

Modeling Heterogeneity of the Level-1 Error Covariance Matrix in Multilevel Models for Single-Case Data

Eunkyeng Baek^a, John J. M. Ferron^b

[a] *Department of Educational Psychology, Texas A&M University, TX, USA.* [b] *Department of Educational and Psychological Studies, University of South Florida, FL, USA.*

Methodology, 2020, Vol. 16(2), 166–185, <https://doi.org/10.5964/meth.2817>

Received: 2018-09-25 • **Accepted:** 2020-02-05 • **Published (VoR):** 2020-06-18

Corresponding Author: Eunkyeng Baek, Texas A&M University, Department of Educational Psychology, 4225 TAMU, College Station, TX 77843-4225. E-mail: baek@tamu.edu

Abstract

Previous research applying multilevel models to single-case data has made a critical assumption that the level-1 error covariance matrix is constant across all participants. However, the level-1 error covariance matrix may differ across participants and ignoring these differences can have an impact on estimation and inferences. Despite the importance of this issue, the effects of modeling between-case variation in the level-1 error structure had not yet been systematically studied. The purpose of this simulation study was to identify the consequences of modeling and not modeling between-case variation in the level-1 error covariance matrices in single-case studies, using Bayesian estimation. The results of this study found that variance estimation was more sensitive to the method used to model the level-1 error structure than fixed effect estimation, with fixed effects only being impacted in the most extreme heterogeneity conditions. Implications for applied single-case researchers and methodologists are discussed.

Keywords

single-case, multilevel modeling, Bayesian estimation, misspecifying level-1 error structure, heterogeneity

Single-case designs facilitate the study of intervention effects at the level of the individual by collecting for each individual repeated observations in each of at least two conditions (e.g., baseline and treatment). The most common type of single-case design is the multiple-baseline design (Shadish & Sullivan, 2011), in which multiple individuals are studied concurrently in a baseline followed by a treatment phase with the restriction that the start of intervention is temporally staggered across the individuals so their baselines have different lengths. Because single-case designs allow researchers to study individuals from populations with low prevalence rates, they have been widely employed



in disciplines like special education, school psychology, applied behavior analysis, and rehabilitation counseling.

The application of statistical models to single-case data requires researchers to make assumptions about the errors in the statistical model. It is difficult to assume independence, because the unmeasured factors that give rise to the errors may impact multiple adjacent observations, such as when a child's illness impacts their behavior several days in a row. When the errors that are closer in time are more similar than errors further apart in time, there is some serial dependency or autocorrelation among the errors. Several autoregressive and moving average models have been considered for the dependency among errors in single-case regression models, such as unstructured, compound symmetry, banded toeplitz or moving average, first-order autoregressive (AR[1]) or independent (σ^2I ; [Wolfinger, 1993](#)), and the most commonly assumed correlated error structure in SCED data is the AR(1) model (i.e., $e_t = \rho e_{t-1} + a_t$; [Pustejovsky, Hedges, & Shadish, 2014](#)).

If positive autocorrelation is ignored, the regression coefficients of single-level models are unbiased, but the standard errors of the regression coefficients would be underestimated ([Neter, Wasserman, & Kutner, 1990](#)) which inflates Type I error rates for significance tests of the treatment effect ([Toothaker, Banz, Noble, Camp, & Davis, 1983](#)). These negative effects of ignoring autocorrelation motivate single-level models that estimate and adjust for autocorrelation (e.g., [McKnight, McKean, & Huitema, 2000](#); [Maggin et al., 2011](#)), but with short series it is difficult to disentangle autocorrelation from trends and estimating both can compound short series estimation bias ([Ferron, 2002](#)). To obtain negligible bias in the estimate of autocorrelation, over 100 observations would be needed using a traditional estimator, or at least 20 observations using a jackknife estimator ([Huitema & McKean, 1994](#)), which makes a challenge in single-level models for single-case studies, which typically have baselines that are less than 20 observations.

Multilevel Modeling in Single-Case Study

When single-case studies involve multiple cases, like in the multiple-baseline design, multilevel modeling (MLM) has been suggested as a method for analyzing the data from multiple cases. MLM allows for possible dependency of the errors to be taken into account ([Baek et al., 2014](#); [Shadish, Rindskopf, Hedges, & Sullivan, 2013](#); [Van den Noortgate & Onghena, 2003](#)). The performance of the MLM approach to analyze single-case data has been examined using several simulation studies. It is generally known that the fixed effects are unbiased, but the variance components have small sample bias ([Ferron, Bell, Hess, Rendina-Gobioff, & Hibbard, 2009](#); [Ferron, Farmer, & Owens, 2010](#); [Moeyaert, Ugille, Ferron, Beretvas, & Van den Noortgate, 2013](#); [Petit-Bois, Baek, Van den Noortgate, Beretvas, & Ferron, 2016](#)).

The following Equations (1) and (2) are for a two-level model for single-case studies.

Level-1 equation:

$$Y_{ij} = \beta_{0j} + \beta_{1j}Phase_{ij} + \beta_{2j}time_{ij} + \beta_{3j}Phase_{ij} * Time'_{ij} + e_{ij} \text{ and } e_{ij} \sim N(0, \Sigma_e) \quad (1)$$

Level-2 equation:

$$\begin{cases} \beta_{0j} = \theta_{00} + u_{0j} \\ \beta_{1j} = \theta_{10} + u_{1j} \\ \beta_{2j} = \theta_{20} + u_{2j} \\ \beta_{3j} = \theta_{30} + u_{3j} \end{cases} \text{ with } \begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \\ u_{3j} \end{bmatrix} \sim N(0, \Sigma_u) \quad (2)$$

where y_{ij} is the observed value (outcome) at the i th observation for the j th case. $Phase_{ij}$ is a dichotomous variable that indicates the phase in which the observation occurred, being 0 indicates the baseline phase and 1 indicates the treatment phase; $time_{ij}$ is an indicator of time coded such that 0 corresponds to the first observation of the study, and $Time'_{ij}$ is $time_{ij}$ recoded such that 0 corresponds to the first intervention observation (Huitema & McKean, 2000). Thus, β_{0j} is the baseline intercept for the j th case at the beginning of the baseline phase, and β_{1j} is the difference between the baseline level and the treatment level (shift in level) for the j th case at the beginning of treatment, β_{2j} is the baseline slope for the j th case, and β_{3j} is the change in slope that occurs with treatment. Finally, e_{ij} is the residual that indicates within case variation (level-1 errors) and is assumed to be multivariate normally distributed $N(0, \Sigma_e)$. The covariance structure Σ_e of the errors can be assumed as any of the error structures mentioned earlier. For the second level equation, θ_{00} is the average baseline intercept, θ_{10} is the average shift in level indexed at the time of the first treatment observation, θ_{20} is the average baseline slope, and θ_{30} is the average shift in slope. The level-2 errors, u_{0j} , u_{1j} , u_{2j} and u_{3j} are assumed to be multivariate normally distributed $N(0, \Sigma_u)$.

Assumption of Between Case Homogeneity in Level-1 Error Structures

The effect of ignoring autocorrelation and other misspecifications of the level-1 error structure have been examined for MLM analyses for single-case data. Researchers found that ignoring autocorrelation does not bias the fixed effect estimates, but the inferences about the fixed effects can be inaccurate due to the underestimate of the corresponding standard errors, and the estimates of the variance parameters become more biased (Ferron et al., 2009; Owens & Ferron, 2012; Petit-Bois et al., 2016).

Although MLM allows autocorrelation among level-1 errors to be taken into consideration in single-case data analyses, this approach holds a critical assumption that the level-1 error structure is the same for all cases. Specifically, it is assumed that (a) autocorrelation is the same for all cases and (b) the level-1 error variance is the same for all cases. Previous single-case research using MLM application as well as misspecification research of level-1 error structures has often assumed the autocorrelation and level-1

error variance to be equal for all cases (e.g., Ferron et al., 2010; Petit-Bois et al., 2016; Van den Noortgate & Onghena, 2003).

However, it is possible that the level-1 error structure may not be homogeneous across cases. For example, the behavior of one child with an emotional/behavioral disorder may vary more substantially from day to day because of intermittent problems in the home, or lapses in medication provision, than that of another child with the same disorder. The findings from previous studies in single-case data support that variations in level-1 error covariance matrices could exist (Baek & Ferron, 2013; Baek, Petit-Bois, Van den Noortgate, Beretvas, & Ferron, 2016). For example, Baek and Ferron (2013) discovered relatively large differences in estimates of autocorrelation and level-1 error variances, when they allowed the level-1 error structure to vary across cases in real datasets from four previously published single-case studies (i.e., Dufrene et al., 2010; Ingersoll & Lalonde, 2010; Koegel, Singh, & Koegel, 2010; Oddo et al., 2010), and the fit indices favored a model with separate estimates of the autocorrelation and the level-1 error variances for some studies. Baek et al. (2016) also found that allowing the level-1 error structure to differ for one of the participants in a single-case study led to estimated individual trajectories that were more consistent with the visually plotted data and improved the model fit.

Because in some studies the cases appear to have different variances, but commonly adopted models assume homogeneity, it is critical to examine the consequences of modeling (and not modeling) the heterogeneity. Thus, the purpose of this study is to extend the MLM modeling in single-case design to allow between case variation in the level-1 error structure, and to identify the consequences of different modeling methods for the level-1 error structure.

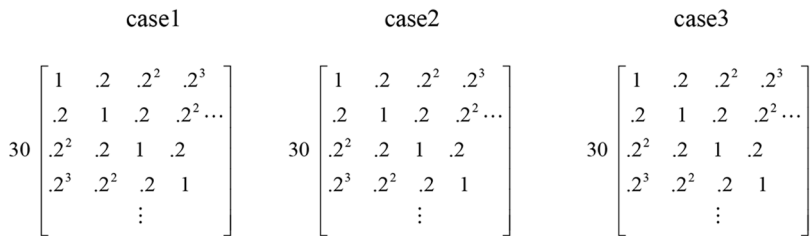
Modeling Between Case Variation in Level-1 Error Structures

The two level model that allows between case variation in the level-1 error structure in single-case design can still be represented by Equations (1) and (2). When modeling between case variation in the level-1 error structure, the covariance structure Σ_e is assumed to have parameters that vary from one case (j) to the next which can be re-expressed as Σ_{ej} . More specifically, if a AR(1) structure is chosen, then the autocorrelation and level-1 error variance are estimated separately for each case.

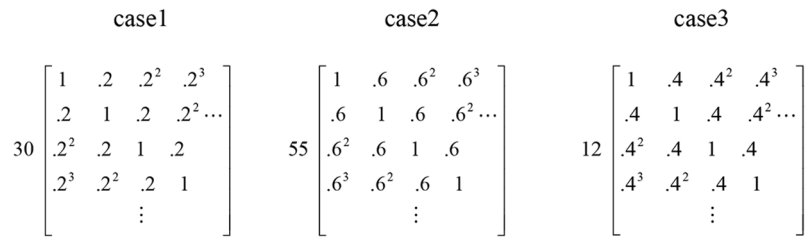
Figure 1 illustrates the results of different ways of modeling the level-1 covariance structure Σ_e . Assume that there are single-case data with three cases and the AR(1) structure is assumed for the covariance structure Σ_e . In scenario (1), the covariance structure Σ_e is assumed to be constant across cases, thus, the autocorrelation ($\rho = .2$) and the variance of level-1 errors ($\sigma^2 = 30$) are estimated and these values are same across all three cases. However, in scenario (2), for the proposed approach where the covariance structure Σ_{ej} is allowed to vary across cases, the autocorrelation and the variance of

level-1 errors are estimated with as many values as cases. Thus, each case has unique autocorrelation and level-1 error variance.

Figure 1
An Example Illustrates the Results of Different Ways of Modeling Level-1 Covariance Structure Σ_e



Scenario (1) Σ_e is assumed to be AR(1) and held constant across three cases



Scenario (2) Σ_e is assumed to be AR(1) and allowed to vary across three cases

Bayesian Method of Estimating Heterogeneous Level-1 Error Structure

Restricted maximum likelihood (REML) estimation is the most commonly used method to analyze multilevel models, and has been implemented by several software procedures that allow easy access. However, REML may encounter estimation problems, such as non-convergence, when analyzing complex multilevel models of SCED data. The Bayesian approach provides a feasible option in computationally intensive scenarios through the use of Markov Chain Monte Carlo (MCMC) procedures (e.g., Browne, 2008; Gilks, Richardson, & Spiegelhalter, 1996) which have the potential to resolve non-convergence issues of REML estimation. It was found that a complex multilevel model of SCED data that failed to converge using REML could be estimated using the Bayesian approach (Baek, Petit-Bois, & Ferron, 2013; Baek, 2015).

Bayesian inference is the process of fitting a probability model, given the observed data, and summarizing the uncertainty of the model parameters with probability distributions (Gelman, 2002). Thus, Bayesian estimation also offers practical advantages be-

cause it takes into account the uncertainty of estimating both fixed effects and variance components (Gelman, Carlin, Stern, & Rubin, 2004). In addition to the computational and practical benefits of using Bayesian estimation, recent studies have indicated that the Bayesian approach has potential benefits in estimating effect sizes, analyzing nonlinear data, and estimating autocorrelation (Baek et al., 2019; Moeyaert, Rindskopf, Onghena, & Van den Noortgate, 2017; Rindskopf, 2014a; Rindskopf, 2014b; Shadish et al., 2013; Swaminathan, Rogers, & Horner, 2014). Baek et al. (2019) examined the impact of REML and Bayesian estimation on average treatment effect inferences of SCED studies using multilevel modeling. They found that Bayesian estimation results were comparable to REML estimation results. However, Moeyaert et al. (2017) found that Bayesian estimation can lead to more precise variance component estimation than ML estimation for certain conditions. Thus, the current study was conducted using Bayesian estimation.

Method

A Monte Carlo simulation study was conducted to examine the performance of the proposed Bayesian analysis of single-case design data, which allows between case variation in the level-1 error variances and autocorrelation. Specifically, two level models where the level-1 error structures were modeled in different ways (i.e., not modeling between case variation vs. modeling between case variation) were examined in terms of the accuracy of the estimates of the parameters using Bayesian estimation. In this study, multiple baseline single-case data were generated using Equations (1) and (2) with the interactive matrix language procedure in SAS version 9.3. The true level-1 error structure was AR(1) and was either homogenous across cases, moderately heterogeneous across cases, or severely heterogeneous across cases. The AR(1) error structure was selected and generated as the true level-1 error structure because it is the most commonly applied correlated level-1 error structure in analyzing single-case data, and it is a relatively simple autocorrelated error structure (Ferron et al., 2009; Matyas & Greenwood, 1996). For the homogeneous conditions, level-1 errors were generated for each case with a level-1 error standard deviation (SD) (σ_e) of 1.0 and an autocorrelation (ρ) of .2. The value of the autocorrelation .2 had been selected based on the literature review of single-case designs (Shadish & Sullivan, 2011). For the heterogeneous conditions, values of autocorrelation (ρ_j) were generated from a normal distribution (Moderately Heterogeneous: $M = 0.2$ and $SD = 0.1$; Severely Heterogeneous: $M = 0.2$ and $SD = 0.2$), whereas the level-1 error SD (σ_{ej}) was generated from a uniform distribution (Moderately Heterogeneous: $M = 1$ and $SD = 0.17$ (lower and upper limit = 0.7, 1.3) ; Severely Heterogeneous: $M = 1$ and $SD = 0.35$ (lower and upper limit = 0.4, 1.6). The level-1 error SD were generated such that the largest level-1 error variance value can be either as much as 3.5 times or as much as 16 times the smallest level-1 error variance value. The motivation was based on the analyses of real single-case design datasets presented by Baek and Ferron (2013) and Baek et al.

(2016). The level-2 errors u_{0j} , u_{1j} , u_{2j} , and u_{3j} were generated independently from normal distributions with mean 0 and variances of either 0.5, 0.5, 0.05, and 0.05, or 2, 2, 0.2, and 0.2, respectively.

The simulation conditions varied in series length (10 or 20) and the number of cases (4 or 8), to cover the sample sizes that are typical in single case studies. In addition, conditions varied in the level-2 error variance (more or less between case variability), and the method used to analyze the data, either allowing level-1 error structure to vary across cases (Model 2, proposed model) or holding the level-1 error variance and autocorrelation constant across cases (Model 1). For each condition, 1000 data sets were simulated, and parameter relative bias, RMSE and credible interval (CI) coverage and width were estimated. The two different models (Model 1 and Model 2) were used to analyze each generated data set using OpenBUGS software. Although various software programs are available to run Bayesian estimation including SAS, OpenBUGS software was selected due to several advantages, including flexibility to handle the complexity of the models, and possible computational efficiency from using the Gibbs sampling method (Spiegelhalter, Thomas, Best, & Lunn, 2014).

Prior Distributions for the Parameters

In Bayesian estimation, various choices of priors can be assumed for each parameter. A noninformative prior distribution is considered as one of the reasonable choices of objective prior distribution because noninformative distributions make the data speak for themselves so that posterior inferences are unaffected by external information (Berger, 2006; Gelman, 2002; Gelman et al., 2004; Goldstein, 2006; Jeffreys, 1961; Morris, 1983). For multilevel models, reasonable noninformative prior distributions have been developed for the fixed effect parameters (i.e., θ_{00} , θ_{10} , θ_{20} , θ_{30}), including the noninformative normal distribution (Gelman, 2006; Gelman et al., 2004). In general, noninformative normal distributions are constructed with large variance (i.e., 1000^2), so that posterior inference is not influenced by the choice of variance value. Thus, in this study, all fixed effect parameters were assumed to follow noninformative normal distributions ($M = 0$ and $\text{Var} = 1000^2$).

Unlike priors for fixed effect parameters, priors for random effect parameters have been more difficult to construct. The choice of noninformative prior distribution for the level-2 error variances (i.e., $\sigma_{u_{0j}}^2$, $\sigma_{u_{1j}}^2$, $\sigma_{u_{2j}}^2$, $\sigma_{u_{3j}}^2$) and the level-1 error variance (σ_e^2) can have a more substantial impact on inferences, especially in the case where j (the number of units of the higher level) is small (Gelman, 2002; Gelman, 2006).

Various noninformative and weakly informative prior distributions have been suggested for the variance parameters in multilevel models, including uniform, inverse-gamma family, inverse-Wishart, half-Cauchy, and half-t distributions (Baek et al., 2019; Berger & Strawderman, 1996; Daniels & Kass, 1999; Gelman, 2006; Moeyaert et al., 2017). Moeyaert et al. (2017) constructed various weakly informative priors for variance components for

SCED data, and the results found biased and less precise variance estimates when the number of cases is small ($J = 3$). By increasing the number of cases, more precise estimates are obtained. Similarly, Gelman (2006) found that when the uniform distribution (noninformative prior) is constructed for the standard deviation (SD) unit rather than variance unit of level-2 error, the uniform distribution generally performs well, as long as $J \geq 3$ which is required to ensure a proper posterior density. Baek et al. (2019) examined the impact of estimation methods in multilevel SCED using noninformative priors for variance parameters, and their study yielded similar results with the previous study that constructed weakly informative priors. Based on these suggestions, noninformative prior distributions for the SD of the level-2 errors ($\sigma_{u_{0j}}, \sigma_{u_{1j}}, \sigma_{u_{2j}}, \sigma_{u_{3j}}$) and the level-1 errors (σ_{e_j}) were assigned to be the uniform distribution (lower limit = 0 and upper limit = 100) in this study. Since all of the level-2 and level-1 error SD values were generated to be less than 2, the upper limit of 100 seems to be large enough to minimally impact the posterior distribution.

Similarly, a noninformative prior for ρ that follows a normal distribution with $M = 0$ and $SD = 1000$ was assigned. However, since ρ is a correlation parameter, the scale of the prior distribution for ρ is stationary restricted so that its range falls between -1 and 1 (Gamerman & Lopes, 2006).

For the proposed model, priors for the level-1 error SD that varies across cases (σ_{e_j}) and the autocorrelation that varies across cases (ρ_j) was constructed as follows:

$$\begin{aligned} \sigma_{e_j} &\sim \text{Uniform}(L_\sigma, U_\sigma) && \text{with } L_\sigma \sim \text{Uniform}(0, 100) \\ & && U_\sigma \sim \text{Uniform}(L_\sigma, 100) \\ \rho_j &\sim \text{Normal}(\mu_\rho, \sigma_\rho^2) I(-1 < \rho_j < 1) && \text{with } \mu_\rho \sim \text{Normal}(0, 1000^2) \\ & && \sigma_\rho \sim \text{Uniform}(0, 100) \end{aligned} \quad (3)$$

The prior for σ_{e_j} was assumed to follow the uniform distribution with the lower limit of L_σ and the upper limit of U_σ . The L_σ and U_σ were further assumed to follow a uniform distribution. In addition, the prior for ρ_j was assumed to follow the normal distribution with a mean of μ_ρ and a variance of σ_ρ^2 but with the restricted range between -1 and 1. The priors for μ_ρ and σ_ρ were further defined as a normal distribution and a uniform distribution, respectively.

Results

A data set per each condition of the design factors (24 conditions) was first generated to test convergence and to make decisions about the number of iterations, and the burn-in period. Based on the test results, it was decided to use a burn-in of 2,000 iterations and to run an additional 500,000 iterations, but to use only 50,000 samples of the 500,000

iterations after thinning to form the posterior distribution for the main analyses. Various diagnostic criteria were used in monitoring convergence, including trace plots, history plots, Kernel density plots, and Brooks–Gelman–Rubin (BGR) plots for the simulated data sets using two different MCMC chains. No signs for non-convergence were found.

Fixed Effects

The relative bias values for treatment effects (shift in level and shift in slope) were compared across the two models. For the treatment effects, the average relative bias values were very minimal and RMSE values were comparable for both models. The average relative bias values were smaller than 5% for both treatment parameters (Model 1: shift in level = .001, shift in slope = .017; Model 2: shift in level < .001, shift in slope = .015) which are considered as minimal, according to [Hoogland and Boomsma \(1998\)](#)'s criteria. The average RMSE value of the shift in level was 0.68 for both models and the average RMSE value of the shift in slope was 0.23 for Model 1 and 0.22 for Model 2 (proposed model). In addition, the average credible interval (CI) coverage and width values for the treatment effects were similar across the two models. The CI coverage exceeded .95 for both models and the average CI width values of the shift in level was 4.96 for Model 1 and 4.95 for Model 2, and the average CI width value of the shift in slope was 1.68 for both models.

Variance Components

The average relative bias values of the variance components across the models are provided in [Table 1](#), along with values for RMSE, CI coverage, and CI width. Consistent with previous findings of SCED with MLM using both REML and Bayesian estimations, the estimates of the variance components were biased. Although the biases were often comparable across models, the average relative bias for the level-1 error *SD* was noticeably smaller in Model 2 (.06) than Model 1 (.10). To further examine bias, RMSE, CI coverage, and CI width, ANOVAs were run to examine the proportion of the variance in these outcomes associated with each of the factors in the simulation study (i.e., η^2 values) and identify factors with a medium effect of .06 or larger. The interaction between the true level-1 error structure and the type of model accounted for substantial variation in the average bias ($\eta^2 = .10$) and the average RMSE ($\eta^2 = .16$) of the level-1 error *SD* ([Figure 2](#)). Specifically, for Model 1, mean bias and RMSE increased consistently and substantially as the true level-1 error structure shifts from homogeneous to moderately and severely heterogeneous. However, for Model 2, changes in the true level-1 error structure led to smaller changes in mean bias and RMSE.

Table 1

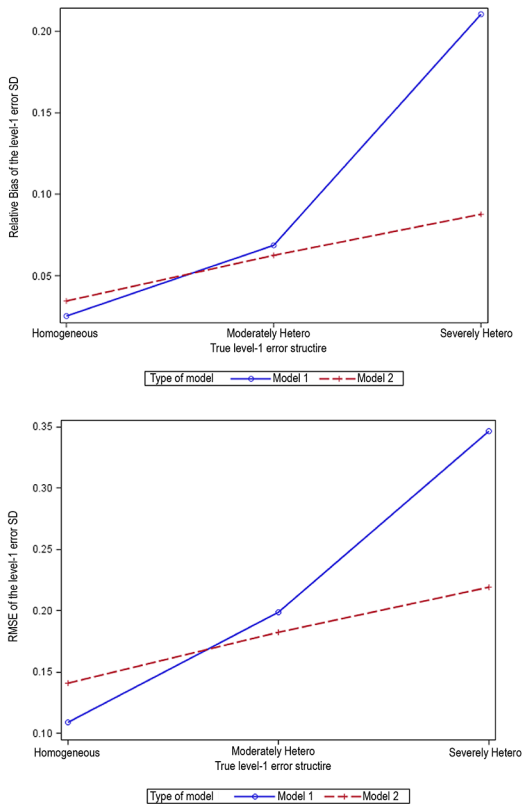
Average Relative Bias, RMSE, CI Coverage and Width for the Variance Components

Outcome	Level-2 error SD				Level-1 error SD		Autocorrelation	
	Shift in level		Shift in slope					
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Bias	0.89	0.88	0.99	0.97	0.10	0.06	-0.60	-0.48
RMSE	1.23	1.21	0.42	0.41	0.22	0.18	0.26	0.25
CI coverage	0.97	0.97	0.97	0.98	0.85	0.97	0.81	0.94
CI width	6.43	6.44	1.62	2.16	0.47	0.81	0.74	1.10

Note. Model 2 is the proposed model.

Figure 2

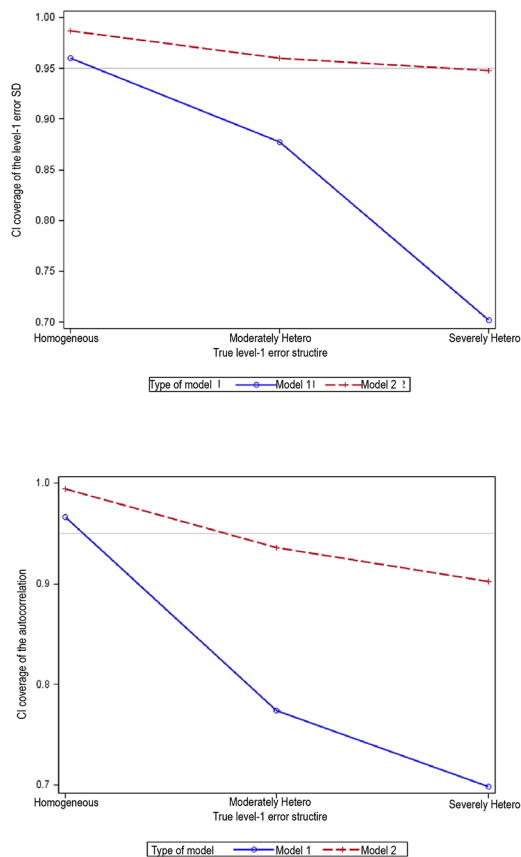
Line Graph Depicting Average Relative Bias and Rmse for the Level-1 Error Standard Deviation as a Function of the Two-Way Interaction Effect Between the Type of Model and the True Level-1 Error Structure



The average CI coverage values of the level-2 error SD for treatment effects were over the nominal value (.95) across the two models. The interaction between type of model and the true level-1 error structure had found that the different modeling of the level-1 error structure had substantial impacts on the average CI coverage of the level-1 error SD and the autocorrelation ($\eta^2 = .19, .07$, respectively). As illustrated in Figure 3, when the true level-1 error structure was homogeneous, the CI coverage was above the nominal level for both models. However, when the true level-1 error structure was one of the heterogeneous error structures, CI coverage was close to the nominal level (.95) for Model 2, while CI coverage decreased substantially below the nominal level for Model 1. However, Model 2 has larger CI widths than Model 1.

Figure 3

Line Graph Depicting Average ci Coverage of the Level-1 Error SD and Autocorrelation as a Function of the Two-Way Interaction Effect Between the Type of Model and the True Level-1 Error Structure



A Follow-Up Study

In the main study, data having the heterogeneous level-1 error structure had been generated in a way that every case had a unique value of the level-1 error *SD* and autocorrelation within a specified range. However, it is conceivable that the values are not evenly spread out in a real dataset, instead, one or two of the cases may have substantially larger variance compared with the other cases (Baek et al., 2016). Therefore, it seemed worthwhile to expand the simulation to include such a condition. Thus, a follow-up study was conducted with a condition where data were generated in a way that one case had a substantial difference in the level-1 error variance (16 times bigger than the other cases) and the autocorrelation (either half [.2] or twice [.4] as large as the other cases). In the follow-up study, there were 8 conditions simulated using the three factors that were used in the main study as shown in Table 2.

Table 2
A Follow-Up Study Conditions

Factor	Level
Autocorrelation	.2 and .4 or .4 and .2
Combination of number of cases and series length per case	4 and 10 or 8 and 20
Method to model the level-1 error structure	Not modeling between case variation (Model 1) or Modeling between case variation (Model 2)

The results of all of the simulated conditions for the fixed treatment effects and variance components are provided in Table 3. Unlike the main study, this study found that the different modeling methods for the level-1 error structure had an impact on the estimates of the both fixed treatment effects and variance components. Although bias estimates for the fixed effects were minimal regardless of the model used, the RMSE values for the fixed treatment effects were generally smaller when estimated by the proposed model. In addition, unlike the main study, this study found that the different modeling methods for the level-1 error structure had an impact on the estimates of the level-2 error *SD*. For the level-2 error *SD*, the proposed model had a generally smaller average bias and smaller RMSE values than Model 1. Similar to the main study, the proposed model generally provided better estimates for the level-1 error *SD* and autocorrelation.

Table 3
Average Relative Bias, RMSE, Ci Coverage and Width for the Fixed Effects and Variance Components for the Follow-Up Study

Parameter	Bias		RMSE		CI coverage		CI width	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Fixed treatment effects								
Shift in level	< -0.01	< -0.01	1.01	0.73	> 0.99	0.99	7.11	5.80
Shift in slope	0.02	0.01	0.39	0.26	0.98	0.97	3.10	2.47
Variance components								
Level-2 error SD								
Shift in level	2.31	1.80	2.12	1.55	0.93	0.98	9.08	7.70
Shift in slope	3.74	2.67	1.09	0.74	0.91	0.98	1.68	3.42
Level-1 error SD	0.60	0.10	1.24	0.51	0.03	0.93	0.85	1.53
Autocorrelation	-1.06	-0.87	0.41	0.33	0.50	0.84	0.67	1.10

Note. Model 2 is the proposed model.

An Empirical Example of the Proposed Model

A published single-case study (Resetar, Noell, & Pellegrin, 2006) was selected to obtain data for an illustrative analysis. Resetar and colleagues (2006) conducted a multiple-base-line design across five first-grade students who were reading below grade level. The purpose of their study was to investigate the effect of parents reading tutoring on child reading fluency, which was measured as words read correctly per minute (WCPM).

The raw data were extracted from the graphs presented in the article using DataThief III (Tummers, 2006). Model 1 and Model 2 (proposed model) were then specified using Equations (1) and (2). OpenBUGS software and the same priors from the simulation study were used. Ten thousand iterations were discarded for a burn-in, and 700,000 iterations for Model 1 and 800,000 iterations for the proposed model were run to form the posterior distribution. The previously mentioned diagnostic criteria found no signs for non-convergence.

The results for Model 1 and Model 2 are compared in Table 4 below. As indicated in the simulation studies that both the fixed effects and the variance components could differ across the two models, results from the empirical data were different across the two models. First, the treatment effect estimates were somewhat different across the two models. The treatment effect for phase or the shift in level was 25.38 for Model 1, and 22.76 for the proposed model (θ_{10}). The treatment effect for the change in trend was -0.31 for Model 1 and -0.69 for the proposed model (θ_{20}). The variance components were also different across the two models. The level-2 error SDs were higher for the proposed

model than Model 1. The *SD* for the shift in level was 11.73 for Model 1 and 14.85 for the proposed model (σ_{u_1}). For the difference in trends (shift in slope), the *SD* was 1.00 for Model 1 and 1.07 for the proposed model (σ_{u_3}). The average level-1 error *SD* (σ_e) was slightly higher in the proposed model than Model 1, 13.51 for Model 1 and 14.00 for the proposed model. In the proposed model, the level-1 error *SD* varied across cases, ranging from 10.66 to 17.23. The estimated average autocorrelation was 0.34 for Model 1 and 0.29 for the proposed model (ρ). In the proposed model, the autocorrelation varied across cases, ranging from 0.13 to 0.36.

Table 4
Parameter Estimates From the MLMs With Level-1 Error Covariance Structures That Were Fixed Across Participants (Model 1) or Varied Across Participants (Model 2)

Parameter	Model 1	Model 2
Variance components		
Intercept	14.24	12.88
Phase	11.73	14.85
Time	0.85	1.08
Interaction	1.00	1.07
Variance	13.51	14.00
AR(1)	0.34	0.29
Variance (Person 1)	-	10.66
AR(1)	-	0.13
Variance (Person 2)	-	11.56
AR(1)	-	0.36
Variance (Person 3)	-	13.07
AR(1)	-	0.30
Variance (Person 4)	-	11.35
AR(1)	-	0.32
Variance (Person 5)	-	17.23
AR(1)	-	0.34
Fixed effects		
Intercept	27.59	24.78
Phase	25.38	22.76
Time	0.31	0.79
Interaction	-0.31	-0.69

Discussion and Recommendations

This study developed a method for estimating between case heterogeneity in level-1 variances and provides insight into how different modeling approaches for the level-1 error

covariance matrices impacts statistical inferences. The results of this study indicate that the different modeling methods in the level-1 error structure can have an impact on the variance components, in some cases, both fixed effects and variance components. This finding is particularly distinguished from previous works that have investigated other misspecifications of the level-1 error structure. The previous studies have found that the fixed effects are generally robust to misspecifications of the level-1 error structure, however, this study found that the misspecification of the level-1 error structure can have an impact on fixed effects when analyzing data that show one or more cases that have substantially different variability than the others. In addition, this study suggests that accuracy and precision of the variance components can be improved by modeling between case variation in the level-1 error structure, and the effectiveness of modeling between case variation increased as the degree of the heterogeneity in the data increased.

The results of this study lead to various implications for applied single-case researchers, as well as for methodologists. Single-case researchers can feel comfortable interpreting the overall treatment effects when they have data that show no or a moderate degree of between case variation in the within-case variance, regardless of whether between-case heterogeneity has been explicitly modeled. However, when the heterogeneity becomes more severe, and particularly when one case has variance that is very different from the others (as in our follow-up study), researchers are encouraged to model the variation in the level-1 error variances. Doing so can lead to less error in the estimate of the average treatment effect, and less bias in the variance estimates. To facilitate modeling between case variation in the level-1 error variances, this study has developed a Bayesian model and corresponding OpenBUGS code, which is accessible to applied researchers for use in their own research (see Supplementary Materials).

This study also provides a few implications for methodologists who study the use of multilevel modeling to conduct single-case data analyses. Although this study was helpful in providing some initial guidance about the use of multilevel modeling for single-case data when there are differences in the within-case variation, this Monte Carlo study was limited by the conditions that were examined. The conditions were chosen based on a review of single-case literature and applied studies that used two-level models to analyze single-case data, but only some of the possible options were included in this study. More simulation work can be done to document the impact of various modeling options across a broader range of heterogeneous data conditions, including those that involve more complex dependent error structures than AR(1). In addition, it is not clear the degree to which biases may be reduced through the choice of alternative priors or may be less pronounced in contexts where simpler models without trends are appropriate. By contrasting conditions and models with and without trends, along with estimation incorporating a wider range of prior distributions, and in particular more informative priors for the random components, future research could make it possible to further refine modeling advice.

Funding: The authors have no funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Acknowledgments: The authors have no support to report.

Author note: This study is initiated based on the first author's dissertation (Baek, 2015).

Supplementary Materials

OpenBUGS code to model between case variation in the level-1 error structure in multilevel SCED data analysis is available via the PsychArchives repository (for access see [Index of Supplementary Materials](#) below).

Index of Supplementary Materials

Baek, E., & Ferron, J. J. M. (2020). *Supplementary materials [code] to: Modeling heterogeneity of the level-1 error covariance matrix in multilevel models for single-case data*. PsychOpen. <https://doi.org/10.23668/psycharchives.2893>

References

- Baek, E. (2015). *Consequences of non-modeled and modeled between case variation in the level-1 error structure in multilevel models for single-case data: A Monte Carlo study* [Unpublished doctoral dissertation]. University of South Florida, Tampa, FL, USA. Retrieved from <https://scholarcommons.usf.edu/etd/5860/>
- Baek, E., Beretvas, S. N., Van den Noortgate, W., & Ferron, J. M. (2019). Brief research report: Bayesian versus REML estimations with noninformative priors in multilevel single-case data. *Journal of Experimental Education*. <https://doi.org/10.1080/00220973.2018.1527280>
- Baek, E., & Ferron, J. M. (2013). Multilevel models for multiple-baseline data: Modeling across participant variation in autocorrelation and residual variance. *Behavior Research Methods*, 45(1), 65-74. <https://doi.org/10.3758/s13428-012-0231-z>
- Baek, E., Moeyaert, M., Petit-Bois, M., Beretvas, S. N., Van den Noortgate, W., & Ferron, J. M. (2014). The use of multilevel analysis for integrating single-case experimental design results within a study and across studies. *Neuropsychological Rehabilitation*, 24(3-4), 590-606. <https://doi.org/10.1080/09602011.2013.835740>
- Baek, E., Petit-Bois, M., & Ferron, M. J. (2013). *A feasible way to vary level-1 error structure across participants in multilevel models for single-case data*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, USA.

- Baek, E., Petit-Bois, M., Van den Noortgate, W., Beretvas, T., & Ferron, J. (2016). Using visual analysis to evaluate and refine multilevel models of single-case studies. *The Journal of Special Education*, 50, 18-26. <https://doi.org/10.1177/0022466914565367>
- Berger, J. O. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, 1, 385-402. <https://doi.org/10.1214/06-BA115>
- Berger, J. O., & Strawderman, W. E. (1996). Choice of hierarchical priors: Admissibility in estimation of normal means. *Annals of Statistics*, 24, 931-951. <https://doi.org/10.1214/aos/1032526950>
- Browne, W. (2008). *MCMC estimation in MLwin*. Bristol, United Kingdom: Centre for Multilevel Modeling.
- Daniels, M. J., & Kass, R. E. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, 94, 1254-1263. <https://doi.org/10.1080/01621459.1999.10473878>
- Dufrene, B. A., Reisener, C. D., Olmi, D. J., Zoder-Martell, K., McNutt, M. R., & Horn, D. R. (2010). Peer tutoring for reading fluency as a feasible and effective alternative in response to intervention systems. *Journal of Behavioral Education*, 19, 239-256. <https://doi.org/10.1007/s10864-010-9111-8>
- Ferron, J. (2002). Reconsidering the use of the general linear model with single-case data. *Behavior Research Methods, Instruments, & Computers*, 34, 324-331. <https://doi.org/10.3758/BF03195459>
- Ferron, J. M., Bell, B. A., Hess, M. R., Rendina-Gobioff, G., & Hibbard, S. T. (2009). Making treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods*, 41, 372-384. <https://doi.org/10.3758/BRM.41.2.372>
- Ferron, J. M., Farmer, J. L., & Owens, C. M. (2010). Estimating individual treatment effects from multiple-baseline data: A Monte Carlo study of multilevel modeling approaches. *Behavior Research Methods*, 42, 930-943. <https://doi.org/10.3758/BRM.42.4.930>
- Gamerman, D., & Lopes, H. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference* (2nd ed.). Boca Raton, FL, USA: Chapman and Hall/CRC.
- Gelman, A. (2002). Prior distribution. *Encyclopedia of Environmetrics*, 3, 1634-1637.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515-534. <https://doi.org/10.1214/06-BA117A>
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL, USA: Chapman and Hall/CRC.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. Boca Raton, FL, USA: Chapman and Hall/CRC.
- Goldstein, M. (2006). Subjective Bayesian analysis: Principles and practice. *Bayesian Analysis*, 1, 403-420. <https://doi.org/10.1214/06-BA116>
- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, 26, 329-367. <https://doi.org/10.1177/0049124198026003003>

- Huitema, B. E., & McKean, J. W. (1994). Reduced bias autocorrelation estimation: Three jackknife methods. *Educational and Psychological Measurement*, 54, 654-665.
<https://doi.org/10.1177/0013164494054003008>
- Huitema, B. E., & McKean, J. W. (2000). Design specification issues in time-series intervention models. *Educational and Psychological Measurement*, 60, 38-58.
<https://doi.org/10.1177/00131640021970358>
- Ingersoll, B., & Lalonde, K. (2010). The impact of object and gesture limitation training on language used in children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research: JSLHR*, 53, 1040-1051. [https://doi.org/10.1044/1092-4388\(2009/09-0043\)](https://doi.org/10.1044/1092-4388(2009/09-0043))
- Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). London, United Kingdom: Oxford University Press.
- Koegel, L. K., Singh, A. K., & Koegel, R. L. (2010). Improving motivation for academics in children with autism. *Journal of Autism and Developmental Disorders*, 40, 1057-1066.
- Maggin, D. M., Swaminathan, H., Rogers, H. J., O'Keefe, B. V., Sugai, G., & Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology*, 49, 301-321.
<https://doi.org/10.1016/j.jsp.2011.03.004>
- Matyas, T. A., & Greenwood, K. M. (1996). Serial dependency in single-case time series. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 215-243). Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- McKnight, S. D., McKean, J. W., & Huitema, B. E. (2000). A double bootstrap method to analyze linear models with autoregressive error terms. *Psychological Methods*, 5, 87-101.
<https://doi.org/10.1037/1082-989X.5.1.87>
- Moeyaert, M., Rindskopf, D., Onghena, P., & Van den Noortgate, W. (2017). Multilevel modeling of single-case data: A comparison of maximum likelihood and Bayesian estimation. *Psychological Methods*, 22, 760-778. <https://doi.org/10.1037/met0000136>
- Moeyaert, M., Ugille, M., Ferron, J., Beretvas, S., & Van den Noortgate, W. (2013). The three-level synthesis of standardized single-subject experimental data: A Monte Carlo simulation study. *Multivariate Behavioral Research*, 48, 719-748. <https://doi.org/10.1080/00273171.2013.816621>
- Morris, C. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78, 47-55. <https://doi.org/10.1080/01621459.1983.10477920>
- Neter, J., Wasserman, W., & Kutner, M. H. (1990). *Applied linear statistical models: Regression, analysis of variance, and experimental designs* (3rd ed.). Homewood, IL, USA: Irwin.
- Oddo, M., Barnett, D. W., Hawkins, R. O., & Musti-Rao, S. (2010). Reciprocal peer tutoring and repeated reading: Increasing practicality using student groups. *Psychology in the Schools*, 47(8), 842-858. <https://doi.org/10.1002/pits.20508>
- Owens, C. M., & Ferron, J. M. (2012). Synthesizing single-case studies: A Monte Carlo examination of a three-level meta-analytic model. *Behavior Research Methods*, 44, 795-805.
<https://doi.org/10.3758/s13428-011-0180-y>

- Petit-Bois, M., Baek, E., Van den Noortgate, W., Beretvas, S. N., & Ferron, J. M. (2016). The consequences of modeling autocorrelation when synthesizing single-case studies using a three level model. *Behavior Research Methods*, 48, 803-812. <https://doi.org/10.3758/s13428-015-0612-1>
- Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics*, 39, 368-393. <https://doi.org/10.3102/1076998614547577>
- Resetar, J., Noell, G., & Pellegrin, A. (2006). Teaching parents to use research-supported systematic strategies to tutor their children in reading. *School Psychology Quarterly*, 21, 241-261. <https://doi.org/10.1521/scpq.2006.21.3.241>
- Rindskopf, D. (2014a). Nonlinear Bayesian analysis for single case designs. *Journal of School Psychology*, 52, 179-189. <https://doi.org/10.1016/j.jsp.2013.12.003>
- Rindskopf, D. (2014b). Bayesian analysis of data from single case designs. *Neuropsychological Rehabilitation*, 24, 572-589. <https://doi.org/10.1080/09602011.2013.866903>
- Shadish, W. R., Rindskopf, D., Hedges, L., & Sullivan, K. (2013). Bayesian estimates of autocorrelations in single-case designs. *Behavior Research Methods*, 45, 813-821. <https://doi.org/10.3758/s13428-012-0282-1>
- Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, 43, 971-980. <https://doi.org/10.3758/s13428-011-0111-y>
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2014). *OpenBUGS User Manual*. Retrieved from <http://www.openbugs.net/Manuals/Manual.html>
- Swaminathan, H., Rogers, H., & Horner, R. (2014). An effect size measure and Bayesian analysis of single-case designs. *Journal of School Psychology*, 52, 213-230. <https://doi.org/10.1016/j.jsp.2013.12.002>
- Toothaker, L. E., Banz, M., Noble, C., Camp, J., & Davis, D. (1983). N = 1 designs: The failure of ANOVA-based tests. *Journal of Educational Statistics*, 8(4), 289-309. <https://doi.org/10.3102/10769986008004289>
- Tummers, B. (2006). DataThief III [software]. Retrieved from <http://datathief.org/>
- Van den Noortgate, W., & Onghena, P. (2003). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly*, 18, 325-346. <https://doi.org/10.1521/scpq.18.3.325.22577>
- Wolfinger, R. (1993). Covariance structure selection in general mixed models. *Communications in Statistics. Simulation and Computation*, 22, 1079-1106. <https://doi.org/10.1080/03610919308813143>



Methodology is the official journal of the European Association of Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology Information (ZPID), Germany.