Marie Wiberg • Dylan Molenaar • Jorge González
Ulf Böckenholt • Jee-Seon Kim
Editors

Quantitative Psychology

The 85th Annual Meeting of the Psychometric Society, Virtual

Springer

Editors
Marie Wiberg
Department of Statistics, USBE
Umeå University
Umeå, Västerbottens Län, Sweden

Jorge González
Facultad de Matemáticas
Pontificia Universidad Católica de Chile
Santiago, Chile

Jee-Seon Kim
Department of Educational Psychology
University of Wisconsin-Madison
Madison, WI, USA

Dylan Molenaar
Department of Psychology
University of Amsterdam
Amsterdam, The Netherlands

Ulf Böckenholt
Kellogg School of Management
Northwestern University
Evanston, IL, USA

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-030-74771-8 ISBN 978-3-030-74772-5 (eBook)
https://doi.org/10.1007/978-3-030-74772-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

A Bayesian Solution to Non-convergence of Crossed Random Effects Models

Mingya Huang and Carolyn Anderson

1 Introduction

Crossed random effects models (CREMs) have become the method of choice in studies in which every subject sees every stimulus and every stimulus is viewed by every subject (Baayen et al., 2008). Researchers often encounter a non-convergence problem when fitting CREMs with Maximum likelihood based methods (MLE/REML) because of the complexity of random effects structure and small sample sizes. A common strategy is to simplify models (i.e., using random intercepts only). We conducted an informal survey of articles from the Journal of Memory and Language from 2015 to 2019 citing Baayen et al. (2008) paper, and found that 43% of these articles utilizing CREMs do not include random slopes and/or removed them to achieve convergence. However, improper model structure will impact the parameter estimates as well as their standard errors. Under-parameterization of the covariance structure invalidates inference, and over-parameterization of the covariance structure leads to inefficient estimation (Molenberghs and Verbeke, 2000). If random slopes are removed from a level, the variance(s) related to that level will be redistributed to other levels and therefore result in inaccurate standard errors (Snijders, 2011). Similarly, omitting incorrect fixed effect structures will also lead to incorrect estimates for both random and fixed effects (Raudenbush and Bryk, 2002). To achieve valid inferences, Barr et al. (2013) proposed the maximal model structure for confirmatory factor analysis with every possible random effects rather than simplifying the models so long as the

M. Huang (✉)
University of Wisconsin-Madison, Madison, WI, USA
e-mail: mhuang233@wisc.edu

C. Anderson
University of Illinois Urbana-Champaign, Champaign, IL, USA
e-mail: cja@illinois.edu

M. Wiberg et al. (eds.), *Quantitative Psychology*, Springer Proceedings in Mathematics & Statistics 353, https://doi.org/10.1007/978-3-030-74772-5_27

design justifies. Bates et al. (2015) also argued that the models should not be too simple or too complex, but just right for optimal statistical inference. An advantage of appropriate modeling of the covariance structure is that it can help explain the random variability captured by the fixed effects.

Estimators from both MLE and REML, two typical methods fitting hierarchical linear models, are both consistent and efficient, but these estimation methods often fail to converge as models become more complex (Snijders, 2011). Unlike MLE and REML, Bayesian approaches can be more flexible when dealing with complex models such as CREMs (Snijders, 2011). Therefore, we investigated whether using a Bayesian method is an efficient alternative that can solve non-convergence problems.

2 Crossed-Random Effects Model (CREMs)

CREMs are used to fit the hierarchical data where units are simultaneously nested in multiple types of clusters (Cho & Rabe-Hesketh, 2011). In psycholinguistic research, there is usually one observation/trial per cell in a design crossed by subject and stimuli. An example of a two-level CREM for this type of cross-classification designs is

$$Y_{ij} = \gamma_{00} + \sum_p \gamma_{0p} x_{pi} + \sum_q \gamma_{q0} z_{qj} + U_{0i} + \sum_p U_{pi} x_{pi} + W_{0j} + \sum_q W_{qj} z_{qj} + R_{ij}, \tag{1}$$

3 Bayesian Approach

In Bayesian estimation, samples of parameter estimates are drawn from their posterior distribution, which are proportional to the product of a marginal probability of the parameter and the conditional probability of the data given the parameters. Let $\boldsymbol{\theta}$ be a vector of model parameters and $\boldsymbol{Y}$ represent data. The posterior distribution of parameters conditional on data is

$$f\left(\boldsymbol{\theta}|\boldsymbol{Y}\right) \propto f\left(\boldsymbol{Y}|\boldsymbol{\theta}\right) f\left(\boldsymbol{\theta}\right), \tag{2}$$

where $f(\boldsymbol{Y}|\boldsymbol{\theta})$ is the likelihood, and $f(\boldsymbol{\theta})$ is the prior distribution which reflects our preceding knowledge of the parameters. We set non-informative priors for both fixed effects (i.e., $N(0, 100)$ which is essentially flat), and variance of the random effects (i.e., Cauchy(0, 5)) based on recommendations from (Gelman, 2006). In simple cases, the posterior distribution can be found analytically (e.g., proportion from a binomial distribution), but for more complex cases, Markov Chain Monte

Carlo (MCMC) is used to sample from the posterior. A Markov chain is a sequence of draws of random variables for which the probability depends only on the previous variable. The sequence of possible estimates for each parameter is known as a "chain," with multiple chains typically run for each parameter. We use Hamiltonian algorithms to iteratively sample the posterior distribution, which is implemented in the brms package (Bürkner, 2017) in R (3.6.2), that function as a wrapper for Stan (Carpenter et al., 2017). Convergence was based on the potential scale reduction factor ($\hat{R}$), which estimates the potential decrease in the between-chains variability relative to the within-chain variability. We expect $\hat{R}$ to be close to 1 at convergence, and Gelman and Rubin (1992) suggests 1.1 as the cutoff value. We also checked the plots of posterior densities, trace plots, and autocorrelation plots to evaluate convergence. The number of Markov chains was set to be 4 with 8,000 iterations per chain where the first 4,000 iterations were warm-ups. The chains were "thinned," a procedure that keeps every k_{th} sample (parameter estimate). We retained every 10th sample, and thus the posterior is only based on 400 sample values for each chain.

4 Simulation

A simulation study with 20 replications was conducted to evaluate the performance of MLE, REML, and Bayesian estimation of CREMs. Data were simulated from CREMs with two or four random slopes and 20 or 50 stimuli and subjects, yielding four conditions (Table 1). For each of the four simulated conditions, we fit an under-specified model (only random intercepts) and the correctly specified model (used to simulate data) using each estimation method. The simplest model fit to data in this study was the random intercepts model,

$$\begin{aligned} Y_{ij} = \gamma_{00} + \gamma_{01}x_{1i} + \gamma_{10}z_{1j} + \gamma_{02}x_{2i} + \gamma_{20}z_{2j} + \gamma_{03}x_{3i} + \gamma_{30}z_{3j} + \gamma_{04}x_{4i} \\ + U_{0i} + W_{0j} + R_{ij}. \end{aligned} \tag{3}$$

The model with two random slopes (i.e. condition 1 and 3) was

$$\begin{aligned} Y_{ij} = \gamma_{00} + \gamma_{01}x_{1i} + \gamma_{10}z_{1j} + \gamma_{02}x_{2i} + \gamma_{20}z_{2j} + \gamma_{03}x_{3i} + \gamma_{30}z_{3j} + \gamma_{04}x_{4i} \\ + U_{0i} + U_{1i}x_{1i} + W_{0j} + W_{1j}z_{1j} + R_{ij}. \end{aligned} \tag{4}$$

Table 1 Summary of four conditions of simulation study

Conditions	# Subject × # Stimuli	# Random Slopes
Condition 1	20 × 20	2 slopes
Condition 2	20 × 20	4 slopes
Condition 3	50 × 50	2 slopes
Condition 4	50 × 50	4 slopes

Table 2 Models parameters for the fixed effects (left) and distributions (right) from which random effects were drawn for i,j=20,50 on each replication of the simulation study

Fixed effects parameters		Random effects parameters	
Subject-specific	Stimuli-specific	Subject-specific	Stimuli-specific
γ_{0p}	γ_{q0}	U_{ri}	W_{rj}
$\gamma_{00} = 1.00$			
$\gamma_{01} = -2.00$	$\gamma_{10} = 2.00$	$U_{0i} \sim N(0, .16)$,	$W_{0i} \sim N(0, .49)$
$\gamma_{02} = .30$	$\gamma_{20} = 2.00$	$U_{1i} \sim N(0, .04)$,	$W_{1i} \sim N(0, .16)$
$\gamma_{03} = .50$	$\gamma_{30} = .20$	$U_{2i} \sim N(0, .36)$,	$W_{2i} \sim N(0, .81)$
$\gamma_{04} = 1.00$			
		$R_{ij} \sim N(0, 1.00)$	

For the more complex case of 4 random slopes (i.e. condition 2 and 4) was

$$Y_{ij} = \gamma_{00} + \gamma_{01}x_{1i} + \gamma_{10}z_{1j} + \gamma_{02}x_{2i} + \gamma_{20}z_{2j} + \gamma_{03}x_{3i} + \gamma_{30}z_{3j} + \gamma_{04}x_{4i} + U_{0i} + U_{1i}x_{1i} + U_{2i}x_{2i} + W_{0j} + W_{1j}z_{1j} + W_{2j}z_{2j} + R_{ij}. \quad (5)$$

Data from models (4) and (5) were simulated for i, j=20, 50. For each replication of the simulated models, values for x_{pi} and z_{qj} were drawn from the following distributions: $x_{1i} \sim N(0, 2.00)$; $x_{2i} \sim N(0, 3.00)$; $x_{3i} \sim N(0, 1.25)$; $x_{4i} \sim$ Bernoulli (0.1); $z_{1j} \sim N(0, 1.75)$; $z_{2j} \sim N(0, 2)$; and $z_{3j} \sim N(0, 2.25)$. The fixed effects parameters and the distributions for random effects are given in Table 2.

5 Results

5.1 Convergence Rate

Table 3 summarized the convergence rates of each condition out of 20 replications. All models using a Bayesian approach yield convergence rates of 100%. In Condition 1 and MLE, 70% of the under-specified models converged and only 10% of the correctly specified models converged. For REML, 82.5% of the under-specified models converged and only 15% of the correctly specified models converged. In Condition 2, both MLE and REML obtained 100% convergence rates for the under-specified models; however, they both failed to converge even with different optimizers such as NEALTHER-MEAD and BOBYQA in all replications when trying to fit the correctly specified model. In Condition 3, the under-specified models fit by either MLE and REML converged in all cases; however, the convergence rates of the correctly specified model for MLE and REML were only 40% and 45%, respectively. In Condition 4, 100% of the under-specified models converged for both MLE and REML. However, only 20% of the correctly specified models fit by MLE converged, and only 32.5% for REML of the correctly specified models converged.

Table 3 Convergence rates of under-specified models (only random intercepts) and the correctly specified models

	#Subjects ×	# Random	Model			
Conditions	#Stimuli	slopes	fit	MLE	REML	Bayesian
Condition 1	20 × 20	2	Under-specified	70%	82.5%	100%
			Correctly specified	10%	15%	100%
Condition 2	20 × 20	4	Under-specified	100%	100%	100%
			Correctly specified	0%	0%	100%
Condition 3	50 × 50	2	Under-specified	100%	100%	100%
			Correctly specified	40%	45%	100%
Condition 4	50 × 50	4	Under-specified	100%	100%	100%
			Correctly specified	20%	32.5%	100%

With 100% convergence rates, the results indicate that a Bayesian approach is a viable alternative of MLE/REML to deal with convergence problems. Note that as the random effects structure of the model used to simulate data became more complex, it was less likely for the correctly specified model to converge with either MLE or REML. As the number of subjects and stimuli increased, the under-specified model using MLE/REML was more likely to converge. In addition, REML yielded a higher convergence rate than MLE in some conditions, suggesting REML as a useful alternative when MLE encounters non-convergence.

5.2 *Parameter Recovery*

We discuss the efficiency and validity of the Bayesian parameter estimates. Tables 4 and 5 summarize the mean of Bayesian estimates, 95% credible intervals, the scale reduction factor $\hat{R}$, root mean squared error (RMSE), and bias for the 20 replications. In Table 4, the $\hat{R}$s are less than 1.20, indicating convergence. The fixed effects estimates are similar to the values used to simulate the data in both 20 × 20 and 50 × 50 cases. In contrast, the random effects estimates in the under-specified model deviate from the values used to generate the data while the correctly specified model yield similar variance estimates which are close to the true ones. The RMSEs and biases are smaller in the correctly specified models. As the sample size increases from 20 × 20 to 50 × 50, the estimated values become closer to the true values.

Similarly, in Table 5, the fixed effects estimates are close to the true values and are more accurate in the correct model than the under-specified model. For the random effects, the under-specified models yield the variance estimates that deviate from the true values, but they are similar in the correct model. These results are supported by smaller RMSEs and biases for the correctly specified model, with no discernible pattern in the biases. Comparing Tables 4 and 5, we find that as the model become more complex (from two to four slopes), the bias and the RMSEs for random effects also increase. Additionally, the deviations between the estimates and true values are

Table 4 Bayesian parameter estimates, 95% credible intervals, intervals widths, $\hat{R}$ values, RMSE and bias for models fit to simulated data for conditions with 20 subjects × 20 stimuli (top) and 50 × 50 (bottom) and CREMs with 2 slopes

Model		True		95%Credible Intervals		Intervals	$\hat{R}$			
Fit	Param	Value	Est.	Lower	Upper	Width	Min	Max	RMSE	Bias
				Condition 1:	20 × 20					
Under-specified	γ_{00}	1.00	0.96	0.35	1.58	1.23	1.00	1.00	0.26	−0.04
	γ_{01}	−2.00	−2.00	−2.17	−1.84	0.33	1.00	1.00	0.11	−0.00
	γ_{10}	2.00	2.00	1.68	2.32	0.64	1.00	1.00	0.16	−0.00
	γ_{02}	0.30	0.30	0.19	0.41	0.22	1.00	1.00	0.13	−0.00
	γ_{20}	0.20	0.22	−0.05	0.50	0.55	1.00	1.00	0.03	0.02
	γ_{03}	0.50	0.50	0.24	0.76	0.52	1.00	1.00	0.08	−0.00
	γ_{30}	1.50	1.47	1.23	1.72	0.51	1.00	1.00	0.14	−0.03
	γ_{04}	1.00	0.81	−0.19	1.81	2.00	1.00	1.00	0.59	−0.19
	$var(U0)$	0.14	0.35	0.14	0.81	0.67	1.00	1.00	0.25	−0.21
	$var(W0)$	0.04	1.14	0.52	2.42	1.90	1.00	1.00	0.79	−0.53
	$var(Rij)$	1.00	1.14	0.93	1.08	0.15	1.00	1.00	0.46	−0.14
Correctly specified	γ_{00}	1.00	0.99	0.41	1.56	1.15	1.00	1.00	0.25	−0.01
	γ_{01}	−2.00	−1.98	−2.20	−1.75	0.45	1.00	1.01	0.10	0.02
	γ_{10}	2.00	2.01	1.58	2.43	0.85	1.00	1.00	0.17	0.01
	γ_{02}	0.30	0.30	0.20	0.41	0.21	1.00	1.00	0.03	0.00
	γ_{20}	0.20	0.22	−0.04	0.48	0.52	1.00	1.00	0.11	0.02
	γ_{03}	0.50	0.50	0.25	0.75	0.50	1.00	1.00	0.07	−0.00
	γ_{30}	1.50	1.48	1.24	1.71	0.47	1.00	1.00	0.12	−0.02
	γ_{04}	1.00	0.82	−0.25	1.90	2.10	1.00	1.01	0.65	−0.18
	$var(U0)$	0.14	0.18	0.01	0.62	0.61	1.00	1.00	0.09	−0.04
	$var(U1)$	0.04	0.06	0.00	0.30	0.30	1.00	1.02	0.04	−0.02
	$var(W0)$	0.48	0.51	0.05	1.69	1.64	1.00	1.03	0.23	−0.03
	$var(W1)$	0.16	0.27	0.01	1.12	1.11	1.00	1.02	0.18	−0.11
	$var(Rij)$	1.00	1.00	0.93	1.05	0.12	1.00	1.00	0.03	0.00

	γ_{10}	2.00	1.96	1.75	2.16	0.41	1.00	1.00	0.10	−0.04
				Condition 3:	50 × 50					
Under-specified	γ_{00}	1.00	1.03	0.70	1.36	0.66	1.00	1.00	0.18	0.03
	γ_{01}	−2.00	−1.99	−2.07	−1.90	0.08	1.00	1.00	0.06	0.01
	γ_{10}	2.00	1.97	1.80	2.15	0.35	1.00	1.00	0.10	−0.03
	γ_{02}	0.30	0.30	0.25	0.36	0.11	1.00	1.00	0.03	0.00
	γ_{20}	0.20	0.20	0.07	0.34	0.27	1.00	1.00	0.07	0.00
	γ_{03}	0.50	0.50	0.36	0.63	0.27	1.00	1.00	0.06	−0.00
	γ_{30}	1.50	1.49	1.36	1.62	0.26	1.00	1.01	0.05	−0.01
	γ_{04}	1.00	1.05	0.51	1.59	1.08	1.00	1.00	0.29	0.04
	$var(U0)$	0.14	0.32	0.20	0.50	0.30	1.00	1.01	0.18	−0.17
	$var(W0)$	0.04	0.99	0.64	1.49	0.85	1.00	1.00	0.55	−0.82
	$var(Rij)$	1.00	0.99	0.97	1.02	0.05	1.00	1.00	0.28	0.00
Correctly specified	γ_{00}	1.00	1.03	0.73	1.33	0.60	1.00	1.00	0.15	0.03
	γ_{01}	−2.00	−1.98	−2.08	−1.87	0.21	1.00	1.00	0.05	0.02
	γ_{02}	0.30	0.30	0.25	0.35	0.10	1.00	1.00	0.03	0.00
	γ_{20}	0.20	0.21	0.08	0.33	0.24	1.00	1.00	0.06	0.01
	γ_{03}	0.50	0.50	0.37	0.62	0.25	1.00	1.00	0.05	−0.00
	γ_{30}	1.50	1.49	1.37	1.60	0.23	1.00	1.00	0.05	−0.01
	γ_{04}	1.00	1.07	0.57	1.56	0.99	1.00	1.01	0.26	0.0
	$var(U0)$	0.14	0.17	0.07	0.35	0.28	1.00	1.00	0.04	−0.02
	$var(U1)$	0.04	0.04	0.00	0.11	0.11	1.00	1.00	0.02	−0.00
	$var(W0)$	0.48	0.53	0.24	1.00	0.76	1.00	1.00	0.16	−0.08
	$var(W1)$	0.16	0.18	0.02	0.44	0.42	1.00	1.01	0.11	−0.01
	$var(Rij)$	1.00	1.00	0.97	1.00	0.03	1.00	1.00	0.01	−0.01

Table 5 Bayesian parameter estimates, 95% credible intervals, intervals widths, $\hat{R}$ values, RMSE and bias for models fit to simulated data for conditions with 20 subjects × 20 stimuli (top) and 50 × 50 (Bottom) and CREMs with 4 slopes

Model		True		95%Credible intervals		Intervals	$\hat{R}$			
fit	Param	value	Est.	Lower	Upper	Width	Min	Max	RMSE	Bias
				Condition 2:	20 × 20					
Under-specified	γ_{00}	1.00	1.15	−0.26	2.56	2.82	1.00	1.00	0.59	0.15
	γ_{01}	−2.00	−2.01	−2.44	−1.57	0.87	1.00	1.00	0.20	−0.01
	γ_{10}	2.00	2.02	1.33	2.71	1.38	1.00	1.00	.35	0.02
	γ_{02}	0.30	0.37	0.09	0.66	0.57	1.00	1.00	0.19	0.07
	γ_{20}	0.20	0.18	−0.38	0.75	1.13	1.00	1.00	0.37	−0.02
	γ_{03}	0.50	0.57	−0.12	1.24	1.36	1.00	1.00	0.31	0.07
	γ_{30}	1.50	1.50	0.99	2.01	1.02	1.00	1.00	0.19	0.00
	γ_{04}	1.00	0.57	−2.06	3.21	5.27	1.00	1.00	1.97	−0.44
	var(U0)	0.14	2.78	1.23	5.75	4.52	1.00	1.00	3.03	−2.64
	var(W0)	0.04	5.24	2.51	10.83	8.32	1.00	1.00	5.05	−4.77
	var(Rij)	1.00	1.00	0.93	1.08	0.15	1.00	1.00	0.03	0.00
Correctly specified	γ_{00}	1.00	1.26	0.16	2.35	2.19	1.00	1.00	0.43	0.26
	γ_{01}	−2.00	−2.00	−2.46	−1.54	0.92	1.00	1.00	0.19	−0.00
	γ_{10}	2.00	2.06	1.38	2.74	1.36	1.00	1.00	0.26	0.06
	γ_{02}	0.30	0.37	0.01	0.73	0.72	1.00	1.01	0.15	0.07
	γ_{20}	0.20	0.13	−0.57	0.82	0.25	1.00	1.00	0.24	−0.07
	γ_{03}	0.50	0.54	−0.04	1.12	1.16	1.00	1.00	0.27	0.04
	γ_{30}	1.50	1.55	1.12	1.96	0.84	1.00	1.00	0.17	0.05
	γ_{04}	1.00	0.62	−2.24	3.54	5.78	1.00	1.01	1.67	−0.38
	var(U0)	0.14	0.45	0.00	2.51	2.51	1.00	1.00	0.44	−0.31
	var(U1)	0.04	0.14	0.00	0.99	0.99	1.00	1.01	0.14	−0.10
	var(U2)	0.30	0.25	0.02	0.85	0.83	1.00	1.00	0.12	0.05
	var(W0)	0.48	0.62	0.00	3.93	3.93	1.00	1.00	0.41	−0.14
	var(W1)	0.16	0.46	0.00	2.44	2.44	1.00	1.00	0.41	−0.30
	var(W2)	0.81	0.97	0.11	3.02	2.91	1.00	1.00	0.42	−0.16
	var(Rij)	1.00	1.00	0.93	1.08	0.15	1.00	1.00	0.03	0.00

	γ_{10}	2.00	1.97	1.65	2.26	0.61	1.00	1.00	0.17	−0.03
				Condition 4:	50 × 50					
Under-specified	γ_{00}	1.00	1.11	0.28	1.93	1.65	1.00	1.00	0.53	0.11
	γ_{01}	−2.00	−1.99	−2.25	−1.73	0.52	1.00	1.00	0.11	0.01
	γ_{10}	2.00	1.91	1.54	2.28	0.74	1.00	1.00	0.23	−0.09
	γ_{02}	0.30	0.35	0.18	0.52	0.34	1.00	1.00	0.17	0.05
	γ_{20}	0.20	0.14	−0.16	0.44	0.60	1.00	1.00	0.26	−0.06
	γ_{03}	0.50	0.50	0.08	0.91	0.82	1.00	1.00	0.17	−0.00
	γ_{30}	1.50	1.52	1.25	1.79	0.54	1.00	1.00	0.11	0.02
	γ_{04}	1.00	0.90	−0.75	2.54	3.29	1.00	1.00	1.00	−0.10
	var(U0)	0.14	3.19	2.02	4.72	2.70	1.00	1.00	3.33	−3.04
	var(W0)	0.04	4.72	3.07	7.02	3.95	1.00	1.00	4.55	−4.27
	var(Rij)	1.00	1.00	0.97	1.02	0.05	1.00	1.00	0.12	−0.01
Correctly specified	γ_{00}	1.00	1.07	0.57	1.59	1.02	1.00	1.00	0.31	0.07
	γ_{01}	−2.00	−1.98	−2.18	−1.79	0.39	1.00	1.00	0.09	0.02
	γ_{02}	0.30	0.34	0.13	0.54	0.41	1.00	1.00	0.12	0.04
	γ_{20}	0.20	0.23	−0.12	0.58	0.70	1.00	1.00	0.17	0.03
	γ_{03}	0.50	0.52	0.25	0.79	0.54	1.00	1.00	0.11	0.02
	γ_{30}	1.50	1.51	1.34	1.69	0.35	1.00	1.00	0.06	0.01
	γ_{04}	1.00	0.94	−0.26	2.12	2.38	1.00	1.00	0.56	−0.06
	var(U0)	0.14	0.24	0.01	0.93	0.92	1.00	1.00	0.18	−0.09
	var(U1)	0.04	0.05	0.00	0.24	0.24	1.00	1.00	0.04	−0.01
	var(U2)	0.30	0.32	0.13	0.59	0.46	1.00	1.00	0.08	0.00
	var(W0)	0.48	0.31	0.01	1.28	1.27	1.00	1.00	0.23	0.14
	var(W1)	0.16	0.20	0.01	0.68	0.67	1.00	1.01	0.20	−0.03
	var(W2)	0.81	0.95	0.42	1.77	1.35	1.00	1.00	0.26	−0.16
	var(Rij)	1.00	1.00	0.97	1.02	0.05	1.00	1.01	0.12	−0.01

larger in the under-specified models than in the correctly-specified models in all conditions.

Overall, for both under-specified and correctly specified models, the fixed effect parameters were well recovered when using Bayesian estimation. However, differences were found with respect to the 95% credible intervals for the fixed effects. The intervals for the correctly specified models were narrower than the ones from the under-specified model, which was even more prominent with larger sample sizes. Similarly, the random effects variance parameters were recovered better in the correctly-specified models. Also, only the 95% credible intervals in the correctly specified models covered the true values used to simulate the data. The variance estimates for under-specified models have poor performance, and variance parameters were over-estimated such that the 95% credible intervals did not cover the true values used to simulate the data. For correctly specified models, the 95% credible intervals were narrowed for larger sample sizes (and given model complexity) and were narrower for simpler models (for given sample size).

6 Conclusion

Although some previous studies have examined the convergence problems of random effects models and promote a Bayesian approach as a solution (Eager & Roy, 2017), none have specifically considered CREMs. This study is the first to do so, and provides solid evidence for a Bayesian approach when fitting the CREMs to data over MLE/REML. Comparing convergence rates of MLE/REML and Bayesian approaches, the latter obtained 100% convergence rates ($\hat{R}s < 1.1$). As the model became more complex with more random effects, the convergence rates decreased under MLE/REML. Furthermore, the Bayesian estimates of both fixed effects and random effects were valid and efficient in the correctly-specified models but not in the under-specified models. This study highlighted three important points: (1) an improper model structure will result in inefficient estimation and invalid results (2) for more complex random effects structures, the models using Bayesian approach can achieve model convergence but not MLE/REML (3) using Bayesian approach to fit the CREMs can obtain efficient estimates. Future studies will explore whether using a Bayesian approach can select an optimal model.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv preprint arXiv:1506.04967

Bürkner, P. C. (2017). Advanced Bayesian multilevel modeling with the R package brms. arXiv preprint arXiv:1705.11123

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*, 1

Cho, S. J., & Rabe-Hesketh, S. (2011). Alternating imputation posterior estimation of models with crossed random effects. *Computational Statistics & Data Analysis, 55*, 12–25

Eager, C., & Roy, J. (2017). Mixed effects models are sometimes terrible. arXiv preprint arXiv:1701.04858

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis, 1*, 515–534

Gelman, A., Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–472

Molenberghs, G., & Verbeke, G. (2000). *Linear mixed models for longitudinal data*. Springer

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (vol. 1). Sage

Snijders, T. (2011). *Multilevel analysis* (pp. 879–882). Heidelberg: Springer