The British
Psychological Society

The relationship between multilevel models and non-parametric multilevel mixture models: Discrete approximation of intraclass correlation, random coefficient distributions, and residual heteroscedasticity

Jason D. Rights* and Sonya K. Sterba

Vanderbilt University, Nashville, Tennessee, USA

Multilevel data structures are common in the social sciences. Often, such nested data are analysed with multilevel models (MLMs) in which heterogeneity between clusters is modelled by continuously distributed random intercepts and/or slopes. Alternatively, the non-parametric multilevel regression mixture model (NPMM) can accommodate the same nested data structures through discrete latent class variation. The purpose of this article is to delineate analytic relationships between NPMM and MLM parameters that are useful for understanding the indirect interpretation of the NPMM as a non-parametric approximation of the MLM, with relaxed distributional assumptions. We define how seven standard and non-standard MLM specifications can be indirectly approximated by particular NPMM specifications. We provide formulas showing how the NPMM can serve as an approximation of the MLM in terms of intraclass correlation, random coefficient means and (co)variances, heteroscedasticity of residuals at level 1, and heteroscedasticity of residuals at level 2. Further, we discuss how these relationships can be useful in practice. The specific relationships are illustrated with simulated graphical demonstrations, and direct and indirect interpretations of NPMM classes are contrasted. We provide an R function to aid in implementing and visualizing an indirect interpretation of NPMM classes. An empirical example is presented and future directions are discussed.

1. Introduction

Multilevel data structures are common in social science research, as when students are nested within schools or patients are nested within clinicians. Such nested data are typically analysed with multilevel models (MLMs, also known as hierarchical linear models or mixed effects models) in which heterogeneity between clusters is modelled by continuous, normally distributed random intercepts and/or slopes (e.g., Goldstein, 2011; Hox, 2010; Raudenbush & Bryk, 2002).

Multilevel mixture models are also used to accommodate nested data. One type of multilevel mixture that has been increasingly applied in practice has been termed the *non-parametric multilevel mixture model* (NPMM; e.g., Asparouhov & Muthén, 2008; Finch & Marchant, 2013; Henry & Muthén, 2010; Kaplan & Keller, 2011; Karakos, 2015;

---

*Correspondence should be addressed to Jason D. Rights, Department of Psychology and Human Development, Vanderbilt University, Peabody #552, 230 Appleton Place, Nashville, TN 37203, USA (email: jason.d. rights@vanderbilt.edu).

Park & Yu, 2015; Van Horn, Feng, Kim, Lamont, Feaster, & Jaki, 2016; Vermunt, 2003, 2004, 2008; Yu & Park, 2014). In this model, latent classes are extracted at multiple levels of a hierarchical data structure (e.g., level 1 *and* level 2) and, unlike the conventional MLM, there are no continuously distributed random effects. In this paper, we consider the *regression*[1] version of the NPMM, where intercepts and slopes of covariates can vary discretely across level-1 and level-2 classes (e.g., Vermunt, 2010; Vermunt & Magidson, 2005). Compared to the MLM, the NPMM provides the advantages of relaxed distributional assumptions (e.g., distributions of effects across clusters are not assumed normal) and the ability to interpret classes both *directly* (as representing literal, discrete subpopulations) and *indirectly* (e.g., as approximating an underlying continuous distribution of effects across clusters; Titterington, Smith, & Makov, 1985). Indeed, the NPMM is termed 'non-parametric' because it has often been motivated for use with an indirect interpretation in which classes can represent a discrete approximation of continuous distributions, such as those assumed in the MLM.

However, precisely *how* this NPMM non-parametrically approximates MLM parameters[2] has not been shown mathematically. Having only a conceptual or incomplete[3] understanding of the relationship between the NPMM and MLM leaves us with several unknowns that limit our understanding of the indirect interpretation of the NPMM: (1) What are the analytic relationships through which the parameters of the NPMM can approximate the parameters of the MLM? (2) Which standard MLM specifications can be approximated by which NPMM specifications? (3) Can the NPMM also indirectly approximate non-standard MLM specifications? In contrast, model parameters from other mixtures have been analytically related to the non-mixture-model parameters they approximate (e.g., Bauer, 2005, 2007; Vermunt & Van Dijk, 2001), which has aided the indirect interpretation of classes (e.g., Muthén & Asparouhov, 2008; Pek, Sterba, Kok, & Bauer, 2009). A recent review of psychology mixture applications noted that 'whereas a direct interpretation of classes is intuitive, an indirect semiparametric function of classes is less so' and called for methodologists to 'provide more concrete details on an indirect interpretation' to facilitate its use in practice (Sterba, Baldasaro, & Bauer, 2012, pp. 593, 625).

Understanding the relationship between the NPMM and MLM (i.e., the indirect interpretation of the NPMM) would be useful to applied researchers for several reasons. First, understanding this relationship can aid both multilevel modelling and mixture modelling researchers in developing competing explanations of their data-generating mechanisms. This can be particularly useful in mixture applications given that many applied researchers currently only consider a direct interpretation of classes – which may be 'seductive' in its simplicity, yet risks reifying classes as literal subpopulations (Bauer,

---

[1] Note that Muthén and Asparouhov (2009) address a different kind of multilevel regression mixture that involves both random effects *and* classes, requires numerical integration, and is most often used for directly interpreting classes. They do not discuss the NPMM. The NPMM involves only classes, does not require numerical integration, and is often used for indirectly interpreting classes.

[2] When the normality assumptions of the MLM hold, the NPMM non-parametrically approximates MLM parameters, including the first moment(s) and second central moment(s) of the random effect(s) distribution. When these normality assumptions do not hold, the NPMM has the flexibility to not only more accurately recover lower-order moments of the random effects distribution (e.g., Asparouhov & Muthén, 2008; Brame, Nagin, & Wasserman, 2006), but also recover its higher-order moments. As a shorthand, throughout this paper we refer to the NPMM as non-parametrically approximating 'MLM parameters' when referring to its discrete approximation of such lower-order moments, regardless of whether MLM assumptions are upheld or not.

[3] Certain relationships have been delineated for simpler mixtures with classes exclusively at level 2 (Vermunt & Van Dijk, 2001), as described in detail later.

2007; Sampson & Laub, 2005). Second, understanding this relationship would illuminate how the NPMM can be used as a preliminary, exploratory tool for multilevel modelling researchers. It will be shown that fitting our general NPMM automatically and flexibly approximates a number of non-standard MLM specifications that would not typically be considered by applied researchers fitting MLMs, yet could diagnostically suggest novel future directions for multilevel model-building. Third, understanding this link can help synthesize or even meta-analyse results across already published studies that have used the MLM and/or NPMM with similar dependent variables and covariates. Without the linkages established in this paper, there would be no way to formally compare empirical results across methods, as is often desired (e.g., Bergman, 2015; Sterba & Bauer, 2014).

The purpose of this paper is to fill this gap by enumerating the relationships between the NPMM and MLM useful for indirectly interpreting the NPMM and, moreover, to highlight the practical importance of understanding these relationships. In doing so, we hope to facilitate the use of the NPMM as a non-parametric approximation of the MLM in practice. We first consider a series of NPMMs of increasing complexity. For each, we (a) define the corresponding MLM and (b) provide formulas underlying how parameters of each MLM are indirectly approximated. We begin by relating null NPMM parameters to the intraclass correlation (ICC) from a random-intercept-only MLM; subsequently, we discuss these relationships for MLM fixed effects, MLM random slope variances, and also non-standard features not previously known to be approximated by the NPMM – namely, MLM heteroscedastic residual variances at level 1 and level 2. To ground these analytic developments, we contrast direct and indirect interpretations of latent classes, employ simulated graphical illustrations, and provide an empirical example involving predicting mathematics achievement for students nested within schools. In our discussion (Section 8), we consider additional novel uses for the indirect interpretation of the NPMM. M*plus* (Muthén & Muthén, 1998–2015) syntax to implement the NPMM and an R function to implement indirect approximation calculations are provided in our Appendix S1.

Regarding our scope, note that subsequent research can use our analytic formulas to investigate via simulation the degree of correspondence between MLM parameters and their NPMM approximations under different design conditions. Though we provide a single-sample empirical illustration, a full-scale simulation, as has been done for other mixtures (e.g., Brame *et al.*, 2006; Muthén & Asparouhov, 2008; Sterba & Bauer, 2014),[4] is beyond our scope here.

Before continuing, it is important to note the ways in which our analytic developments build on and differ from those provided for other non-parametric mixtures with classes *only* at level 2 (Vermunt & Van Dijk, 2001). First, the additional presence of level-1 classes in the NPMM not only changes computational aspects of these analytic relationships (see Sections 2 and 3), but also introduces new features (see Sections 4 and 5). Second, previous work has not related the variety of possible NPMM specifications each to particular MLMs, nor focused on approximating the ICC. Third, we show steps to derive each indirect approximation formula. This will allow our analytic approach to be generalized to non-parametric versions of other multilevel mixtures (e.g., multilevel latent class analysis) and to be used to clarify indirect relationships in other contexts.

---

[4] In general, the NPMM would be expected to achieve a closer approximation of first moments and second central moments of MLM's random effects distribution with greater numbers of classes. Also note that the NPMM can fit aspects of the data that MLM cannot (i.e., higher-order moments). See, for example, Bauer and Curran (2003).

2. Relationship of the NPMM to the random-intercept-only MLM and ICC

To begin, we consider a null NPMM. In this model, there are no covariates; rather, the distribution of the outcome variable, y_{ij}, is allowed to differ across latent classes at level 1 and level 2. Let i denote a level-1 unit, or observation within cluster ($i = 1, \ldots, N_j$), and j denote a level-2 unit or cluster ($j = 1, \ldots, J$). Latent class membership at level 1 and level 2 is denoted c_{ij} and d_j, respectively, with particular level-1 classes denoted by k ($k = 1, \ldots, K$) and level-2 classes by h ($h = 1, \ldots, H$).[5] For instance, suppose we were analysing scores on a mathematics achievement test for students (level-1 units) nested within schools (level-2 units). For the distribution of maths scores, we may specify a model with multiple *school-level latent classes* (e.g., level-2 classes of schools with primarily high-achieving students vs. level-2 classes of schools with primarily low-achieving students) and, within each school-level class, multiple *student-level latent classes* (e.g., level-1 classes of high-achieving students vs. low-achieving students). More formally, this null NPMM is specified in equation (1) in Table 1[6] as $y_{ij}|_{c_{ij} = k, d_j = h} = \gamma_0^{kh} + \varepsilon_{ij}$. In equation (1), the outcome variable y_{ij}, where $c_{ij} = k$ and $d_j = h$, is modelled by a kh class-combination intercept, γ_0^{kh}, and residual, ε_{ij}. The latter is assumed normally distributed with class-combination-specific[7] variance θ^{kh}, that is, $\varepsilon_{ij} \sim N(0, \theta^{kh})$. Thus, the γ_0^{kh} and θ^{kh} parameters are specific to a particular combination of level-1 class and level-2 class.

There are three types of probabilities important to the NPMM. The marginal probability of level-2 class membership, $p(h_j = d) = \pi^{\bullet h}$, represents the probability of a randomly selected observation belonging to level-2 class h. It is modelled by a multinomial regression, $p(d_j = h) = \pi^{\bullet h} = \exp(\varpi^h)/\sum_{h=1}^{H} \exp(\varpi^h)$, with multinomial intercept ϖ^h fixed at 0 in class H for identification. Additionally, the conditional probability of level-1 class membership given level-2 class membership, $p(c_{ij} = k|h_j = d) = \pi^{k|h}$, represents the probability of belonging to level-1 class k for a randomly selected observation within level-2 class h. It is modelled by another multinomial regression, $p(c_{ij} = k|d_j = h) = \pi^{k|h} = \exp(\omega^k + \delta^{kh})/\sum_{k=1}^{K} \exp(\omega^k + \delta^{kh})$, with multinomial intercept ω^k and slope (k on h) δ^{kh}. For identification, $\omega^K = \delta^{Kh} = \delta^{kH} = 0$. Finally, the kh class-combination probability π^{kh} represents the joint probability of a randomly selected observation belonging to level-1 class k and level-2 class h. This is computed as the product of $\pi^{\bullet h}$ and $\pi^{k|h}$.

Under a direct interpretation of equation (1), one could consider intercept and residual variance differences between classes to represent substantively meaningful differences between distinct subpopulations. For example, average maths achievement scores may differ across distinct school-level (i.e., level-2) classes, and we may be particularly interested in examining membership in kh class combinations such as a high-achieving student-level (i.e., level-1) class within a low-performing school-level class, or a low-achieving student-level class within a high-performing school-level class.

---

[5] Though selecting K and H is not our focus, we revisit this topic later in the context of an empirical example.
[6] Equations (1)–(14) are contained in Table 1, and equations (15) and (16) in Table 2.
[7] In empirical research, it is common for some or all class-combination residual variances to be constrained equal across class combination, that is $\theta^{kh} = \theta$, for estimation stability (for discussion, see McLachlan & Peel, 2000).

Table 1. Multilevel model (MLM) specifications with corresponding non-parametric multilevel regression mixture model (NPMM) specifications

MLM		NPMM[a]	
Random-intercept-only model			
$y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$ $u_{0j} \sim N(0,\ \tau_{00})\ e_{ij} \sim N(0,\ \sigma^2)$	(2)	$y_{ij}\big|_{c_{ij}=k,\,d_j=b} = \gamma_0^{kb} + \varepsilon_{ij}$ $\varepsilon_{ij} \sim N(0,\ \theta^{kb})$	(1)
+ Fixed slope of x_{ij}			
$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + e_{ij}$ $u_{0j} \sim N(0,\ \tau_{00})\ e_{ij} \sim N(0,\ \sigma^2)$	(4)	$y_{ij}\big|_{c_{ij}=k,\,d_j=b} = \gamma_0^{kb} + \ddot{\gamma}_1^{kb} x_{ij} + \varepsilon_{ij}$ $\varepsilon_{ij} \sim N(0,\ \theta^{kb})$	(3)
+ Random slope of x_{ij}			
$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + u_{1j}x_{ij} + e_{ij}$ $\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\tau_{00} & \\ \tau_{01} & \tau_{11}\end{bmatrix}\right) e_{ij} \sim N(0,\ \sigma^2)$	(6)	$y_{ij}\big|_{c_{ij}=k,\,d_j=b} = \gamma_0^{kb} + \gamma_1^{\bullet b} x_{ij} + \varepsilon_{ij}$ $\varepsilon_{ij} \sim N(0,\ \theta^{kb})$	(5)
+ Level-1 heteroscedasticity on x_{ij}			
$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + u_{1j}x_{ij} + e_{0ij} + e_{1ij}x_{ij}$ $\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\tau_{00} & \\ \tau_{01} & \tau_{11}\end{bmatrix}\right) \begin{bmatrix} e_{0ij} \\ e_{1ij} \end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_0^2 & \\ \sigma_{0,1} & \sigma_1^2\end{bmatrix}\right)$	(8)	$y_{ij}\big|_{c_{ij}=k,\,d_j=b} = \gamma_0^{kb} + \gamma_1^{kb} x_{ij} + \varepsilon_{ij}$ $\varepsilon_{ij} \sim N(0,\ \theta^{kb})$	(7)
+ Fixed slope of w_j			
$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}w_j + u_{0j} + u_{1j}x_{ij} + e_{0ij} + e_{1ij}x_{ij}$ $\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\tau_{00} & \\ \tau_{01} & \tau_{11}\end{bmatrix}\right) \begin{bmatrix} e_{0ij} \\ e_{1ij} \end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_0^2 & \\ \sigma_{0,1} & \sigma_1^2\end{bmatrix}\right)$	(10)	$y_{ij}\big|_{c_{ij}=k,\,d_j=b} = \gamma_0^{kb} + \gamma_1^{kb} x_{ij} + \ddot{\gamma}_2 w_j + \varepsilon_{ij}$ $\varepsilon_{ij} \sim N(0,\ \theta^{kb})$	(9)

Continued

Table 1. (*Continued*)

MLM	NPMM[a]

+ Level-2 heteroscedasticity on w_j

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}w_j + u_{2j} + u_{3j}w_j + u_{1j}x_{ij} + e_{0ij} + e_{1ij}x_{ij}$$

$$\begin{bmatrix} u_{2j} \\ u_{3j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}\tau_{22} & & \\ \tau_{32} & \tau_{33} & \\ \tau_{12} & \tau_{13} & \tau_{11}\end{bmatrix}\right)$$

$$\begin{bmatrix}e_{0ij}\\e_{1ij}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_0^2 & \\ \sigma_{0,1} & \sigma_1^2\end{bmatrix}\right) \tag{12}$$

$$y_{ij}\big|_{c_{ij}=k,d_j=b} = \gamma_0^{kb} + \gamma_1^{kb}x_{ij} + \gamma_2^{\cdot b}w_j + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0,\theta^{kb}) \tag{11}$$

+ Level-1 heteroscedasticity on w_j

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}w_j + u_{2j} + u_{3j}w_j + u_{1j}x_{ij} + e_{0ij} + e_{1ij}x_{ij} + e_{2ij}w_j$$

$$\begin{bmatrix} u_{2j} \\ u_{3j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}\tau_{22} & & \\ \tau_{32} & \tau_{33} & \\ \tau_{12} & \tau_{13} & \tau_{11}\end{bmatrix}\right)$$

$$\begin{bmatrix}e_{0ij}\\e_{1ij}\\e_{2ij}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}\sigma_0^2 & & \\ \sigma_{0,1} & \sigma_1^2 & \\ \sigma_{0,2} & \sigma_{1,2} & \sigma_2^2\end{bmatrix}\right) \tag{14}$$

$$y_{ij}\big|_{c_{ij}=k,d_j=b} = \gamma_0^{kb} + \gamma_1^{kb}x_{ij} + \gamma_2^{kb}w_j + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0,\theta^{kb}) \tag{13}$$

Note. [a]For all NPMM models:

$$p(d_j = b) = \pi^{\cdot b} = \exp(\varpi^b)/\sum_{b=1}^{H}\exp(\varpi^b)$$

$$p(c_{ij} = k|d_j = b) = \pi^{k|b} = \exp(\omega^k + \delta^{kb})/\sum_{k=1}^{K}\exp(\omega^k + \delta^{kb})$$

For identification, $\varpi^H = \omega^K = \delta^{Kb} = \delta^{kH} = 0$

Table 2. Matrix specification of corresponding general NPMM (equation 15) and MLM (equation 16)

MLM general specification	NPMM general specification
$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\gamma} + \mathbf{w}'_{ij}\mathbf{u}_j + \mathbf{v}'_{ij}\mathbf{e}_{ij}$ $\quad \mathbf{u}_j \sim N(\mathbf{0}, \mathbf{T})$ $\quad \mathbf{e}_{ij} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ $\qquad\qquad (16)$	$y_{ij}\|_{c_{ij}=k,d_j=b} = \mathbf{z}'_{ij}\boldsymbol{\gamma}^{kb} + \varepsilon_{ij}$ $\qquad\qquad\quad \varepsilon_{ij} \sim N(0, \theta^{kb})$ $\quad (15)$ $p(d_j = b) = \pi^{\cdot b}$ $p(c_{ij} = k \mid d_j = b) = \pi^{k\|b}$

where:

$\mathbf{x}_{ij}$ vector of 1 and all L1 or L2 covariates with fixed components

$\boldsymbol{\gamma}$ vector of corresponding fixed components

$\mathbf{v}_{ij}$ vector of 1 and L1 or L2 covariates affecting L1 residuals

$\mathbf{e}_{ij}$ vector of L1 random effects modelling L1 heteroscedasticity

$\boldsymbol{\Sigma}$ covariance matrix of $\mathbf{e}_{ij}$

$\mathbf{w}_{ij}$ vector of 1 and all covariates with random components

$\mathbf{u}_j$ vector of corresponding L2 random effects

$\mathbf{T}$ covariance matrix of $\mathbf{u}_j$

$\boldsymbol{\Phi}$ submatrix of $\mathbf{T}$ containing covariances of random effects modelling L2 heteroscedasticity (see equation 37)

$\mathbf{w}_j$ submatrix of $\mathbf{w}_{ij}$ containing 1 and L2 covariates affecting L2 residuals (see equation 37)

where:

$\mathbf{z}_{ij}$ vector containing 1 and *all* L1 and L2 covariates

$\boldsymbol{\gamma}^{kb}$ vector of coefficients for class combination kb

θ^{kb} residual variance for class combination kb

$\pi^{\cdot b}$ and $\pi^{k\|b}$: see Table 1

Notes. L1 = level-1; L2 = level-2.

An indirect interpretation of equation (1), however, could consider this NPMM a discrete approximation of a corresponding MLM. For example, rather than the school-level and student-level latent classes representing distinct subpopulations, they may be approximating continuous distributions of deviations in maths scores at each level. To illustrate, we begin by defining the corresponding MLM in equation (2), the random-intercept-only (null) MLM. Specifically, in equation (2) we have $y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$, where outcome y_{ij} is modelled by a mean intercept, γ_{00}, cluster-specific deviation, $u_{0j} \sim N(0, \tau_{00})$, and level-1 residual, $e_{ij} \sim N(0, \sigma^2)$. In contrast to the NPMM, all variance (within-cluster and between-cluster) is modelled by continuously distributed residuals, rather than through discrete latent class variation.

Researchers fitting MLMs typically use this null MLM exclusively to calculate and interpret the ICC (e.g., Snijders & Bosker, 2012), which involves partitioning outcome variance into within-cluster (σ^2) and between-cluster (τ_{00}) components. Specifically, the ICC is the proportion of variance that is between clusters:

$$\text{ICC} = \frac{\tau_{00}}{\tau_{00} + \sigma^2}. \qquad (17)$$

Because the ICC is typically the only quantity of interpretive interest from the null MLM, to mathematically relate the null NPMM and null MLM – thus aiding in an indirect interpretation of NPMM – we first will focus on how the ICC would be computed from NPMM parameters.

Preliminarily, it is important to consider why one would want to compute the ICC from NPMM parameters. In an MLM context, researchers typically compute the ICC using equation (17) to determine if their degree of nesting is great enough to necessitate fitting MLMs and, if so, report their non-trivial ICC as a rationale for using the MLM. In an NPMM context, no such dependency measure currently exists that uses parameters of a fitted null NPMM model, and consequently there is often no rationale given for the necessity of fitting multilevel data with an NPMM as opposed to, say, fitting a single-level latent class regression model (see Park & Yu, 2015). Having such a measure of ICC for the NPMM would allow researchers to understand the degree of nesting in a familiar correlation metric, without having to fit a second model (i.e., a null MLM).

In MLMs, calculation of the ICC is straightforward. In NPMMs, however, there is no explicit (i.e., directly estimated) within-cluster or between-cluster variance component. Therefore, we must first partition variance analytically in the NPMM to obtain these quantities, and then use them in constructing the implied ICC. Because all clusters vary discretely in the NPMM (i.e., by level-2 class), partitioning variance into within-cluster and between-cluster components amounts to partitioning variance into *within level-2 class* and *between level-2 class* components, respectively. Fig 1 illustrates the concept of partitioning within-cluster and between-cluster variance in a null NPMM. In this figure, both panels depict a mixture of four kh class-combination densities generated from a null NPMM with two level-1 classes ($K = 2$) and two level-2 classes ($H = 2$).

As shown in Fig 1a, to examine *within-cluster variance*, we examine how parameters vary by level-1 class k *within* level-2 class h. Specifically, in Fig 1a, each vertical line represents a kh class-combination intercept, γ_0^{kh}. Comparing the two leftmost vertical lines (denoted γ_0^{11} and γ_0^{21}, respectively), we can see that there is class-combination intercept variance within level-2 class 1 (i.e., the intercepts γ_0^{11} and γ_0^{21}, both belonging to $h = 1$, differ). Assessing *within level-2 class variance* in class-combination intercepts (and/or residuals) parallels assessing within-cluster variance in MLM. Though γ_0^{kh} is specific to class combination kh, we can marginalize across k within h (denoted $E_{k|h}()$ where E represents expectation). We thus obtain *marginal level-2 class* parameters, denoted with a $\bullet h$ superscript: $E_{k|h}(\gamma_0^{kh}) = \gamma_0^{\bullet h}$. To examine *between-cluster variance*, we examine how marginal level-2 class parameters vary by h, that is, *between* level-2 classes, as shown in Fig 1b. There, each vertical line represents a marginal level-2 class intercept, $\gamma_0^{\bullet h}$. Comparing the two vertical lines (denoted $\gamma_0^{\bullet 1}$ and $\gamma_0^{\bullet 2}$), we can see how there is marginal level-2 class intercept variance between level-2 classes. Assessing this *between level-2 class variance* in marginal level-2 class intercepts parallels assessing between-cluster variance in MLMs.

With this background, we proceed to first calculate within level-2 class variance from the null NPMM. Because we are relating it to σ^2 in the MLM, it's termed an *implied* σ^2, denoted σ_{NP}^2. Specifically, we take the variance of the intercepts and residuals from equation (1) across level-1 classes within each level-2 class, denoted $\text{var}_{k|h}()$, and pool across level-2 classes, denoted $E_h()$, to obtain an overall variance,

$$\begin{aligned}
\sigma_{NP}^2 &= E_h[\text{var}_{k|h}(\gamma_0^{kh} + \varepsilon_{ij})] \\
&= E_h[E_{k|h}[(\gamma_0^{kh} - \gamma_0^{\bullet h})^2 + \theta^{kh}]].
\end{aligned} \tag{18}$$

The population expression in equation (18) can be converted to a sample estimate by replacing expectations with weighted sums and replacing parameters with corresponding estimates:
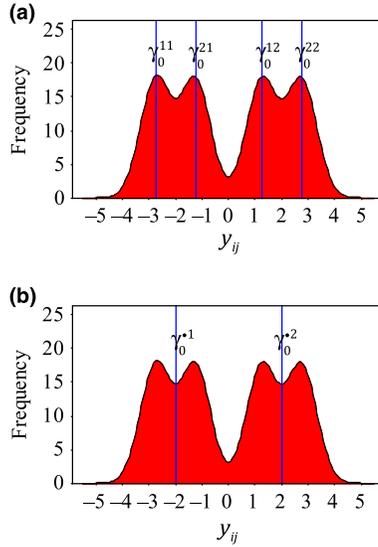
Figure 1. Conceptualizing within-cluster and between-cluster variance in the null NPMM (equation 1) through weighted within and between level-2 class intercept differences. (a) Comparing class-combination intercepts within level-2 class to assess within-cluster variance. (b) Comparing marginal level-2 class intercepts between level-2 classes to assess between-cluster variance. *Notes*. These plots show the kernel density of data generated (in SAS 9.4; SAS Institute Inc., Cary, NC, USA) from the null NPMM in equation (1). There are two level-1 classes ($K = 2$) and two level-2 classes ($H = 2$), making four kh class combinations. Each plot is a mixture of four normal densities; each density corresponds to a kh class combination. In (a), each vertical line represents a class-combination intercept, γ_0^{kh}. In (b), each vertical line represents a marginal level-2 class intercept, $\gamma_0^{\bullet h}$. By comparing these two plots, we can see conceptually how class-combination parameters (here intercepts, but in other contexts also/instead level-1 class probabilities) can vary within level-2 classes (by level-1 class). We can also see how marginal level-2 class parameters can vary between level-2 classes. Comparing parameters across level-1 classes within level-2 class parallels assessing within-cluster variance in a fitted MLM; comparing parameters between level-2 classes parallels assessing between-cluster variance in a fitted MLM.

$$\hat{\sigma}_{\mathrm{NP}}^2 = \sum_{b=1}^{H} \hat{\pi}^{\bullet b} \sum_{k=1}^{K} \hat{\pi}^{k|b} ((\hat{\gamma}_0^{kb} - \hat{\gamma}_0^{\bullet b})^2 + \hat{\theta}^{kb})$$

$$= \sum_{b=1}^{H} \sum_{k=1}^{K} \hat{\pi}^{kb} ((\hat{\gamma}_0^{kb} - \hat{\gamma}_0^{\bullet b})^2 + \hat{\theta}^{kb}). \tag{19}$$

We now calculate the between level-2 class variance from the null NPMM, which we term the *implied* τ_{00}, denoted $\tau_{00(\mathrm{NP})}$. To do so, we take the variance of marginal level-2 class intercepts between level-2 classes,

$$\tau_{00(\mathrm{NP})} = \mathrm{var}_b(E_{k|b}[\gamma_0^{kb}]) = \mathrm{var}_b(\gamma_0^{\bullet b}) = E_b[(\gamma_0^{\bullet b})^2] - E_b[\gamma_0^{\bullet b}]^2. \tag{20}$$

This population expression also can be converted to a sample estimate,

$$\hat{\tau}_{00(\mathrm{NP})} = \mathrm{var}_b \left(\sum_{k=1}^{K} \hat{\pi}^{k|b} \hat{\gamma}_0^{kb} \right) = \sum_{b=1}^{H} \hat{\pi}^{\cdot b} (\hat{\gamma}_0^{\cdot b})^2 - \left(\sum_{b=1}^{H} \hat{\pi}^{\cdot b} \hat{\gamma}_0^{\cdot b} \right)^2. \tag{21}$$

Using σ_{NP}^2 and $\tau_{00(\mathrm{NP})}$ from the NPMM, we then compute the implied ICC:[8]

$$\mathrm{ICC}_{\mathrm{NP}} = \frac{\tau_{00(\mathrm{NP})}}{\tau_{00(\mathrm{NP})} + \sigma_{\mathrm{NP}}^2}. \tag{22}$$

With these implied variance components now enumerated for the NPMM, we can better understand how NPMM parameters relate to the degree of nesting, that is, the magnitude of the ICC if an MLM were fitted. This can aid researchers in indirectly interpreting classes from a null NPMM. Specifically, examining equation (20), we see that greater spread of marginal level-2 class intercepts in NPMM implies larger between-cluster variance in a fitted MLM (increasing ICC). In other words, with more level-2 class separation, there is more distinction between clusters and, thus, greater effects of nesting. This is illustrated by comparing Fig 2a,b. Furthermore, examining equation (18), we see that smaller kb class-combination residual variance in NPMM implies smaller within-cluster variance in a fitted MLM (increasing ICC). This is illustrated by comparing Fig 2a,c. Also, from equation (18) we see that less spread of intercepts across level-1 classes within level-2 class in the NPMM implies smaller within-cluster variance in MLM (increasing ICC). This indicates less distinction across level-1 classes within level-2 class and, thus, less overall variability within clusters. This pattern is illustrated by comparing Fig 2a,d.[9]

In sum, in addition to directly interpreting kb class-combination parameter differences in the null NPMM, researchers can better understand how the NPMM accounts for nestedness by indirectly interpreting these parameters in relation to the familiar ICC from the MLM. That is, it is now possible for researchers to simultaneously consider two options: that nesting is driven by some continuous distribution of intercepts across clusters, or that nesting is driven by patterns of discrete class differences. We will revisit both in the empirical example in Section 7.

3. Relationship of NPMM to random slope MLM

Suppose now that a researcher is interested in fitting a more complex NPMM, involving a single level-1 covariate, x_{ij}, whose effect on y_{ij} is allowed to differ across level-2 classes, as in equation (5).[10] In equation (5), $\gamma_1^{\cdot b}$ denotes the level-2 class b slope of x_{ij}; this slope does not vary across k within b. In equation (5), the intercept γ_0^{kb} is again allowed to differ across all kb class combinations.

---

[8] Note that each term in equation (22) is a variance and is thus bounded by 0. The implied ICC is thus also bounded by 0.

[9] In the Fig 2 demonstration, between level-2 class differences are driven by intercepts that vary freely across class combination. For simplicity, generating kb class-combination probabilities are equal. Alternatively, even if parameters (e.g., intercepts) were equal across b for all k (i.e., $\gamma_0^{11} = \gamma_0^{12}$ and $\gamma_0^{21} = \gamma_0^{22}$), between level-2 class differences could still be driven by level-1 class *proportions* varying by level-2 class. Similar ICC patterns could be found.

[10] Covariates can also/instead predict class membership (e.g., Muthén & Asparouhov, 2009; Vermunt, 2003), but we do not include this here because our focus is on non-parametric multilevel mixture *regression* models, where covariates predict outcomes within class.

From a direct interpretation of classes, the marginal level-2 class slopes ($\gamma_1^{\cdot b}$) could be interpreted as substantively meaningful slopes corresponding to different subpopulations of clusters. For example, suppose the number of hours students spend studying at home (x_{ij}) has a strong effect on maths achievement in one school-level class but a weak effect in another school-level class, perhaps because these school-level classes differ in how effectively course material is taught in the classroom.

Alternatively, we can indirectly interpret these slopes as approximating an underlying continuous distribution of slopes across clusters. We delineate the indirect interpretation by first defining the MLM corresponding to the NPMM in equation (5) – namely, the random slope MLM in equation (6). In equation (6), there are normally distributed random effects at level-2 for the intercept, $u_{0j} \sim N(0, \tau_{00})$, and slope of x_{ij}, $u_{1j} \sim N(0, \tau_{11})$, which are allowed to covary (τ_{01}). This yields a 2×2 covariance matrix of random effects, $\mathbf{T}$. There is also a fixed component for the slope, denoted γ_{10}, and intercept, denoted γ_{00}. Let γ denote the vector of fixed effects. Here, $\gamma' = [\gamma_{00}\ \gamma_{10}]$. Next, to aid in an indirect interpretation, we compute the implied γ and implied $\mathbf{T}$ from equation (5) NPMM parameters.

To approximate γ using NPMM parameters, we first define γ^{kb} as the vector of all coefficients in the NPMM and take its expectation across k and b:

$$\gamma_{\mathrm{NP}} = E_b[E_{k|b}[\gamma^{kb}]] = \gamma^{\cdot\cdot}. \tag{23}$$

The corresponding sample estimate of equation (23) is

$$\hat{\gamma}_{\mathrm{NP}} = \sum_{b=1}^{H} \sum_{k=1}^{K} \hat{\pi}^{kb} \hat{\gamma}^{kb} = \hat{\gamma}^{\cdot\cdot}. \tag{24}$$

To approximate $\mathbf{T}$ using NPMM parameters, we first define $\gamma^{\cdot b}$ as a vector of marginal level-2 class coefficients in the NPMM with *implied* random effects. Here, $\gamma^{\cdot b'} = [\gamma_0^{\cdot b}\ \gamma_1^{\cdot b}]$. Because the intercept (γ_0^{kb}) is allowed to differ across all kb in equation (5), the marginal level-2 class b intercept, $\gamma_0^{\cdot b}$, is obtained as $E_{k|b}[\gamma_0^{kb}] = \gamma_0^{\cdot b}$. Because the slope of x_{ij} differs only across b in equation (5), the marginal level-2 class b slope $\gamma_1^{\cdot b}$ is directly estimated.[11] We can then take the variance of $\gamma^{\cdot b}$ across level-2 classes to obtain the implied $\mathbf{T}$, approximated using NPMM parameters:

$$\mathbf{T}_{\mathrm{NP}} = \mathrm{var}_b(\gamma^{\cdot b}) = E_b[\gamma^{\cdot b}\gamma^{\cdot b'}] - E_b[\gamma^{\cdot b}]E_b[\gamma^{\cdot b'}]. \tag{25}$$

Expanding this expression for our example yields a 2×2 implied $\mathbf{T}$ consisting of the implied intercept variance (as in equation 20), implied slope variance, and implied intercept–slope covariance:

$$\mathbf{T}_{\mathrm{NP}} = \begin{bmatrix} E_b[(\gamma_0^{\cdot b})^2] - E_b[\gamma_0^{\cdot b}]^2 \\ E_b[\gamma_0^{\cdot b}\gamma_1^{\cdot b}] - E_b[\gamma_0^{\cdot b}]E_b[\gamma_1^{\cdot b}] & E_b[(\gamma_1^{\cdot b})^2] - E_b[\gamma_1^{\cdot b}]^2 \end{bmatrix}. \tag{26}$$

The matrix expression in equation (25) would apply generally, that is, to any number of level-1 covariates whose effects in $\gamma^{\cdot b}$ differ across level-2 classes. The corresponding sample estimate of equation (25) is

---

[11] If the slope of x_{ij} does vary across kb, it would be computed as $\gamma_1^{\cdot b} = E_{k|b}[\gamma_1^{kb}]$, as done later.

(a) versus **(b)**: Increased marginal level-2 class intercept separation in NPMM increases ICC in fitted MLM

(a)

Fitted MLM $\widehat{\text{ICC}} = 0.81$
$\hat{\tau}_{00} = 7.56 \quad \hat{\sigma}^2 = 1.81$

(b)

Fitted MLM $\widehat{\text{ICC}} = 0.87$
$\hat{\tau}_{00} = 14.04 \quad \hat{\sigma}^2 = 1.81$

(a) versus **(c)**: Decreased class-combination residual variance in NPMM increases ICC in fitted MLM

(c)

Fitted MLM $\widehat{\text{ICC}} = 0.85$
$\hat{\tau}_{00} = 7.55 \quad \hat{\sigma}^2 = 1.33$

(a) versus **(d)**: Decreased class-combination intercept separation within L2 classes in NPMM increases ICC in fitted MLM

(d)

Fitted MLM $\widehat{\text{ICC}} = 0.91$
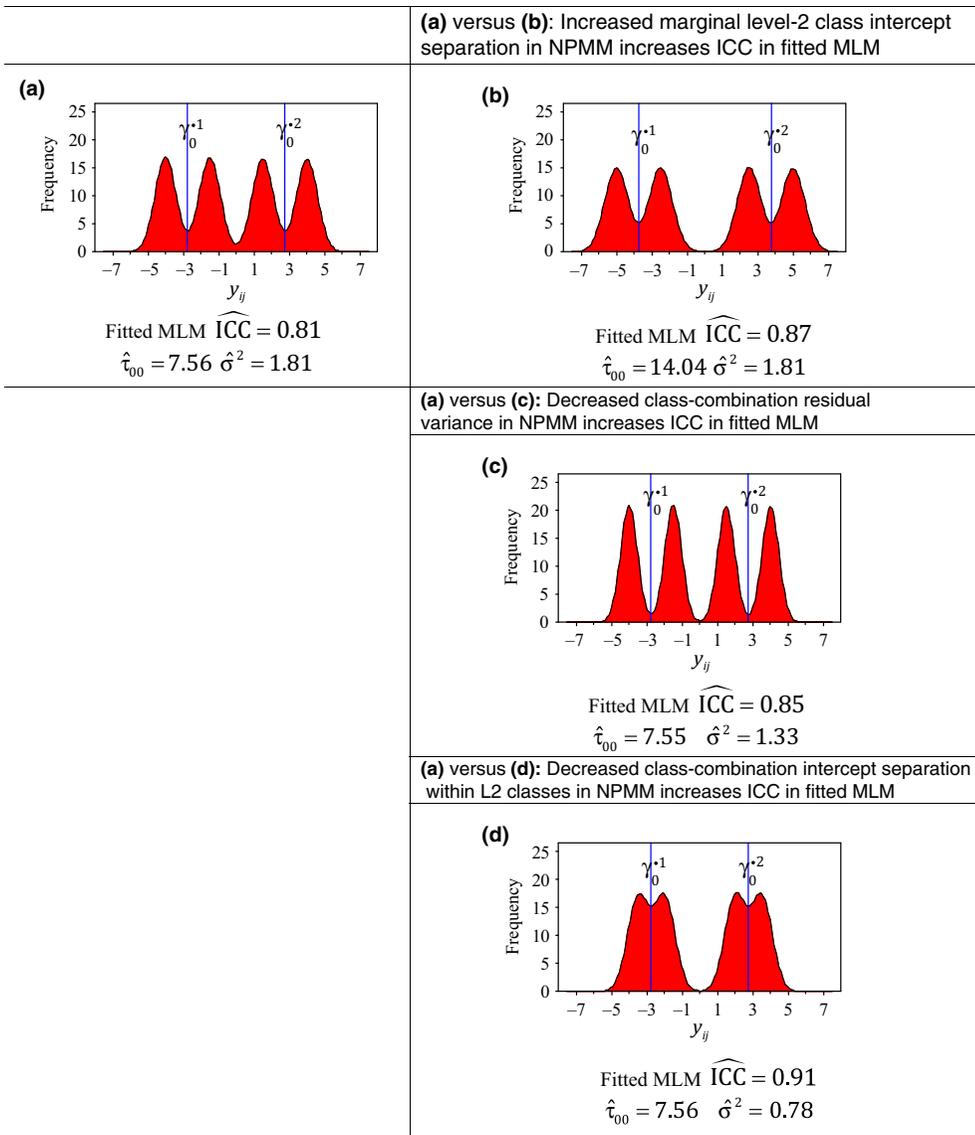$\hat{\tau}_{00} = 7.56 \quad \hat{\sigma}^2 = 0.78$

Figure 2. Relationship between NPMM parameters and the ICC from a fitted MLM: generating model = null NPMM (equation 1) and fitted model = null MLM (equation 2). (a) versus (b) Increased marginal level-2 class intercept separation in NPMM increases ICC in fitted MLM. (a) versus (c) Decreased class-combination residual variance in NPMM increases ICC in fitted MLM. (a) versus (d) Decreased class-combination intercept separation within level-2 classes in NPMM increases ICC in fitted MLM. *Notes.* Plots show kernel densities of data generated from a $K = 2$ (level-1 classes), $H = 2$ (level-2 classes) null NPMM and fitted with a random-intercept-only MLM. For illustrations in Figures 2–5, data were generated in SAS 9.4 and MLM estimates were computed using Proc MIXED. Note that the same pattern of results would hold with smaller ICC values; the conditions here were chosen for clarity of graphical illustration.

$$\hat{\mathbf{T}}_{\text{NP}} = \text{var}_b(\hat{\gamma}^{\cdot b}) = \sum_{b=1}^{H} \hat{\pi}^{\cdot b}\hat{\gamma}^{\cdot b}\hat{\gamma}^{\cdot b'} - \left(\sum_{b=1}^{H} \hat{\pi}^{\cdot b}\hat{\gamma}^{\cdot b}\right)\left(\sum_{b=1}^{H} \hat{\pi}^{\cdot b}\hat{\gamma}^{\cdot b'}\right). \qquad (27)$$

A key result from equations (25) to (27) is that the implied random effect variance from the NPMM will be non-zero only when there is variance in the marginal level-2 class parameters. This is illustrated in Fig 3a. On the other hand, if the marginal level-2 class parameters are equal, the implied random effect variance will be zero. This is illustrated in Fig 3b.

In the context of fitting an NPMM, if marginal level-2 class parameters are explicitly *constrained* equal, this yields an approximation of a fixed slope, as shown in equations (3) and (4). In the discussion in Section 8, we consider the potential for utilizing this knowledge in significance testing of implied random effect variances using NPMM parameters.

4. Relationship of NPMM to MLM with heteroscedasticity of residuals at level 1

Suppose now that a researcher is interested in fitting an NPMM where the effect of a level-1 covariate, x_{ij}, is allowed to vary by level-1 class within level-2 class, as in equation (7). That is, in the equation (7) NPMM, the slope of x_{ij} is allowed to vary not only across b (as in equation 5), but also across k within each b. The slope of x_{ij} in class combination kb is denoted γ_1^{kb}.

From a direct interpretation standpoint, we could consider x_{ij} slope variation by level-1 class within level-2 class to be substantively meaningful. For instance, within the school-level class where the number of hours spent studying (x_{ij}) strongly affects maths achievement, the effect of studying may still vary across student-level classes (perhaps because student-level classes differ in ability levels and test anxiety).

Alternatively, we could consider an indirect interpretation for this x_{ij} slope variation within level-2 class. However, no previous sources have addressed precisely what this slope variation in the NPMM would be approximating in MLM. Here we show that, under an indirect interpretation, slope variation within level-2 class in the NPMM approximates *heteroscedastic level-1 residual variance in MLM*, in which MLM residuals (e_{ij}) depend linearly on x_{ij}. To explain this, we start by defining the MLM corresponding to the NPMM in equation (7). This corresponding MLM is provided in equation (8), where $e_{ij} = e_{0ij} + e_{1ij} x_{ij}$ (e.g., Goldstein, 2011; Snijders & Berkhof, 2007). This non-standard MLM specification is conceptually equivalent to having a random effect of x_{ij} at level 1, meaning that the effect of x_{ij} varies across level-1 units (unlike a conventional random effect that varies across clusters). The equation (8) MLM[12] where residuals depend linearly on x_{ij} yields the following quadratic residual variance expression (Goldstein, 2011; Snijders & Bosker, 2012):

---

[12] Note that residuals can also depend linearly on a covariate in single-level contexts. Thus, the approximation outlined in this section is also relevant for relating single-level regression mixture models to conventional single-level regression models with residual heteroscedasticity.
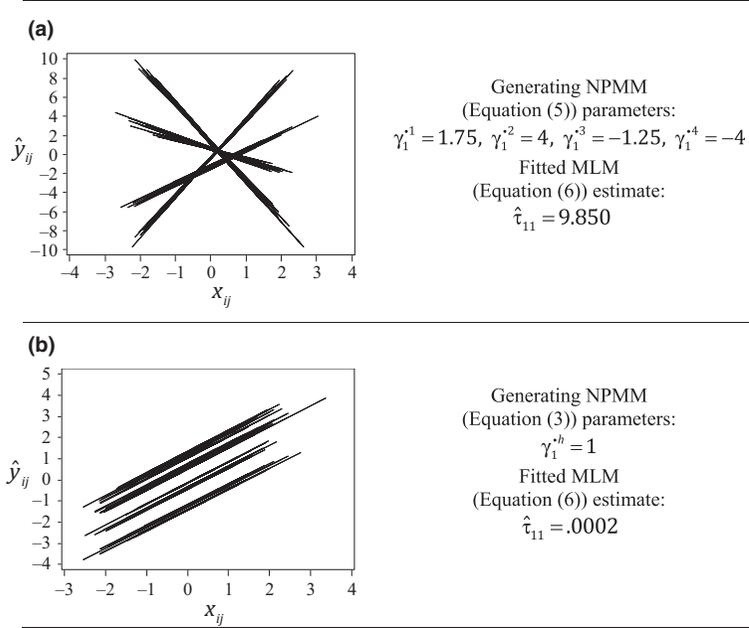
(a)

Generating NPMM
(Equation (5)) parameters:
$\gamma_1^{\cdot 1} = 1.75, \ \gamma_1^{\cdot 2} = 4, \ \gamma_1^{\cdot 3} = -1.25, \ \gamma_1^{\cdot 4} = -4$
Fitted MLM
(Equation (6)) estimate:
$\hat{\tau}_{11} = 9.850$

(b)

Generating NPMM
(Equation (3)) parameters:
$\gamma_1^{\cdot h} = 1$
Fitted MLM
(Equation (6)) estimate:
$\hat{\tau}_{11} = .0002$

Figure 3. Relationship between NPMM parameters and the random slope from a fitted MLM. (a) Implied random slope and (b) implied fixed slope. *Notes.* Both plots show predicted scores $(\hat{y}_{ij}|x_{ij}, \mathbf{u}_j)$ in a fitted MLM. Each line represents the regression line for a single cluster. Data were generated from a $K = 4, H = 4$ NPMM with a single cluster-mean-centred level-1 covariate. Analyses in (a) are of data generated from an NPMM with varying marginal level-2 class slopes of x_{ij}, whereas analyses in (b) are of data generated from an NPMM with constant marginal level-2 class slopes of x_{ij}. As such, the data for (a) are consistent with a random slope in a fitted MLM and the data for (b) are consistent with a fixed slope in a fitted MLM.

$$\sigma_{ij}^2 = \text{var}_{ij}(e_{0ij} + e_{1ij}x_{ij}) = \sigma_0^2 + x_{ij}^2\sigma_1^2 + 2x_{ij}\sigma_{0,1}. \qquad (28)$$

Variances σ_0^2 and σ_1^2 and covariance $\sigma_{0,1}$ are defined in equation (8). For instance, this MLM allowing for level-1 heteroscedasticity can be useful when within-cluster variability in maths achievement is expected to be smaller for students who spend moderate hours studying, but larger for students who spend many or few hours studying. Within cluster, achievement can be highly variable for students who study intensively (some perform well whereas others perform poorly due to test anxiety) or study infrequently (some perform poorly whereas others perform well due to high ability). Taken together, equation (28) clarifies that the residual variance in the equation (8) MLM is heteroscedastic: σ_{ij}^2 is specific to the x_{ij} score for observation i in cluster j.

To further clarify how the equation (7) NPMM indirectly approximates the MLM's level-1 heteroscedasticity in equation (28), we compute the corresponding quadratic variance expression with NPMM parameters. Specifically, we take the variance of $y_{ij}|_{c_{ij}=k,d_j=b}$ across level-1 classes within level-2 class, conditional on x_{ij}, and pool across level-2 classes:

$$\sigma^2_{ij(\mathrm{NP})} = E_b[\mathrm{var}_{k|b}(\gamma^{kb}_0 + \gamma^{kb}_1 x_{ij} + \varepsilon_{ij})]$$
$$= E_b[E_{k|b}[(\gamma^{kb}_0 - \gamma^{\bullet b}_0)^2 + x^2_{ij}(\gamma^{kb}_1 - \gamma^{\bullet b}_1)^2 + 2x_{ij}(\gamma^{kb}_0 - \gamma^{\bullet b}_0)(\gamma^{kb}_1 - \gamma^{\bullet b}_1) + \theta^{kb}]].$$

$$(29)$$

The sample estimate of equation (29) is

$$\hat{\sigma}^2_{ij(\mathrm{NP})} = \sum_{b=1}^{H} \sum_{k=1}^{K} \hat{\pi}^{kb}((\hat{\gamma}^{kb}_0 - \hat{\gamma}^{\bullet b}_0)^2 + x^2_{ij}(\hat{\gamma}^{kb}_1 - \hat{\gamma}^{\bullet b}_1)^2 + 2x_{ij}(\hat{\gamma}^{kb}_0 - \hat{\gamma}^{\bullet b}_0)(\hat{\gamma}^{kb}_1 - \hat{\gamma}^{\bullet b}_1) + \hat{\theta}^{kb}).$$

$$(30)$$

Comparing equation (29) to (28), the expectation of $(\gamma^{kb}_0 - \gamma^{\bullet b}_0)^2 + \theta^{kb}$ approximates σ^2_0, the expectation of $(\gamma^{kb}_1 - \gamma^{\bullet b}_1)^2$ approximates σ^2_1, and the expectation of $(\gamma^{kb}_0 - \gamma^{\bullet b}_0)(\gamma^{kb}_1 - \gamma^{\bullet b}_1)$ approximates $\sigma_{0,1}$.

More generally, with q covariates affecting level-1 residuals, the MLM level-1 residual variance in equation (28) is expressed as

$$\sigma^2_{ij} = \mathbf{v}'_{ij}\boldsymbol{\Sigma}\mathbf{v}_{ij} \qquad (31)$$

where the vector $\mathbf{v}_{ij}$ contains 1 and q covariates affecting level-1 residuals and $\boldsymbol{\Sigma}$ is a $(q + 1) \times (q + 1)$ covariance matrix of level-1 random effects used in modelling σ^2_{ij}. The corresponding expression implied by the NPMM generalizes to

$$\sigma^2_{ij(\mathrm{NP})} = E_b[\mathrm{var}_{k|b}(\mathbf{v}'_{ij}\boldsymbol{\upsilon}^{kb} + \varepsilon_{ij})]$$
$$= \mathbf{v}'_{ij}\big(E_b[E_{k|b}[(\boldsymbol{\upsilon}^{kb} - \boldsymbol{\upsilon}^{\bullet b})(\boldsymbol{\upsilon}^{kb} - \boldsymbol{\upsilon}^{\bullet b})' + \mathbf{a}\theta^{kb}\mathbf{a}']]\big)\mathbf{v}_{ij} \qquad (32)$$
$$= \mathbf{v}'_{ij}\boldsymbol{\Sigma}_{\mathrm{NP}}\mathbf{v}_{ij},$$

where $\mathbf{a} = \mathbf{v}_{ij}(\mathbf{v}'_{ij}\mathbf{v}_{ij})^{-1}$ and $\boldsymbol{\upsilon}^{kb}$ is the $(q + 1) \times 1$ subvector of γ^{kb} containing coefficients for $\mathbf{v}_{ij}$. In the sample, this yields

$$\hat{\sigma}^2_{ij(\mathrm{NP})} = \mathbf{v}'_{ij}\left(\sum_{b=1}^{H} \sum_{k=1}^{K} \hat{\pi}^{kb}((\hat{\boldsymbol{\upsilon}}^{kb} - \hat{\boldsymbol{\upsilon}}^{\bullet b})(\hat{\boldsymbol{\upsilon}}^{kb} - \hat{\boldsymbol{\upsilon}}^{\bullet b})' + \mathbf{a}\hat{\theta}^{kb}\mathbf{a}')\right)\mathbf{v}_{ij}$$
$$= \mathbf{v}'_{ij}\hat{\boldsymbol{\Sigma}}_{\mathrm{NP}}\mathbf{v}_{ij}.$$

$$(33)$$

From equations (29) and (30), a key result is evident: if there is *no* slope variation across level-1 class within level-2 class (e.g., $\gamma^{kb}_1 = \gamma^{\bullet b}_1$ for all b, as in equation 5), the linear and quadratic terms drop out. All that remains is a constant, implied homoscedastic residual variance, given by equation (19). This point is illustrated in Fig 4b in which, under these conditions, the variance of residuals $\hat{e}_{ij}$ from a fitted MLM is shown to be homoscedastic – constant across x_{ij}. Fig 4a depicts the alternative situation: when level-1 class slopes vary within level-2 class ($\gamma^{kb}_1 \neq \gamma^{\bullet b}_1$ for some b and k) the variance of residuals $\hat{e}_{ij}$ from a fitted MLM is heteroscedastic – greater at the extremes of x_{ij}.

In sum, in NPMM, slope differences by level-1 class within level-2 class afford not only a direct interpretation, but also an indirect interpretation in terms of accommodating level-1 heteroscedasticity in MLM. Given that the latter MLM specification is
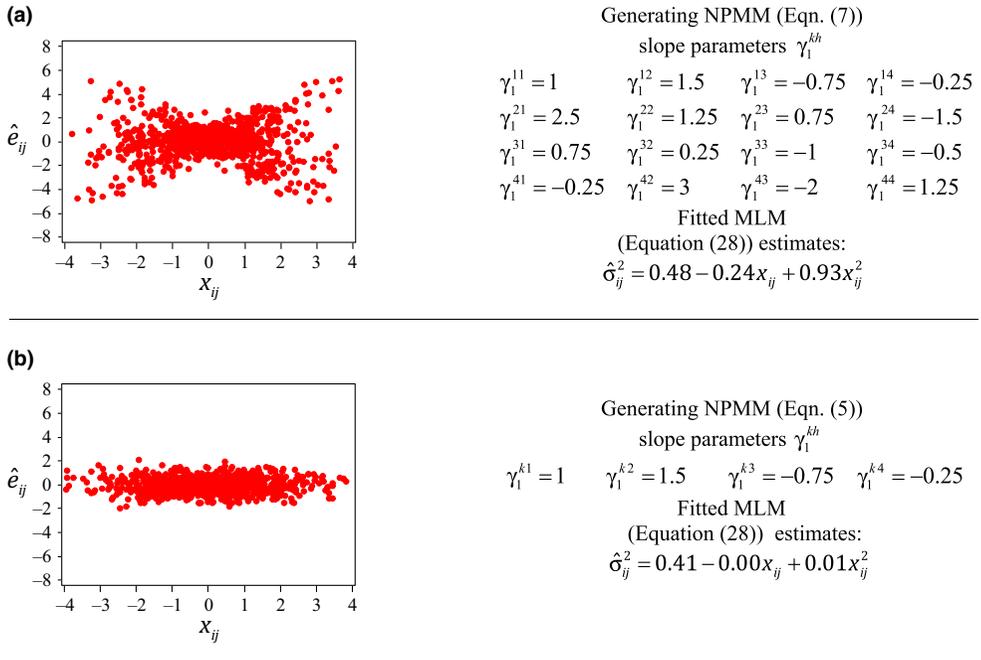
(a)

Generating NPMM (Eqn. (7))
slope parameters γ_1^{kh}

$\gamma_1^{11} = 1 \qquad \gamma_1^{12} = 1.5 \qquad \gamma_1^{13} = -0.75 \quad \gamma_1^{14} = -0.25$

$\gamma_1^{21} = 2.5 \qquad \gamma_1^{22} = 1.25 \quad \gamma_1^{23} = 0.75 \qquad \gamma_1^{24} = -1.5$

$\gamma_1^{31} = 0.75 \qquad \gamma_1^{32} = 0.25 \quad \gamma_1^{33} = -1 \qquad \gamma_1^{34} = -0.5$

$\gamma_1^{41} = -0.25 \quad \gamma_1^{42} = 3 \qquad \gamma_1^{43} = -2 \qquad \gamma_1^{44} = 1.25$

Fitted MLM
(Equation (28)) estimates:
$\hat{\sigma}_{ij}^2 = 0.48 - 0.24 x_{ij} + 0.93 x_{ij}^2$

(b)

Generating NPMM (Eqn. (5))
slope parameters γ_1^{kh}

$\gamma_1^{k1} = 1 \qquad \gamma_1^{k2} = 1.5 \qquad \gamma_1^{k3} = -0.75 \quad \gamma_1^{k4} = -0.25$

Fitted MLM
(Equation (28)) estimates:
$\hat{\sigma}_{ij}^2 = 0.41 - 0.00 x_{ij} + 0.01 x_{ij}^2$

Figure 4. Relationship between NPMM parameters and heteroscedasticity of residuals at level-1 in MLM. (a) Level-1 heteroscedasticity in fitted MLM and (b) Level-1 homoscedasticity in fitted MLM. *Notes.* Both plots show estimated level-1 residuals from a fitted MLM, $\hat{e}_{ij}$, plotted by x_{ij}. Data were generated from a $K = 4$, $H = 4$ NPMM with a single group-mean-centred level-1 covariate. Analyses in (a) are of data generated from an NPMM with varying slopes across level-1 class within level-2 class, while analyses in (b) are of data generated from an NPMM with constant slopes across level-1 class within level-2 class. As such, the data for (a) are consistent with level-1 heteroscedasticity in a fitted MLM and the data for (b) are consistent with level-1 homoscedasticity in a fitted MLM.

underused, unfamiliar (Goldstein, 2011; Snijders & Bosker, 2012) and, moreover, unavailable in some MLM software, an indirect interpretation of NPMM can be used to highlight or investigate level-1 heteroscedasticity.

5. Relationship of NPMM to MLM with heteroscedasticity of residuals at level-2

Now suppose a researcher is interested in fitting a more complex NPMM that also includes a level-2 covariate, w_j, such as student–teacher ratio. If we fit the NPMM in equation (9) where the slope of w_j does not vary across level-2 *or* level-1 classes $(\gamma_2^{\cdot\cdot})$, the corresponding MLM is defined as having a *fixed slope* of w_j, as shown in equation (10). Suppose instead we fit the NPMM in equation (11) where the slope of w_j (i.e., $\gamma_2^{\cdot b}$) varies across level-2 classes.

From a direct interpretation standpoint, one could consider variance in the marginal level-2 class slope of w_j between level-2 classes in equation (11) to represent substantively

meaningful slope differences between cluster-level subpopulations. For example, perhaps the effect of student-teacher ratio (w_j) on maths achievement is stronger for some school-level classes (e.g., those with more focus on class participation) compared to other school-level classes.

If we instead desired an indirect interpretation of equation (11), previous sources have not addressed what the w_j slope variation across level-2 class in the NPMM would be approximating in the MLM. Under an indirect interpretation, w_j slope variation across level-2 classes in the NPMM approximates *heteroscedastic level-2 residual variance in the MLM*, where MLM intercept residuals, u_{0j}, depend linearly on w_j. To illustrate this, we first define the MLM corresponding to the NPMM in equation (11). This non-standard MLM is provided in equation (12), where $u_{0j} = u_{2j} + u_{3j} w_j$ (see Goldstein, 2011; Snijders & Berkhof, 2007). Modelling level-2 heteroscedasticity in this way in the MLM essentially adds a random effect, u_{3j}, for a level-2 covariate at level 2, only u_{3j} is not interpreted as a conventional random effect, but as a means to model heteroscedasticity. The equation (12) MLM yields the following quadratic variance expression for intercept residuals (e.g., Goldstein, 2011; Snijders & Bosker, 2012):

$$\tau_{00j} = \text{var}(u_{2j} + u_{3j} w_j) = \tau_{22} + w_j^2 \tau_{33} + 2 w_j \tau_{32}. \tag{34}$$

The variances τ_{22} and τ_{33} and covariance τ_{32} were defined in equation (12). Such an expression can be useful when, for example, larger intercept residual variance is anticipated for more extreme student–teacher ratios. This heteroscedastic intercept variance, τ_{00j}, is specific to cluster j because it depends on the value of w_j.

To further illustrate this indirect interpretation of the equation (11) NPMM, we compute the quadratic variance expression corresponding to equation (34) that is implied by NPMM parameters, as shown in the following equation, which computes the variance of the sum of the marginal level-2 class intercept and slope of conditional on w_j:

$$\begin{aligned}
\tau_{00j(\text{NP})} &= \text{var}_b(\gamma_0^{\bullet b} + \gamma_2^{\bullet b} w_j) \\
&= E_b[(\gamma_0^{\bullet b} - \gamma_0^{\bullet\bullet})^2 + w_j^2(\gamma_2^{\bullet b} - \gamma_2^{\bullet\bullet})^2 + 2 w_j(\gamma_0^{\bullet b} - \gamma_0^{\bullet\bullet})(\gamma_2^{\bullet b} - \gamma_2^{\bullet\bullet})].
\end{aligned} \tag{35}$$

Comparing equation (35) to (34), the expectation of $(\gamma_0^{\bullet b} - \gamma_0^{\bullet\bullet})^2$ approximates the variance τ_{22}, the expectation of $(\gamma_2^{\bullet b} - \gamma_2^{\bullet\bullet})^2$ approximates the variance τ_{33}, and the expectation of $(\gamma_0^{\bullet b} - \gamma_0^{\bullet\bullet})(\gamma_2^{\bullet b} - \gamma_2^{\bullet\bullet})$ approximates the covariance τ_{32}. The sample counterpart to equation (35) is:

$$\hat{\tau}_{00j(\text{NP})} = \sum_{b=1}^{H} \hat{\pi}^{\bullet b}((\hat{\gamma}_0^{\bullet b} - \hat{\gamma}_0^{\bullet\bullet})^2 + w_j^2(\hat{\gamma}_2^{\bullet b} - \hat{\gamma}_2^{\bullet\bullet})^2 + 2 w_j(\hat{\gamma}_0^{\bullet b} - \hat{\gamma}_0^{\bullet\bullet})(\hat{\gamma}_2^{\bullet b} - \hat{\gamma}_2^{\bullet\bullet})). \tag{36}$$

We now consider a more general expression with p level-2 covariates affecting intercept residuals. The equation (34) MLM expression generalizes to

$$\tau_{00j} = \mathbf{w}_j' \mathbf{\Phi} \mathbf{w}_j, \tag{37}$$

where the vector $\mathbf{w}_j$ contains 1 and p level-2 covariates affecting intercept residuals and $\mathbf{\Phi}$ is a $(p + 1) \times (p + 1)$ covariance matrix, a submatrix of $\mathbf{T}$, used in modelling the

heteroscedastic intercept variance, τ_{00j}. The corresponding expression implied by the NPMM generalizes to

$$
\begin{aligned}
\tau_{00j(\text{NP})} &= \text{var}_b(\mathbf{w}_j'\boldsymbol{\alpha}^{\cdot b}) \\
&= \mathbf{w}_j' E_b[(\boldsymbol{\alpha}^{\cdot b} - \boldsymbol{\alpha}^{\cdot\cdot})(\boldsymbol{\alpha}^{\cdot b} - \boldsymbol{\alpha}^{\cdot\cdot})']\mathbf{w}_j \\
&= \mathbf{w}_j'\boldsymbol{\Phi}_{\text{NP}}\mathbf{w}_j,
\end{aligned}
\tag{38}
$$

where $\boldsymbol{\alpha}^{\cdot b}$ is a $(p + 1) \times 1$ subvector of $\boldsymbol{\gamma}^{\cdot b}$ containing marginal level-2 coefficients for $\mathbf{w}_j$. In the sample this yields

$$
\begin{aligned}
\hat{\tau}_{00j(\text{NP})} &= \mathbf{w}_j'\left(\sum_{b=1}^{H} \hat{\pi}^{\cdot b}(\hat{\boldsymbol{\alpha}}^{\cdot b} - \hat{\boldsymbol{\alpha}}^{\cdot\cdot})(\hat{\boldsymbol{\alpha}}^{\cdot b} - \hat{\boldsymbol{\alpha}}^{\cdot\cdot})'\right)\mathbf{w}_j \\
&= \mathbf{w}_j'\hat{\boldsymbol{\Phi}}_{\text{NP}}\mathbf{w}_j.
\end{aligned}
\tag{39}
$$

From equations (35) and (36) a key result is evident: if there is no marginal level-2 class slope variation for level-2 covariates in NPMM (e.g., $\gamma_2^{\cdot b} = \gamma_2^{\cdot\cdot}$ for all h, as in equation 9), the linear and quadratic terms drop out. This leaves the implied homoscedastic intercept variance expression given by equation (20). This point is illustrated in Fig 5b in which, under these conditions, the variance of level-2 intercept residuals in a fitted MLM, $\hat{u}_{0j}$, remains constant across w_j (i.e., homoscedastic). Fig 5a illustrates the opposite situation: provided $\gamma_2^{\cdot b} \neq \gamma_2^{\cdot\cdot}$ for some h, there is implied heteroscedasticity of level-2 intercept residuals in a fitted MLM. Thus, an indirect interpretation of NPMM offers the ability to explore and accommodate MLM level-2 intercept heteroscedasticity – another underused and unfamiliar MLM specification (Snijders & Bosker, 2012).

6. Relationship of NPMM to general matrix formulation of MLM

In previous sections, we demonstrated specific relationships between particular NPMM and MLM specifications in isolation. In this section, we integrate these results to (a) describe a *general* NPMM expression, (b) define its counterpart MLM, and (c) summarize how this general NPMM indirectly accommodates any number of covariates and *any combination* of implied fixed or random effects and implied homoscedasticity or heteroscedasticity of level-1 and/or level-2 residuals in MLM.

The general NPMM specification is given in equation (15) in Table 2. Here, intercepts and both level-1 and level-2 covariate slopes are allowed to freely vary across all kh class combinations. This general NPMM in equation (15) corresponds with the MLM specification in equation (16) when $\mathbf{v}_{ij}$ contains all covariates (i.e., $\mathbf{v}_{ij} = \mathbf{z}_{ij}$) and $\mathbf{w}_j$ contains all level-2 covariates. Specifically, the general equation (15) NPMM indirectly approximates the following in the corresponding MLM: random intercept, random slopes for all level-1 covariates (because their $\gamma^{\cdot b}$ vary across h), level-1 heteroscedasticity for all covariates (because their γ^{kb} vary across k within h), and level-2 heteroscedasticity for all level-2 covariates (because their $\gamma^{\cdot b}$ vary across h). To summarize how the MLM parameters in equation (16) are indirectly approximated with general NPMM parameters: γ is indirectly approximated by equation (24), $\mathbf{T}$ by equation (27), $\boldsymbol{\Sigma}$ by equation (33), and $\boldsymbol{\Phi}$ by equation (39). An R function for computing elements of each matrix is provided in our Appendix S1 to aid in indirect interpretation of the NPMM.
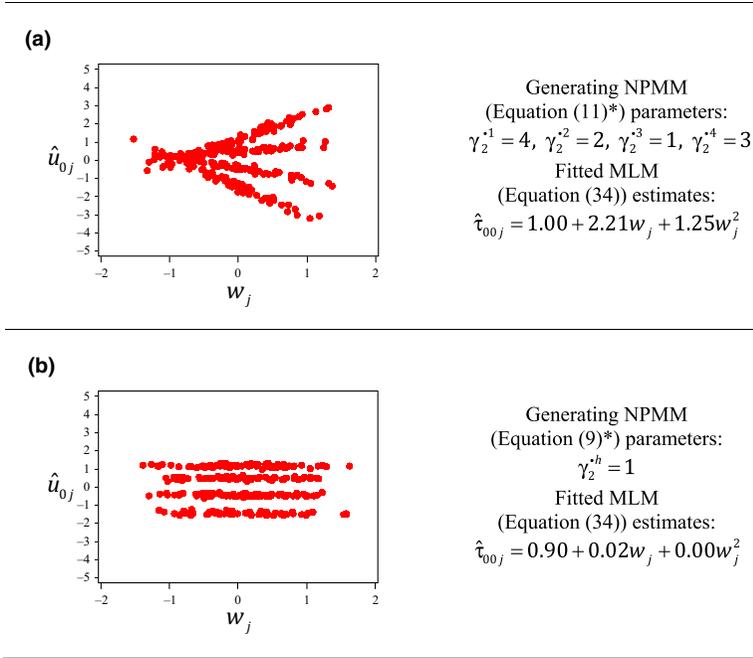
(a)

Generating NPMM
(Equation (11)*) parameters:
$\gamma_2^{\cdot 1} = 4, \ \gamma_2^{\cdot 2} = 2, \ \gamma_2^{\cdot 3} = 1, \ \gamma_2^{\cdot 4} = 3$
Fitted MLM
(Equation (34)) estimates:
$\hat{\tau}_{00j} = 1.00 + 2.21w_j + 1.25w_j^2$

(b)

Generating NPMM
(Equation (9)*) parameters:
$\gamma_2^{\cdot h} = 1$
Fitted MLM
(Equation (34)) estimates:
$\hat{\tau}_{00j} = 0.90 + 0.02w_j + 0.00w_j^2$

Figure 5. Relationship between NPMM parameters and heteroscedasticity of intercept residuals at level 2 in MLM. (a) Level-2 heteroscedasticity in fitted MLM and (b) Level-2 homoscedasticity in fitted MLM. *Notes.* Both plots show estimated level-2 intercept residuals in a fitted MLM, $\hat{u}_{0j}$, plotted by the level-2 covariate w_j. Data were generated from an NPMM. Analyses in (a) are of data generated from an NPMM with varying marginal level-2 class slopes, while analyses in (b) are of data generated from an NPMM with constant marginal level-2 class slopes. As such, the data for (a) are consistent with level-2 heteroscedasticity in a fitted MLM and the data for (b) are consistent with level-2 homoscedasticity in a fitted MLM. *Generating models included a single level-2 covariate, w_j, which is slightly different than equations (9) and (11), which also include x_{ij}.

7. Empirical example

We now present an empirical application implementing the indirect interpretation of NPMM and, further, demonstrate how this interpretation can provide additional utility beyond simply implementing either an MLM or a direct interpretation of the NPMM. Using a data set of students nested within schools from the Trends in International Mathematics and Science Study (TIMSS; Mullis, Martin, Gonzalez, & Chrostowski, 2004), we investigate effects of mathematics confidence (MATHCONF, cluster-mean-centred level-1 x_{ij}) and percentage of students receiving free/reduced lunch (PERLUNCH, level-2 w_j) on mathematics exam scores involving data/probability (MATH) for eighth graders.

Confidence should be positively associated with maths achievement (Pajares & Miller, 1994) as it is thought to enable students to better demonstrate their knowledge. School-level socioeconomic deprivation should be negatively associated with maths achievement (Bradley & Corwyn, 2002) as it is a proxy for disadvantages that impede learning (for further discussion, see Lubinski, 2009). Level-1 residual heteroscedasticity by confidence could arise if students with high or low confidence levels have more performance variability

within cluster (perhaps because some optimistic students misjudge their ability and underprepare whereas some pessimistic students overprepare). Hence, our illustrative fitted MLM consists of a random intercept, fixed slope of MATHCONF, fixed component of the slope of PERLUNCH, and residual heteroscedasticity at level 1 by MATHCONF and at level 2 by PERLUNCH. Our corresponding NPMM is most similar to equation (11) with two additional constraints for parsimony: class-combination residual variances equal and conditional probabilities of k given h equal across h, which also ensures an approximated fixed slope of MATHCONF. Using two versions of the Bayesian information criterion (BIC1 with $N = J$ (see Lukočienė, Varriale, & Vermunt, 2010) and BIC2 with $N =$ number of individuals), we compared the fit of all K and H ranging from 2 to 6 for this NPMM. Both BICs preferred $K = 2$, $H = 4$ (BIC1 = 78535.37, BIC2 = 78600.44).

7.1. Results

A researcher intending to fit MLMs may first report the ICC obtained from the null MLM ($\widehat{ICC} = .286$) to verify that clustering needs to be accounted for; slightly over a quarter of the total variance in performance is attributable to between-school differences. A researcher fitting NPMMs no longer needs to resort to *also* fitting a null MLM to quantify the degree of nesting, since (utilizing Section 2 equations) they can now compute the ICC from null NPMM parameters with $K = 2$, $H = 4$ ($\widehat{ICC}_{NP} = .272$). Note that the approximation of the ICC could depend on factors such as sample size. For instance, if sample size is small and adding classes leads to convergence problems even before the BIC stops improving, the approximation of the ICC may not be as good as for a larger sample size. Explicitly investigating such quality of approximation with differing numbers of classes in NPMM is noted as a future direction in the discussion in Section 8.

MLM results from the full model indicated significant ($p < .05$) fixed effects of MATHCONF and PERLUNCH in predicting maths achievement, significant level-1 residual heteroscedasticity, and non-significant level-2 residual heteroscedasticity. MLM results correspond to each horizontal black line in Figs 6 and 7. Figs 6 and 7 also show the NPMM approximation of each MLM parameter, for models of varying K and H (recall that the best BIC was $K = 2$, $H = 4$). Fig 7b shows how NPMM estimates can be interpreted as indirectly approximating the MLM's level-1 heteroscedastic residual variance, which depends on levels of MATHCONF (more extreme values indicating higher residual variance than more moderate values). Alternatively, from a direct interpretation, here this implied heteroscedasticity reflects the fact that within each school-level class, MATHCONF positively predicts MATH in some student-level classes (e.g., $k = 1$, where $\gamma_1^{1b} = 39.628$) but negatively predicts MATH in other student-level classes (e.g., $k = 2$; where $\gamma_1^{2b} = -15.097$). In the latter case, where $k = 2$, certain students may be adversely affected by overconfidence.

Consideration of the indirect interpretation of the NPMM offers us several practical benefits in this empirical setting. If we had purely considered a direct interpretation of the NPMM, results would suggest a typologically based intervention strategy, such as intervening with tutoring in maths and/or perception of skills for low-confidence students in $k = 1$ but high-confidence students in $k = 2$. The indirect interpretation forces researchers to consider the possibility that the observed student-level latent class slope differences are, instead, simply indicative of residual heteroscedasticity at level 1 and encourages researchers to design follow-up studies that would distinguish between these two generation processes (e.g., by investigating whether distinct etiological processes underlie $k = 1$ vs. $k = 2$). If we had purely considered a conventional *homoscedastic* MLM (as would be most typical of applied practice), we would have overlooked the
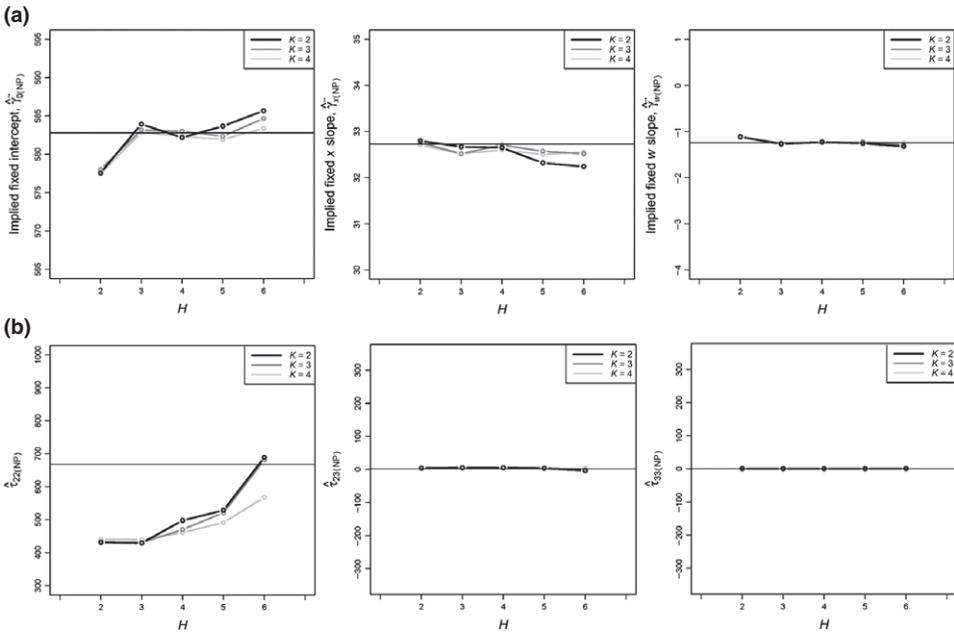
Figure 6. Empirical example results for implied fixed effects and implied level-2 heteroscedastic residual variance from the NPMM. (a) NPMM implied fixed effect estimates, plotted across level-2 class H by level-1 class K. (b) NPMM implied heteroscedastic intercept variance component estimates (see equation 34), plotted across H by K. *Note.* Best BIC is where $K = 2, H = 4$. Horizontal line = MLM estimate.

presence of level-1 heteroscedasticity and perhaps considered an oversimplified intervention strategy only targeting low-confidence students for maths tutoring. Using an indirectly interpreted NPMM as a preliminary exploratory tool in this context could spur further MLM model-building to account for level-l heteroscedasticity.

Although outside the scope of the current paper, note that Fig 6 shows that increasing H beyond best BIC can continue to improve the correspondence in level-2 variance components for the MLM (e.g., τ_{22}) and the NPMM approximations, as also found for other mixtures (Sterba *et al.*, 2012). Whereas fewer classes may aid in a parsimonious, substantive explanation of classes under direct interpretation, future research can investigate whether more NPMM classes continue to be useful for indirect approximation.

8. Discussion

Multilevel mixtures are increasingly used in analysing nested data. We focused on one such model, the non-parametric multilevel regression mixture model. Although the NPMM has been motivated for use in indirect interpretation of classes, the analytic relationships between alternative specifications of NPMMs and MLMs had not been enumerated. The current paper filled this gap by delineating how NPMM parameters relate to, and can be used to approximate, the (a) ICC in MLMs, (b) implied random coefficient means and (co)variances in MLMs, and (c) implied heteroscedasticity of
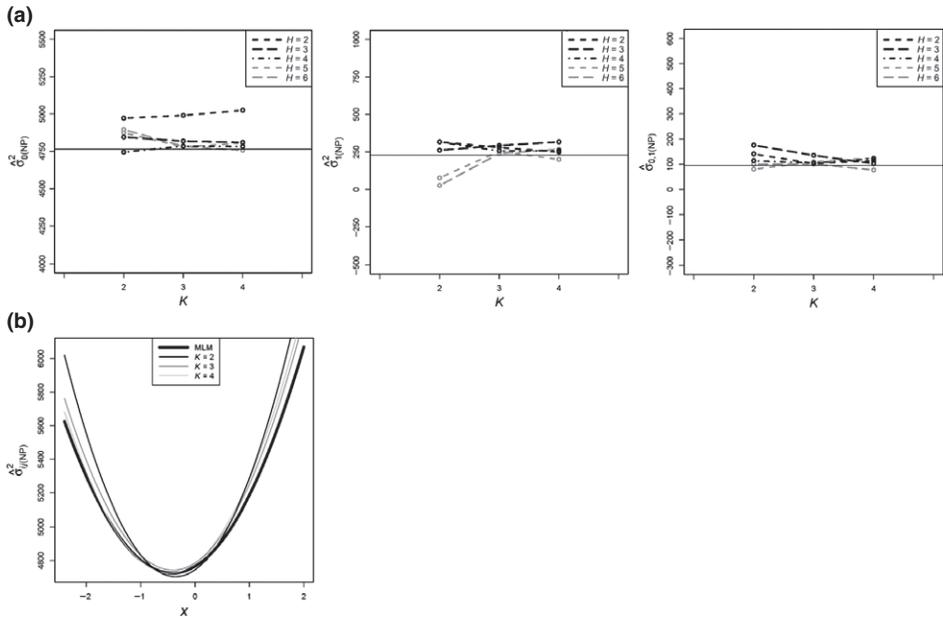
(a)

(b)

Figure 7. Empirical example results for implied level-1 heteroscedastic residual variance. (a) NPMM implied heteroscedastic level-1 residual variance component estimates (see equation 28), plotted across level-1 class K by level-2 class H. (b) NPMM implied heteroscedastic level-1 residual variance estimate, plotted across x_{ij} (math confidence) by K (for $H = 2$). *Note.* Best BIC is where $K = 2$, $H = 4$. Horizontal line = MLM estimate.

Table 3. Modifying the general NPMM (equation 15) with constraints to mirror corresponding MLM specifications

MLM specification	Corresponding constraint on the general NPMM in equation (15)	
Fixed intercept	All marginal level-2 class intercepts held equal	$\gamma_0^{\bullet 1} = \ldots = \gamma_0^{\bullet H}$
Fixed slope	All marginal level-2 class slopes of level-1 covariates held equal	$\gamma_p^{\bullet 1} = \ldots = \gamma_p^{\bullet H}$
Homoscedasticity of residuals at level 1	All level-1 slopes held equal within all level-2 classes	$\gamma_p^{11} = \ldots = \gamma_p^{K1}; \ldots; \gamma_p^{1H} = \ldots = \gamma_p^{KH}$
Homoscedasticity of residuals at level 2	All marginal level-2 class slopes of level-2 covariates held equal	$\gamma_p^{\bullet 1} = \ldots = \gamma_p^{\bullet H}$

residuals at level 1 and level 2 in MLMs. Level-1 and level-2 residual heteroscedasticity is currently underinvestigated in MLMs (Goldstein, 2011; Korendijk, Maas, Moerbeek, & Van der Heijden, 2008; Snijders & Berkhof, 2007). We showed how the NPMM offers a novel approach to exploring and modelling phenomena that would manifest as heteroscedasticity in the MLM. Even if researchers are primarily interested in directly interpreting NPMM classes, the indirect interpretation can be

considered simultaneously and contrasted with substantively driven explanations of classes. Next, we provide a practical summary of fitting special cases of NPMM using model constraints. We then address extensions, software, limitations, and future directions.

8.1. Constrained special cases of the general NPMM

In previous sections, we used a *build-up* approach to describe increasingly complex NPMM specifications, from Tables 1 and 2. More parsimonious special case mixture models can also be obtained by placing certain constraints on the general NPMM in equation (15). These constraints are described in Table 3, alongside a listing of the corresponding MLM specifications, to aid indirect interpretation of the NPMM. The NPMM can also be constrained to yield other types of mixture models: a single-level regression mixture when $H = 1$ and $K > 1$ (e.g., DeSarbo & Cron, 1988; Wedel & DeSarbo, 1994) and a non-parametric random coefficient mixture when $K = 1$ and $H > 1$ (e.g., Vermunt & Van Dijk, 2001).

8.2. Software syntax and tools

The Appendix S1 provides M*plus* syntax for the general NPMM. It also provides an R function, npmmApproximation, that reads in M*plus* results with any number of x_{ij} and w_j and outputs approximation calculations for equations (24), (27), (33), and (39) (implied fixed effects, random effect (co)variances, and level-1 residual variance components) for a given NPMM, or a range of NPMMs with different K and H. Plots similar to Figs 6 and 7 can be automatically generated.

8.3. When would using an MLM be preferred to fitting an NPMM?

It is important to consider when an MLM would be preferred to an indirectly interpreted NPMM. First, pragmatically, an MLM may be preferred with modest sample sizes. Because the MLM is fully parametric, there tend to be fewer parameters to be estimated in an MLM, as compared to its counterpart NPMM, in each row of Table 1. Second, NPMM may also be less useful when a researcher is confident that the distributional assumptions of the MLM are upheld (e.g., normality of random effects and correct specification of the random effects distribution); in this case the relaxed distributional assumptions of NPMM may be less appealing. Lastly, researchers are presently unable to use an indirectly interpreted NPMMs for multilevel designs in which indirect approximations of MLM parameters have not yet been analytically developed. For instance, in the current paper, we did not address the approximation of MLM parameters in cross-classified, multiple-membership, dynamic group membership, or partial nesting designs (e.g., Cafri, Hedeker, & Aarons, 2015; Goldstein, 2011; Sterba, 2016). Future work can derive such approximations to facilitate the use of indirectly interpreted NPMMs for more complex multilevel specifications.

8.4. When would using an NPMM as a non-parametric approximation be preferred to fitting an MLM?

Conversely, it is important to consider when an indirectly interpreted NPMM would be preferred to the more widely used MLM. First, the NPMM can potentially give a more

accurate representation of random effect variances when the random effects are markedly non-normal in the population – thus violating distributional assumptions of the MLM (Asparouhov & Muthén, 2008; Brame *et al.*, 2006; Wall, Guo, & Amemiya, 2012). Second, the NPMM can be advantageous in terms of estimability and computation time when y_{ij} is discrete, and particularly when between-cluster variability is anticipated in many slopes (Vermunt, 2004, 2008, 2010). In this situation, numerical integration is required for fitting an MLM, with one dimension of integration needed per random effect. Such integration is not necessary for fitting an NPMM. Third, the NPMM would be preferred for researchers explicitly interested in considering not only an indirect but also a direct interpretation of classes, since there are no classes in MLM. Fourth, as mentioned previously, since the NPMM can approximate non-standard MLM specifications (i.e., heteroscedasticity at level 1 and level 2), it might be preferred at an exploratory phase of modelling when researchers have no *a priori* hypotheses regarding residual variance structure. Fitting an NPMM may provide justification for exploring residual heteroscedasticity. Note that if, instead, residual heteroscedasticity is anticipated *a priori*, it could be parametrically modelled and tested within an MLM framework using some MLM software packages (e.g., MLwiN) but not others (e.g., lme4 in R). Finally, as explained in the next subsection, NPMM provides an alternative for testing the significance of implied random effect variances that can avoid limitations of approaches commonly implemented in MLM software (e.g., z-tests).

8.5. Extension: Testing random effect variances implied by the NPMM

The significance of indirectly approximated random effects, ICC, and heteroscedasticity could be tested using the NPMM by comparing models with and without particular sets of constraints (using Wald or likelihood ratio tests [LRTs]). Here, it is particularly intriguing to discuss this approach for testing implied random effect variances using an NPMM, given the methodological difficulties arising when testing the null hypothesis that a random effect variance equals 0 using an MLM. Specifically, using the MLM, this null hypothesis lies on a boundary of the parameter space (0), violating regularity conditions of conventional LRTs.[13] Although alternative testing procedures exist – for example, adjusted alpha-levels (Fitzmaurice, Laird, & Ware, 2004), unconstrained variance estimation (Dijkstra, 1992), or modified reference distributions (Stoel, Garre, Dolan, & van den Wittenboer, 2006) – they have additional drawbacks in generality or availability (Savalei & Kolenikov, 2008). Using the NPMM, however, we could test the significance of *implied* random effect variances using conventional LRT or Wald tests because 0 does not lie on the boundary of the parameter space. Testing the implied intercept variance in the NPMM involves testing $H_0 : \gamma_0^{\cdot 1} = \ldots = \gamma_0^{\cdot H}$, where df = $H - 1$ (i.e., testing the equality of H marginal level-2 class intercepts). Testing an implied slope variance in the NPMM involves testing $H_0 : \gamma_p^{\cdot 1} = \ldots = \gamma_p^{\cdot H}$ where df = $H - 1$. We avoid the boundary issue because each pairwise difference in these coefficients theoretically could be any value. To illustrate, we tested random slope variances for the two generated data sets from Fig 3, using Wald tests. As expected, the test of $H_0 : \gamma_1^{\cdot 1} = \ldots = \gamma_1^{\cdot H}$ was significant for Fig 3a ($\chi_3^2 = 13665.76$; $p < .0001$) and non-significant for Fig 3b ($\chi_3^2 = 0.35$; $p = .95$). Syntax for these tests is in the

---

[13] The asymptotic distribution of the LRT statistic is instead a mixture of chi-square distributions (Shapiro, 1985).

Appendix S1. Future research can further investigate the utility of such testing procedures.

8.6. Future directions

Several limitations can be noted which serve as future directions. First, future research could relate to other variations of the NPMM and MLM. For example, cross-level interactions could be included in both the general MLM and NPMM expressions (equations 16 and 15). Second, although we distinguished theoretically between direct and indirect interpretations, it is difficult in practice to determine which function classes are serving (especially under model misspecification). Future simulations can address this issue, as has been done for other mixtures (Lubke & Neale, 2006, 2008). Regardless, it is now more feasible for researchers to understand and consider both interpretations of the NPMM. Throughout this paper, we have discussed the possibility of considering both simultaneously, as has been suggested in other contexts (Nagin, 2005).

Third, it is important to consider that there are ways to non-parametrically approximate continua during the model estimation phase (using non-parametric maximum likelihood estimation [NPMLE]; Skrondal & Rabe-Hesketh, 2004) rather than during the model specification phase, as in the NPMM. The quality of the indirect approximation afforded by both NPMLE and the NPMM depends to some degree on the number of components (i.e., the number of points of support in NPMLE or number of latent classes in NPMM), though this number is obtained differently in each approach. The quality of the indirect approximation by NPMLE should be investigated and compared with the indirect application of the NPMM in future work. Note, however, that our NPMM approach has the advantage of also affording a direct interpretation of classes, which can be substantively compelling.

8.7. Conclusions

With our analytic approach, researchers can understand and visualize how the NPMM serves as a non-parametric approximation of the MLM. Researchers are now better able to consider two possibilities: that classes in NPMM represent distinct subpopulations or that classes can approximate underlying continuous distributions of effects, such as those modelled in the MLM.

Acknowledgements

References

Asparouhov, T., & Muthén, B. (2008). Multilevel mixture models. In G. R. Hancock & K. M. Samuelson (Eds.), *Advances in latent variable mixture models* (pp. 27–51). Charlotte, NC: Information Age Publishing.

Bauer, D. J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling*, *12*, 513–535. doi:10.1207/s15328007sem1204_1

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, *42*, 757–786. doi:10.1080/00273170701710338

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338–363. doi:10.1037/1082-989x.8.3.338

Bergman, L. R. (2015). Challenges for person-oriented research: Some considerations based on Laursen's article 'I Don't Quite Get it…: Personal Experiences with the Person-Oriented Approach'. *Journal for Person-Oriented Research*, *1*, 163–169. doi:10.17505/jpor.2015.17

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371–399. doi:10.1146/annurev.psych.53.100901.135233

Brame, R., Nagin, D. S., & Wasserman, L. (2006). Exploring some analytical characteristics of finite mixture models. *Journal of Quantitative Criminology*, *22*, 31–59. doi:10.1007/s10940-005-9001-8

Cafri, G., Hedeker, D., & Aarons, G. A. (2015). An introduction and integration of cross-classified, multiple membership, and dynamic group random-effects models. *Psychological Methods*, *20*, 407–421. doi:10.1037/met0000043

DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, *5*, 249–282. doi:10.1007/bf01897167

Dijkstra, T. K. (1992). On statistical inference with parameter estimates on the boundary of the parameter space. *British Journal of Mathematical and Statistical Psychology*, *45*, 289–309. doi:10.1111/j.2044-8317.1992.tb00994.x

Finch, W. H., & Marchant, G. J. (2013). Application of multilevel latent class analysis to identify achievement and socio-economic typologies in the 20 wealthiest countries. *Journal of Educational and Developmental Psychology*, *3*, 201–221. doi:10.5539/jedp.v3n1p201

Fitzmaurice, G., Laird, N., & Ware, J. (2004). *Applied longitudinal analysis*. Hoboken, NJ: Wiley.

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). New York, NY: Wiley.

Henry, K. L., & Muthén, B. (2010). Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling*, *17*, 193–215. doi:10.1080/10705511003659342

Hox, J. J. (2010). *Multilevel analysis. Techniques and applications* (2nd ed.). New York, NY: Routledge.

Kaplan, D., & Keller, B. (2011). A note on cluster effects in latent class analysis. *Structural Equation Modeling*, *18*, 525–536. doi:10.1080/10705511.2011.607071

Karakos, L. H. (2015). *Understanding civic engagement among youth in diverse contexts* (Unpublished doctoral dissertation). Vanderbilt University, Nashville, TN.

Korendijk, E. J., Maas, C. J., Moerbeek, M., & Van der Heijden, P. G. (2008). The influence of misspecification of the heteroscedasticity on multilevel regression parameter and standard error estimates. *Methodology*, *4*, 67–72. doi:10.1027/1614-2241.4.2.67

Lubinski, D. (2009). Cognitive epidemiology: With emphasis on untangling cognitive ability and socioeconomic status. *Intelligence*, *37*, 625–633. doi:10.1016/j.intell.2009.09.001

Lubke, G., & Neale, M. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research*, *41*, 499–532. doi:10.1207/s15327906mbr4104_4

Lubke, G., & Neale, M. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models. *Multivariate Behavioral Research*, *43*, 592–620. doi:10.1080/00273170802490673

Lukočienė, O., Varriale, R., & Vermunt, J. K. (2010). The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis. *Sociological Methodology*, *40*, 247–283. doi:10.1111/j.1467-9531.2010.01231.x

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.

Mullis, I. V., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Muthén, B., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In F. G. Davidian, M. G. Verbeke & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC Press.

Muthén, B., & Asparouhov, T. (2009). Multilevel regression mixture analysis. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, *172*, 639–657. doi:10.1111/j.1467-985x.2009.00589.x

Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus (version 7.31) [computer software]*. Los Angeles, CA: Muthén & Muthén.

Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.

Pajares, F., & Miller, M. D. (1994). The role of self-efficacy and self-concept beliefs on mathematical problem-solving: A path analysis. *Journal of Educational Psychology*, *86*, 193–203. doi:10.1037/0022-0663.86.2.193

Park, J., & Yu, H.-T. (2015). The impact of ignoring the level of nesting structure in a multilevel latent class model. *Educational and Psychological Measurement*. Advance online publication. doi:10.1177/0013164415618240

Pek, J., Sterba, S. K., Kok, B. E., & Bauer, D. J. (2009). Estimating and visualizing non-linear relations among latent variables: A semiparametric approach. *Multivariate Behavioral Research*, *44*, 407–436. doi:10.1080/00273170903103290

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Sampson, R. J., & Laub, J. H. (2005). Seductions of method: Rejoinder to Nagin and Tremblay's 'Developmental trajectory groups: Fact or fiction?'. *Criminology*, *43*, 905–913. doi:10.1111/j.1745-9125.2005.00027.x

Savalei, V., & Kolenikov, S. (2008). Constrained versus unconstrained estimation in structural equation modeling. *Psychological Methods*, *13*, 150–170. doi:10.1037/1082-989x.13.2.150

Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika*, *72*, 133–144. doi:10.1093/biomet/72.1.133

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.

Snijders, T. A., & Berkhof, J. (2007). Diagnostic checks for multilevel models. In J. de Leeuw & M. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 139–173). New York, NY: Springer.

Snijders, T. A., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, UK: Sage.

Sterba, S. K. (2016). Partially nested designs in psychotherapy trials: A review of modeling developments. *Psychotherapy Research*. Advance online publication. doi:10.1080/10503307.2015.1114688

Sterba, S. K., Baldasaro, R. E., & Bauer, D. J. (2012). Factors affecting the adequacy and preferability of semiparametric groups-based approximations of continuous growth trajectories. *Multivariate Behavioral Research*, *47*, 590–634. doi:10.1080/00273171.2012.692639

Sterba, S. K., & Bauer, D. J. (2014). Predictions of individual change recovered with latent class or random coefficient growth models. *Structural Equation Modeling*, *21*, 342–360. doi:10.1080/10705511.2014.915189

Stoel, R. D., Garre, F. G., Dolan, C., & van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods*, *11*, 439–455. doi:10.1037/1082-989x.11.4.439

Titterington, D. M., Smith, A. F., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York, NY: Wiley.

Van Horn, M. L., Feng, Y., Kim, M., Lamont, A., Feaster, D., & Jaki, T. (2016). Using multilevel regression mixture to identify level-1 heterogeneity in level-2 effects. *Structural Equation Modeling: A Multidisciplinary Journal*,, *23*, 259–269. doi:10.1080/10705511.2015.1035437

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, *33*, 213–239. doi:10.1111/j.0081-1750.2003.t01-1-00131.x

Vermunt, J. K. (2004). An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. *Statistica Neerlandica*, *58*, 220–233. doi:10.1046/j.0039-0402.2003.00257.x

Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, *17*, 33–51. doi:10.1177/0962280207081238

Vermunt, J. K. (2010). Mixture models for multilevel datasets. In J. Hox & J. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 59–81). New York, NY: Routledge.

Vermunt, J. K., & Magidson, J. (2005). Hierarchical mixture models for nested data structures. In W. Weihs & G. Wolfgang (Eds.), *Classification – the ubiquitous challenge* (pp. 240–247). Berlin, Germany: Springer.

Vermunt, J. K., & Van Dijk, L. (2001). A nonparametric random-coefficients approach: The latent class regression model. *Multilevel Modelling Newsletter*, *13*, 6–13. doi:10.1080/00273171.2012.658339

Wall, M. M., Guo, J., & Amemiya, Y. (2012). Mixture factor analysis for approximating a nonnormally distributed continuous latent factor with continuous and dichotomous observed variables. *Multivariate Behavioral Research*, *47*, 276–313. doi:10.1080/00273171.2012.658339

Wedel, M., & DeSarbo, W. S. (1994). A review of latent class regression models and their applications. In R. P. Bagozzi (Ed.), *Advanced methods for marketing research* (pp. 353–388). Cambridge, MA: Blackwell.

Yu, H. T., & Park, J. (2014). Simultaneous decision on the number of latent clusters and classes for multilevel latent class models. *Multivariate Behavioral Research*, *49*, 232–244. doi:10.1080/00273171.2014.900431

Supporting Information

The following supporting information may be found in the online edition of the article:

Appendix S1. R function to calculate approximations and Mplus code to fit and test NPMMs.