in the BCSM this (covariance) parameter can also be zero or even negative. This ensures that the BCSM allows correlations within clusters to be positive, zero, or negative. Making it able to analyse complex data, without knowing the dependence structure beforehand.

Additionally, by avoiding the random effects modelling approach, the number of trials and/or participants will not increase the number of model parameters, which is one of the strengths of the BCSM. Another strength lies in its ability to provide posterior credible intervals and posterior standard deviations for all model parameters. Note that the LME only provides standard errors and credible intervals for the fixed effect parameters, but not for the random factor components.

Simulation Study Design

Data was simulated under a balanced design according to a four-factor nested design. The random factor variances of each of the four components was varied leading to a total of six conditions. In each condition, the four nested factors had different numbers of levels (but equal across the levels of higher factors). The factor levels for the four factors were m = {5; 4; 3; 2}. Therefore, the number of observations per level for the four factors one through four were n = {5, 20, 60, 120}. The number of levels for the four factors one through four were o = {120, 24, 6, 2}, in this simulation study this will be referred to as sample size. This number of four random factors ensured the model to be of sufficient complexity to show where and when the fit of the LME is problematic. The data was simulated for twenty participants in all conditions, the number of data replications was set at 10,000.

Simulation Study Procedure

For every condition, the covariances were estimated for both models. In the LME model, the random factor variances are estimated, and in the BCSM the covariance parameters. These estimates were through several statistics compared with the true covariance values (τ), and subsequently the models' performances were compared to each other. To do this, data was generated for six conditions, where every condition corresponded to four covariance values of differing values. All covariance parameters were equal. These ranged from -.015 to .4, where every value occurred once for every sample size. However, the value -.015 could not be set for the lowest sample size due to the positive definite restriction. The covariance was set to -.005 for this condition, τ = {-.015/-.005, 0, .05, .1, .2, .4}. Sample size was defined as the number of levels for a factor, thus the lowest sample size belonged to the largest cluster in a nested design. For every condition, twenty participants were generated with 10,000 data replications to obtain accurate and reliable results. A participant can be seen as one reproduction of the 10,000 data replications.

The posterior standard deviation (SD) was estimated for the BCSM. Here, the SD is a measure of precision of the estimated covariances. Due to the Bayesian nature of the BCSM, the SD represents the posterior standard deviation of the model parameter. It is the square root of the average squared difference between the posterior parameter values and the posterior mean.

The root mean squared error (RMSE) – which indicates the discrepancy between the estimated and the true value of each model parameter – was estimated for both models. For every data simulation the average squared difference between the estimated and true parameters was computed, and subsequently the square root was taken. The average of all these numbers were then taken and interpreted as the RMSE. A higher sample size was predicted to result in lower RMSEs.

The 95% posterior credible intervals were calculated for the BCSM, to provide an estimation of the Type-I error. These give a sense of the precision of the parameter estimates. When a covariance was in the range of the calculated 95% credible interval of a simulation, a one was prescribed to that simulation. These were then added to each other and divided by the total amount of simulations per factor level (= 10,000). This resulted in a number between zero and one, which represented the retrieved 95% credible interval of the BCSM. Thus, it could be seen whether the posterior credible interval truly was at the 95% level. By subtracting the resulting value from 1, the .05 Type-I error could be calculated.

Results Simulation Study

Parameter Estimates of the Covariances

The covariances were estimated by the LME model, as well as the BCSM. For covariances above zero (τ = {.4, .2, .1, .05}) for both models accurately estimated values were obtained. However, both models' estimations converged slightly less good when the sample size decreased. Here, the number of levels for the four factors was defined as sample size. Sample size influenced the random factor variance estimations (LME) more than the covariance parameter estimates (BCSM).

When the covariance was set to zero, the parameter estimates of both models were similar. However, it should be noted that as the estimated random effect parameter(s) of the LME model approached zero, this led to a problematic fit. The LME model cannot assign a value of zero to its estimates; when the true covariance is close to zero, zero or negative, the model is not (properly) supported by the data. A convergence problem in obtaining the parameter value was observed, this entails that the covariances were non-estimable under the LME. This resulted in a parameter estimate of zero, which in this case of a covariance of zero was correct. However, it was not an indication of the model functioning properly. In contrast, the estimates of the BCSM did not seem to be affected by sample size.

For the negative covariance values, the BCSM tended to have relatively accurate estimates. Particularly when comparing the estimates with the estimates provided by the LME model, since these could not be negative. A convergence problem often occurred for the LME model, resulting in estimates of zero. For the estimates of the BCSM, a higher sample size resulted in a more accurate estimation (see Table 1; for a visualisation, see Figure A1).

Table 1. Parameter estimates (averaged over 10,000 replications) for different sample sizes and covariances for the BCSM and LME

Sample size and model	Covariances					
	.4	.2	.1	.05	0	-.015
120						
LME	.400	.200	.100	.042	.002	. 001
BCSM	.401	.201	.101	.050	.000	-.015
24						
LME	.396	.199	.099	.043	.001	. 000
BCSM	.403	.202	.101	.051	.000	-.015
6						
LME	.392	.193	.095	.041	.002	.000
BCSM	.415	.211	.106	.053	.001	-.021
2						
LME	.395	.201	.106	.074	.010	.000 [a]
BCSM	.412	.200	.099	.049	-.000	-.005 [a]

[a] Here, the covariance value was -.005 instead of -.015.

Quality of Estimates (Root Mean Squared Error)

For the positive covariances for higher sample sizes, the root mean squared errors (RMSEs) for both models behaved approximately the same. However, when the sample size reduced, the RMSEs of the estimated covariance parameters (BCSM) tended to become slightly higher than the RMSEs of the estimated random factor variances (LME model). The RMSEs went up for both models when the sample size decreased.

A major dip could be seen in the RMSEs for the estimations of the LME model when the covariance reached zero (see Figure 1). Here, the RMSEs of the LME model approached zero as well. The LME model searched for an effect that was not supported by the data. This convergence problem resulted in the model prescribing zero to the random factor variances. While this value of zero was in this instance correct, the process behind it was not ideal.

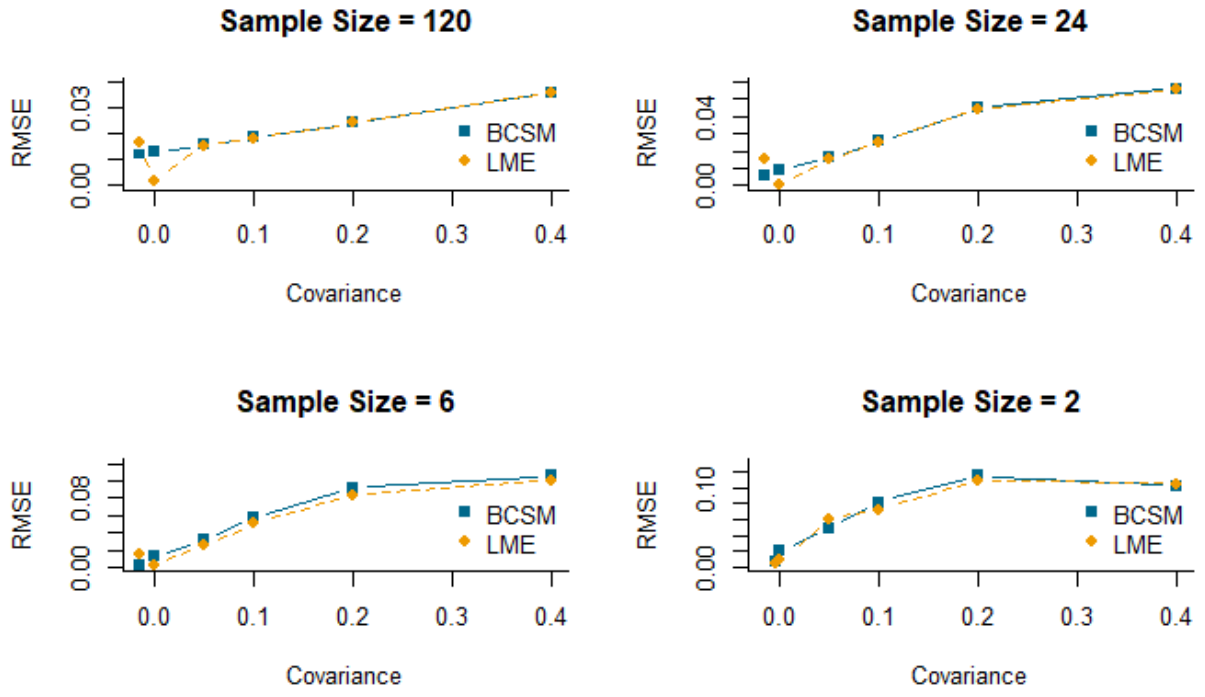


Figure 1. Root mean squared errors (RMSE) (averaged over 10,000 replications) for different sample sizes for the BCSM and LME

The RMSEs of the LME model went up to around .015 for the negative covariance of -.015. This was in line with the random factor variance estimates being around zero for this covariance. It should be noted however, that this did not occur for the smallest sample size with a covariance of -.005. Here, the RMSE for the LME model was around .005. Thus, the RMSEs for the LME was approximately equivalent to the difference between the true value and the estimated value when the highest sample size was concerned. The trend in RMSEs of the estimated covariance parameters (BCSM) across true values of the covariance parameters did not show inconsistencies when the sign changed of the covariance values. This decline became steeper when the sample size decreased (see Figure 1; for all root mean squared errors see Table B1).

Precision of the Estimates (Posterior Standard Deviation)

The posterior standard deviations (SDs) of the BCSM parameters were calculated. They could not be calculated under the LME model, due to restrictions of the model. The posterior distribution is based on a product of a prior and a likelihood component. The contribution of the prior will lead to more uncertainty – a higher SD – when the sample size is small. Next to this, a positive covariance will result in more homogeneity within clusters, subsequently there is less homogeneity between clusters, resulting in a higher SD. On the contrary, a negative covariance will result in a higher homogeneity between clusters, and thus a lower SD.

When concerning larger sample sizes, the SDs formed an almost straight line, when the covariances went up, the standard deviations did as well. For smaller sample sizes, however, the SDs seemed to drop to zero when the covariances became negative. A higher sample size reduced the values of the SDs, thus, the precision of the estimated parameters increased. Whereas a higher covariance value led to an increase of the SDs, indicating a

lowering of precision. Both effects were expected (see Figure 2; for all standard deviations, see Table B2).

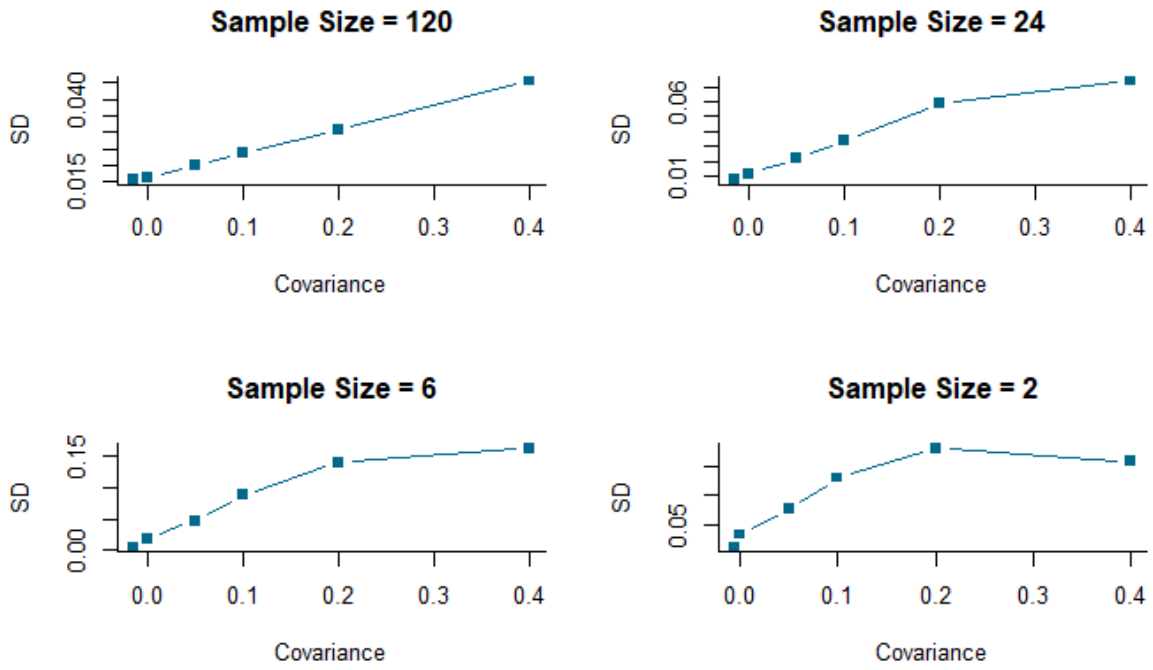


Figure 2. (Posterior) Standard deviations (SDs) (averaged over 10,000 replications) for different sample sizes for the BCSM

Reliability of the Estimates (Credible Intervals and Type-I Error Rates)

The posterior credible intervals (CIs) of the estimates were only provided by the BCSM. For the highest sample sizes of 120 and 24, a recovery rate of the CIs around the preferred .95 was obtained. They did not differ much with ranges of CI = {.947; …; .949} and CI = {.948; …; .950} respectively. A sample size of six resulted in a CI recovery rate of about .95, except for the smallest covariance values zero and -.015. Here, the recovery rate was higher. For the smallest sample size of two, the recovery rate of the credible intervals laid above .95, except for the CI for a covariance of .4, where it was roughly .95 and for a covariance of -.005, where it was as low as .817 (see Figure 3). Lastly, the range between the lower- and upper-bound values of the 95% CI increased in distance from each other when the

sample size went down, and/or the covariance increased (see Figure A2; for all retrieved posterior credible interval values see Table B3; for all lower- and upper-bound values see Table B4).

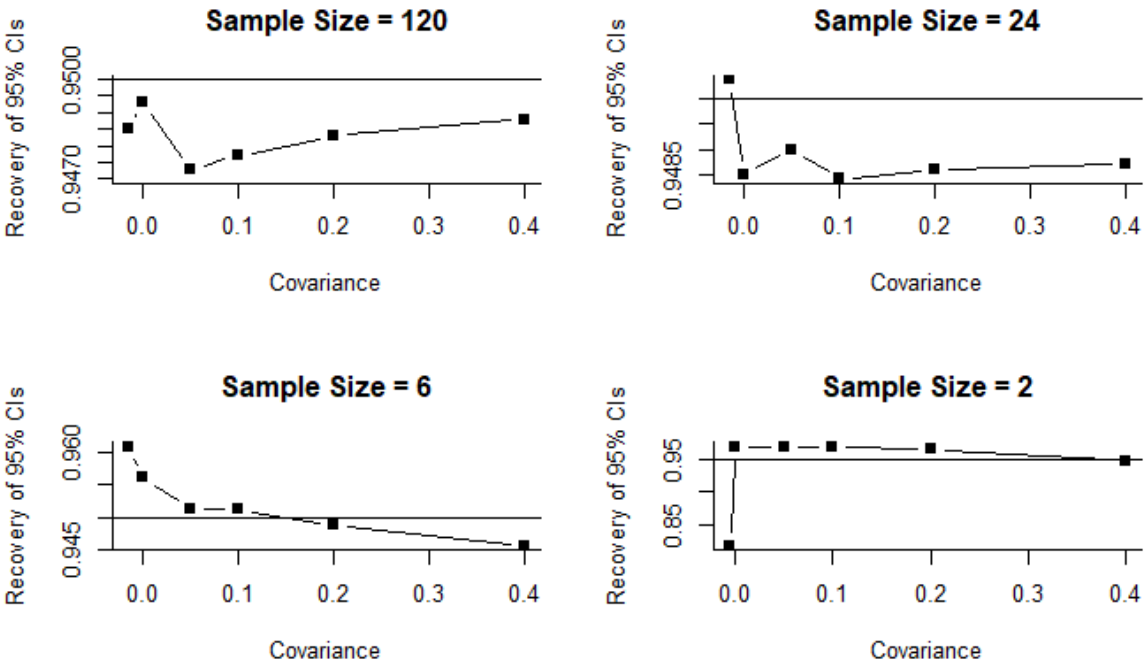


Figure 3. Recovery rate of the 95% (posterior) credible intervals (calculated over 10,000 replications) for different sample sizes for the BCSM

Discussion Simulation Study

The performance of a Bayesian Covariance Structure Model (BCSM) was examined for conditions where the LME is expected to perform poorly. The LME uses a random factor to model dependence between clustered observations, where the random factor variance represents the covariance among clustered observations. Thus, under the LME observations are always expected to be positively correlated, since the random factor variance cannot be zero or negative under the model. The BCSM avoids this problem by modelling the implied covariance structure by the random factor(s) directly. The estimations of the covariance parameters are not restricted to be positive under the BCSM. It allows the covariance

parameters to be negative, zero, or positive as long as the covariance matrix remains positive definite. It was expected that the lower-bound for the random factor variance led to estimation problems for the LME, when considering conditions where the true value for the dependence among observations is small, zero, or even negative.

The purpose of this study was thus to see whether the BCSM would perform better than the LME model in estimating covariances in data that represented a nested clustering. Special attention was put on their performances when the sample sizes were low and/or covariances were close to zero, zero or negative. As could happen in EEG (trial) data. This was done through a comparison of the point estimates of the LME model and the BCSM (random factor variance estimates and covariance parameter estimates respectively). The root mean squared errors (RMSEs) were calculated and compared for both models as well. This statistic indicates the quality of a model. Furthermore, (posterior) standard deviations were calculated for the BCSM, along with (posterior) credible intervals. These indicated the precision of the estimated covariances and the reliability of the model, respectively. These statistics strengthen the BCSM, since the BCSM's point estimates could be higher than the covariances. The proper functioning of these is of importance for a correct statistical analysis.

The quality of the parameter estimates of both models were increased for higher sample sizes and higher (positive) covariances. The sample size was defined as the minimum over the number of levels for the four factors. The RMSEs in these conditions were comparable as well. When the sample size decreased, the estimates became less accurate, and the RMSEs increased. This can have two causes. The first is that statistical models need more data to increase statistical power. The second lies in the data itself. When the sample size of a cluster is low, it is significantly tougher to simulate data with the true factor variances (Guadagnoli & Velicer, 1988; MacCallum et al., 1999). Thus, the imposed structure of the data could not (always) be realized in the simulated data. However, since both models had

some difficulty with smaller sample sizes, and the LME model has been rigorously tested, it can be concluded that the BCSM is up to par to the customary standards under these conditions.

An inconsistency in the trend of sample size decreasing and RMSEs increasing was identified for a covariance of zero. Here, the LME estimate converged to zero, and so did the RMSEs, which represented a poor fit of the model, since it was due to convergence problems. For a negative covariance, the RMSEs of the LME model increased again for the higher sample sizes. This can be explained by the LME's inability to support and estimate negative values, as here the RMSEs took on the squared root value of the covariance. In stark contrast, the inconsistency in the trend was not visible for the BCSM for larger sample sizes. The parameter estimates for the BCSM remained accurate here.

When sample size decreased, the estimates under the BCSM were less good. If the sample size decreases the effect of the prior on the estimation increases. This is problematic, since the data was generated without the use of a prior. This was in line with the LME model, since this model does not include a prior in its estimations. Thus, the effect of prior uncertainty was not present in the data. This discrepancy between the data and the model affected the BCSM to perform less good in smaller sample size conditions.

The additional (posterior) standard deviations of the BCSM model followed clear trends, where the SDs decreased for smaller to negative covariances and higher sample sizes. Both effects were to be expected. The recovered 95% posterior credible intervals tended to be .95 for higher sample sizes. For the smaller sample sizes in combination with smaller covariances, the CIs tended to differ from the .95 mark. Additionally, when sample size decreased and/or the covariance increased, the lower- and upper-bound values of the 95% CI increased. The estimates were less precise in these conditions.

Method EEG Study

EEG Trial Data Study Design and Procedure

The electroencephalogram (EEG) data consisted of 16 trials for two conditions. These conditions were L and R, which represented a left and a right hand respectively. Every trial contained 8,000 data points and was either one of two conditions: one for the left hand and one for the right. All data was obtained through one participant, making it a repeated measurements design. The data was structured so that both the left- and right-hand conditions were stacked together for every x^{th} trial. Resulting in a 'trial' set of 16,000 measurements, with in total 16 trials. The redefined 'trial' will be referred to as a single trial. The BCSM correlated observations within trials and conditions with each other. Then, it compared the conditions with each other. Instead of estimating the effects of variables, it estimated the correlations between measurements.

To see whether an effect was present for the different conditions, a factor structure with four factors was defined. The first two factors modelled the serial correlation present in the data. This factor structure of the serial correlation was based on a covariance matrix proposed by von Rosen (2011), but all blocks were magnified by 10. This matrix contained crossed and nested factors, defining a (serial) correlational structure for 100 observations. The correlational structure was repeated 160 times to describe the data of one trial. The third factor modelled the effect of the experimental condition whether there was an effect present for which hand the participant used. Lastly, the fourth factor modelled the trial specific covariance. Thus, whether within a trial any effect was left unmodelled between the two conditions. This also reflected whether the trials were independently distributed, given the nested factor structure, which is represented by a non-positive within-trial correlation.

The $Z_{iq}Z_{iq}^{t}$-matrix component of Equation 1.2 was constructed in line with the previously mentioned covariance matrix proposed by von Rosen (2011). This was done for

the first two factors used in the analysis of the EEG data. The nested and crossed combinations of the two factors lead to four τ parameters which defined the serial correlation component in the covariance matrix.

To indicate whether there was an effect of the conditions on the data, the mean and (posterior) standard deviations were calculated of the estimated covariances. Particular attention was set on the statistics of the third and fourth factor. Then, the posterior distribution was calculated. This indicated how likely it was that the effect was bigger than zero. Lastly, a density plot was made to give a visual representation of the spread of the estimations of the covariance effect.

Results EEG Study

The first two factors modelled significant amounts of covariance, when considering the means and the (posterior) standard deviations (SD). Because of these effects, the covariance of factor three – which indicated the effect of the variable hand – could be properly analysed as well. This effect was estimated at .149 with an SD of .063. Additionally, in all sub-samples the covariance of this factor was estimated above zero (for all covariance estimates, see Table B5). This was evident from a density plot as well (see Figure A3). It indicated that there was an effect present for the variable hand in the data. Lastly, the fourth factor which indicated a between condition effect was analysed. The estimated covariance was -.035, with an SD of .037. The posterior probability of a positive effect being present in the data was .126. Entailing it was 6.9 times more likely there was a negative or no effect present in the data for this factor, than a positive effect. This was an indication of the previous three factors having addressed all between-trial correlations in the data, and only showing some light support of any within-trial correlation.

Discussion EEG Study

The Bayesian Covariance Structure Model (BCSM) provided plausible results. Not only did it find a positive effect for the variable hand in the data, but it also found strong evidence for no additional between-trial heterogeneity. This was an indication of the model successfully modelling the correlational structure of the data.

It should be mentioned that there is currently no benchmark for how well the model fits. Whereas the results seem promising, one should be careful with drawing conclusions from them for now. Thus, a way of validating the model and its fit on the data should be considered. This is immediately a possible direction for future research. If this validation can be found, the model can be applied in research. This would reduce not only the effort, but also the costs associated with this type of research.

Next to this, electroencephalogram (EEG) data can be non-normally distributed based on the data-observation duration (Hsieh et al, 2022; Sugimoto et al, 1978). By not correcting for this non-normality, the BCSM may have violated some of the assumptions used for the proper analysis of this data. This raises the question whether the results and conclusions drawn from this type of analysis can be assumed to be correct. A relatively straightforward manner to deal with this non-normality is to transform the data with a Box-Cox transformation to a (more) normal distribution. Other methods of normalisation are z-score normalization, min-max normalization, and percentile normalisation.

Conclusion

Electroencephalogram (EEG) data is difficult to analyse. A conventional way of doing this analysis is using a linear mixed effect (LME) model. It represents covariances among clustered observations with random factor variances, which must be positive. Whereas the Bayesian Covariance Structure Model (BCSM) directly models the implied covariance structure, allowing for negative covariances. In the simulation study, for the higher

covariances the models obtained roughly the same values. The BCSM was better able to estimate the covariances for the lower, negative, and non-existent covariances. This is ideal for analysing EEG data where covariances can be negative and it is not certain whether there will be a covariance supported by the data. Both the posterior standard deviations as the credible intervals seemed to function well. The latter indicating a good Type-I error rate. Considering these results, the BCSM can be said to function relatively accurately, even in situations where the LME model does not perform smoothly. The BCSM seemed to work well on the EEG data as well. However, since there is no benchmark for quality, conclusions should be drawn with great care.

References

Al-Fahoum, A. S., & Al-Fraihat, A. (2014). *Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains*. ISRN Neuroscience. https://doi.org/10.1155/2014/730218

Eldridge, S. M., Ukoumunne, O. C., & Carlin, J. B. (2009). The intra-cluster correlation coefficient in cluster randomized trials: A review of definitions. *International Statistical Review, 77*(3), 378-394. https://doi.org/10.1111/j.1751-5823.2009.00092.x

Frömer, R., Maier, M. & Rahman, R. A. (2018). Group-Level EEG-Processing Pipeline for Flexible Single Trial-Based Analyses Including Linear Mixed Models. *Frontiers in Neuroscience, 12*(48), https://doi.org/10.3389/fnins.2018.00048

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103*(2), 265–275. https://doi.org/10.1037/0033-2909.103.2.265

Heise, M. J., Mon, S. K. & Bowman, L. C. (2022). Utility of linear mixed effects models for event-related potential research with infants and children. *Developmental Cognitive Neuroscience, 54*. https://doi.org/10.1016/j.dcn.2022.101070

Hsieh, P., Lin, T., Al-Nashash, H., Mir, H., Muaz, M. & Wong, K. (2022). On the Statistical Normality Rate of EEG Ambient Signal of Healthy Subjects and Its Dependence on Data-Observation Duration. *IEEE Sensors Journal, 22*(22), 21769 – 21779, https://doi.org/10.1109/JSEN.2022.3194699

Kretzschmar, F. & Alday, P. M. (2020). Principles of statistical analyses: old and new tools. *Language Electrified. Techniques, Methods, Applications, and Future Perspectives in the Neurophysiological Investigation of Language.* Springer. https://doi.org/10.31234/osf.io/nyj3k

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor

analysis. *Psychological Methods, 4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

Matuschek, H., Kliegl, R., Vasishth, S., Baayen & H., Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*. *94*, 305-315, https://doi.org/10.1016/j.jml.2017.01.001

Event Related Potentials. (n.d.). NeurotechEDU. http://learn.neurotechedu.com/erp/

Robinson, W.S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review, 15*(3), 351–357. https://doi.org/10.2307/2087176

Sugimoto, H., Ishii, N., Iwata, A., Suzumura, N. & Tomita, T. (1978). On the stationarity and normality of the electroencephalographic data during sleep stages. *Computer Programs in Biomedicine, 8*(3-4), 224-34, https://doi.org/10.1016/0010-468x(78)90029-6

Visalli, A., Montefinese, M., Viviani, G., Finos, L., Vallesi, A. & Ambrosini, E. (2023). *lmeEEG: Mass linear mixed-effects modeling of EEG data with crossed random effects*. bioRxiv https://doi.org/10.1101/2023.01.18.524560

Von Rosen, T. (2011). On the inverse of certain patterned sums of matrices with Kronecker product structures. *Linear and Multilinear Algebra*, *59*(5), 595-606, https://doi.org/10.1080/03081081003718000

Appendix A

Plots of the Performance Statistics

Figure A1. Parameter estimations (averaged over 10,000 replications) for different sample sizes for the BCSM and LME

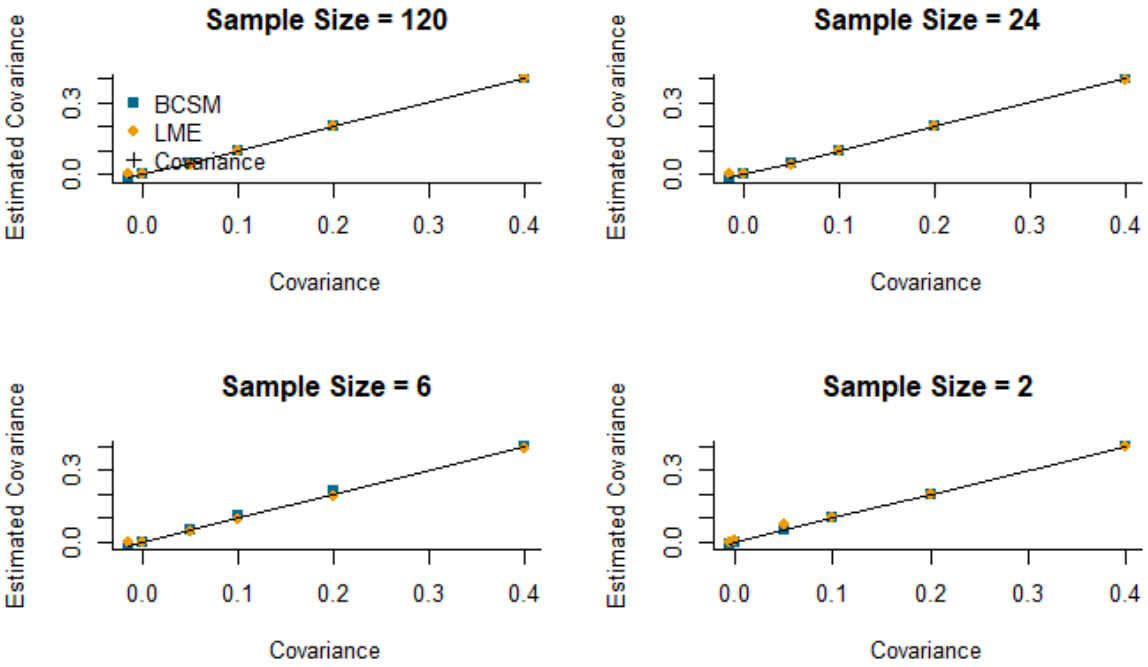


Figure A2. Upper- and lower-bound estimates of 95% (posterior) credible intervals of the covariance parameters for different sample sizes averaged across 10,000 replications

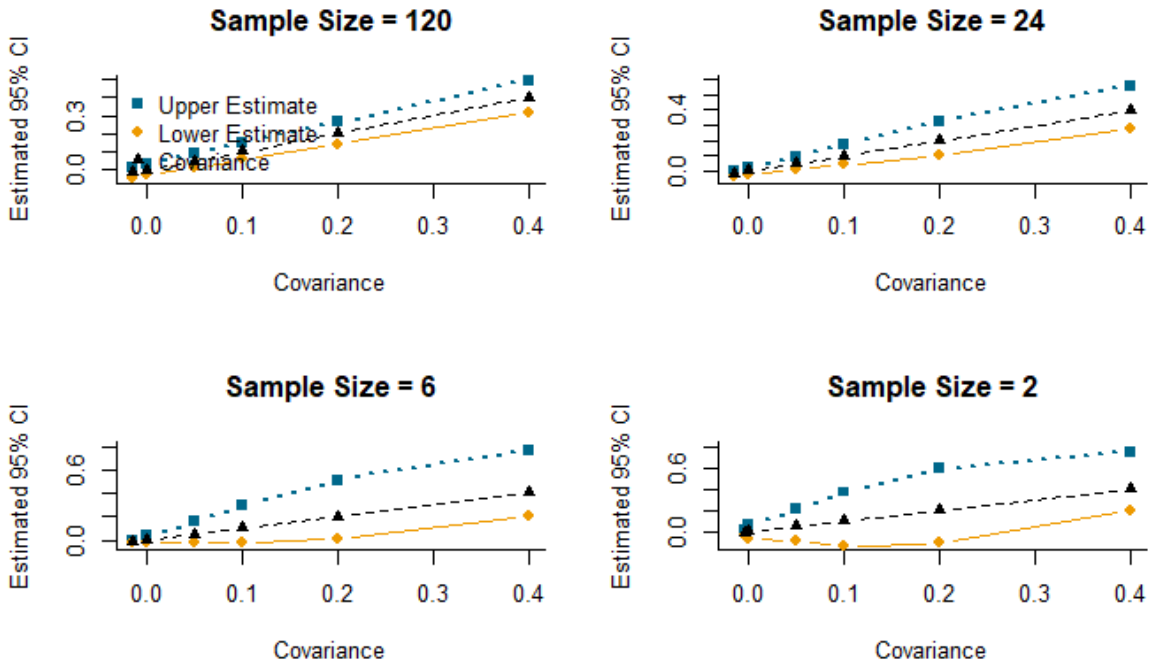
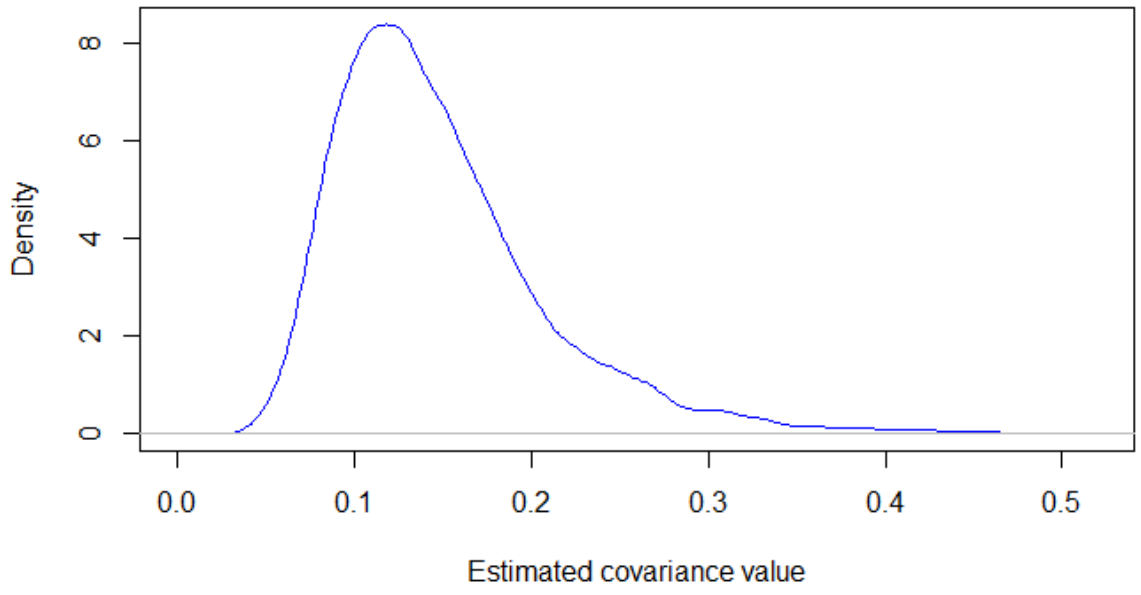


Figure A3. Density plot of estimated covariances of the third factor (which models the effect of hand on the data) across replications



Appendix B

Tables of Performance Statistics

Table B1. Root mean squared errors (RMSE) (averaged over 10,000 replications) for different sample sizes and covariances for the BCSM and LME

Sample size and model	Covariances					
	.4	.2	.1	.05	0	-.015
120						
LME	.036	.024	.018	.015	.002	.016
BCSM	.036	.024	.018	.016	.013	.012
24						
LME	.056	.044	.025	.015	.001	.015
BCSM	.056	.045	.026	.016	.009	.005
6						
LME	.100	.083	.050	.025	.002	.015
BCSM	.106	.092	.058	.031	.012	.003
2						
LME	.105	.109	.073	.060	.010	.005 [a]
BCSM	.102	.115	.082	.048	.021	.007 [a]

[a] Here, the covariance value was -.005 instead of -.015.

Table B2. (Posterior) Standard deviations (SD) (averaged over 10,000 replications) for different covariances for the BCSM.

Sample size	Covariances					
	.4	.2	.1	.05	0	-.015
120	.046	.031	.024	.020	.016	.015
24	.074	.059	.034	.022	.011	.007
6	.162	.140	.088	.047	.018	.004
2	.157	.181	.130	.077	.033	.010 [a]

[a] Here, the covariance value was -.005 instead of -.015.

Table B3. Recovery rate of the (posterior) credible intervals (CI) (averaged over 10,000 replications) for different covariances for the BCSM

Sample size	Covariances					
	.4	.2	.1	.05	0	-.015
120	.949	.948	.948	.947	.949	.949
24	.949	.949	.948	.949	.949	.950
6	.946	.949	.951	.951	.956	.961
2	.949	.966	.970	.971	.970	.817 [a]

[a] Here, the covariance value was -.005 instead of -.015.

Table B4. The 95% (posterior) credible intervals (CI) (averaged over 10,000 replications) for different covariances for the BCSM

Sample size and estimate	Covariances					
	.4	.2	.1	.05	0	-.015
120						
Lowerbound	.317	.144	.057	.014	-.030	-.043
Upperbound	.493	.263	.148	.090	.033	.016
24						
Lowerbound	.278	.101	.043	.014	-.020	-.028
Upperbound	.558	.326	.172	.096	.024	-.001
6						
Lowerbound	.201	.017	-.018	-.014	-.027	-.023
Upperbound	.764	.513	.296	.156	.040	-.007
2						
Lowerbound	.203	-.102	-.130	-.089	-.065	-.027 [a]
Upperbound	.750	.589	.373	.210	.064	.014 [a]

[a] Here, the covariance value was -.005 instead of -.015.

Table B5. Covariance estimates (averaged over 10,000 sub-samples) and respective posterior standard deviations (SD) for different factors

	τ_1	τ_2	τ_3	τ_4	Factor difference conditions; L and R	Between trial heterogeneity
Mean	.001	.140	.001	.760	.149	-.035
SD	.000	.001	.000	.022	.063	.037