Psychological Methods
2018, Vol. 23, No. 3, 434–457

1082-989X/18/$12.00 http://dx.doi.org/10.1037/met0000139

A Framework of R-Squared Measures for Single-Level and Multilevel Regression Mixture Models

Jason D. Rights and Sonya K. Sterba
Vanderbilt University

Abstract

Psychologists commonly apply regression mixture models in single-level (i.e., unclustered) and multilevel (i.e., clustered) data analysis contexts. Though researchers applying nonmixture regression models typically report R-squared measures of explained variance, there has been no general treatment of R-squared measures for single-level and multilevel regression mixtures. Consequently, it is common for researchers to summarize results of a fitted regression mixture by simply reporting class-specific regression coefficients and their associated *p* values, rather than considering measures of effect size. In this article, we fill this gap by providing an integrative framework of R-squared measures for single-level regression mixture models and multilevel regression mixture models (with classes at Level-2 or both levels). Specifically, we describe 11 R-squared measures that are distinguished based on what the researcher chooses to consider as *outcome variance* and what sources the researcher chooses to contribute to *predicted variance*. We relate these measures analytically and through simulated graphical illustrations. Further, we demonstrate how these R-squared measures can be decomposed in novel ways into substantively meaningful sources of explained variance. We describe and demonstrate new software tools to allow researchers to compute these R-squared measures and decompositions in practice. Using 2 empirical examples, we show how researchers can answer distinct substantive questions with each measure and can gain insights by interpreting the set of measures in juxtaposition to each other.

Translational Abstract

Regression mixture models allow regression coefficients (intercepts and slopes) to vary by unobserved group, or latent class. Psychologists commonly apply such models in single-level (i.e., unclustered) and multilevel (i.e., clustered, such as students nested within schools) data analysis contexts. Though researchers applying nonmixture regression models typically report an R-squared (defined as the proportion of variance that is explained by the model), there has been no general treatment of R-squared measures for single-level and multilevel regression mixtures. Consequently, it is common for researchers to summarize results of a fitted regression mixture by simply reporting class-specific regression coefficients and their associated *p* values, rather than considering measures of effect size. In this article, we fill this gap by providing an integrative framework of R-squared measures for single-level regression mixture models and multilevel regression mixture models (with classes at Level-2 or both levels). Specifically, we describe 11 R-squared measures that are distinguished based on what the researcher chooses to consider as *outcome variance* and what sources the researcher chooses to contribute to *predicted variance*. We relate these measures analytically and through simulated graphical illustrations. Further, we demonstrate how these R-squared measures can be decomposed in novel ways into substantively meaningful sources of explained variance. We describe and demonstrate new software tools to allow researchers to compute these R-squared measures and decompositions in practice. Using 2 empirical examples, we show how researchers can answer distinct substantive questions with each measure and can gain insights by interpreting the set of measures in juxtaposition to each other.

Keywords: mixture modeling, R-squared, multilevel modeling, regression mixtures

Supplemental materials: http://dx.doi.org/10.1037/met0000139.supp

Psychologists commonly apply regression mixture models in single-level (i.e., unclustered) data analysis contexts (e.g., Ding, 2006; Dyer, Pleck, & McBride, 2012; Fagan, Van Horn, Hawkins, & Jaki, 2013; George et al., 2013; Kliegel & Zimprich, 2005; Montgomery, Vaughn, Thompson, & Howard, 2013) and increasingly in multilevel (i.e., clustered) contexts (e.g., Karakos, 2015; Muthén & Asparouhov, 2009; Rights & Sterba, 2016; Van Horn et al., 2016; Vermunt, 2010; Vermunt & Magidson, 2005). Whereas traditional single-level regression analysis assumes that a set of regression coefficients characterize a homogeneous population, single-level regression mixtures allow these coefficients (intercepts and slopes) to vary by unobserved group, or latent class (DeSarbo & Cron, 1988; Wedel & DeSarbo, 1994). Relatedly, whereas traditional multilevel regression analysis assumes that a set of regression coefficients are continuously distributed across clusters, multilevel regression mixtures can, for instance, allow these coefficients to vary discretely across cluster-level (Level-2) latent classes (e.g., Nagin, 2005; Vermunt & Van Dijk, 2001) or across combinations of both observation-level (Level-1) and cluster-level latent classes (e.g., Muthén & Asparouhov, 2009; Rights & Sterba, 2016;

This article was published Online First March 16, 2017.

Jason D. Rights and Sonya K. Sterba, Department of Psychology and Human Development, Vanderbilt University.

We thank Kristopher J. Preacher for helpful comments.

Correspondence concerning this article should be addressed to Jason D. Rights, Quantitative Methods Program, Department of Psychology and Human Development, Vanderbilt University, Peabody 552, 230 Appleton Place, Nashville, TN 37203. E-mail: Jason.D.Rights@Vanderbilt.edu

Vermunt & Magidson, 2005). These latent classes are often thought to represent distinct subpopulations of individuals, in which case across-class heterogeneity in regression coefficients would reflect substantively meaningful group differences (McLachlan & Peel, 2000).

When reporting results of *nonmixture* single-level regression analyses and *nonmixture* multilevel regression analyses, researchers commonly provide a total R-squared (R^2) measure (e.g., Cohen, Cohen, West, & Aiken, 2003; Draper & Smith, 1998; LaHuis, Hartman, Hakoyama, & Clark, 2014; O'Grady, 1982; Orelien & Edwards, 2008; Vonesh & Chinchilli, 1997). R^2 measures indicate the proportion of total outcome variance that is explained by the model, often computed as the variance of the predicted scores[1] divided by the total variance of the outcome variable. R^2's are useful effect size measures because: (a) they let a researcher know how well a given model's predictions match the observed data; (b) they have an easily interpretable metric (proportion) with meaningful bounds (0 and 1); and (c) they can be compared across studies[2] (e.g., Gelman & Hill, 2007; Xu, 2003). As summarized by Magee (1990),

> There is a natural appeal for a number that can be computed for a fitted model, lies between 0 and 1, is invariant to units of measurement, and becomes larger as the model 'fits better' [in the sense of correspondence between observed and predicted values] (p. 250).

However, despite single-level regression mixtures being used for several decades and multilevel regression mixtures being used for over a decade, there has been no general treatment of R^2 measures for these contexts. Consequently, it is typical for researchers to summarize results of a fitted regression mixture by simply reporting class-specific regression coefficients and their associated *p* values, along with qualitative class-labels (e.g., DeSarbo et al., 2001; Dyer et al., 2012; Fagan et al., 2013; Grewal et al., 2013; Manchia et al., 2010; Nowrouzi et al., 2013; Schmiege et al., 2009). In contrast, there are general and widespread methodological recommendations to look beyond *statistical* significance to also examine *practical* significance in terms of effect size indices, such as R^2 (e.g., American Psychological Association, 2009; Cumming, 2012; Harlow, Muliak, & Steiger, 1997; Kelley & Preacher, 2012; Panter & Sterba, 2011; Wilkinson, 1999).

In this article, we fill this gap by providing an integrative framework of R^2 measures for single-level regression mixture models as well as multilevel regression mixture models (with classes at Level-2, or at both levels). Specifically, we provide 11 different R^2 measures that can be distinguished based on what the researcher chooses to consider as *outcome variance* (total vs. class-specific variance, delineated in the columns of Tables 1 and 2) and based on what sources[3] the researcher chooses to consider as contributing to *predicted variance* (three options, delineated in the rows of Tables 1 and 2). In this article, the R^2 measures listed in Tables 1 and 2 will be given detailed definitions and will be related both analytically and through graphical illustrations. We will contrast the respective substantive interpretations of these R^2 measures, explain what research question is answered by each measure, and explain what insights can be gained by interpreting the entire suite of R^2 measures for a given model, in juxtaposition to each other, rather than simply reporting one. Furthermore, we show how each R^2 measure can be analytically decomposed into distinct, substantively meaningful sources of explained variance.

Table 1
Measures for Single-Level Regression Mixtures or Multilevel Regression Mixtures With Classes at Level-2

	Outcome variance	
Source of explained variance	Total	Class specific
• By predictors via marginal slopes • By predictors via class variation in slopes • By means via class variation	$R^{2(fvm)}$	
• By predictors via marginal slopes • By predictors via class variation in slopes	$R^{2(fv)}$	
• By predictors via marginal slopes	$R^{2(f)}$	$R_k^{2(f)}$

Though analogues to a few of these 11 measures have been incorporated into existing commercial software (Latent GOLD, Vermunt & Magidson, 2016; and M*plus*, Muthén & Muthén, 1998–2016, as described later in our Software Implementation section), the majority of these measures are newly proposed here, and the entire suite is not available in existing software. Moreover, R^2's for regression mixtures in general (whether with classes at a single level, at Level-2 only, or at both levels) have not been systematically described nor interrelated in the literature. In this article, we provide a matrix-based approach for calculating the suite of R^2 measures; formulas are provided in the manuscript text whereas derivations are deferred to Appendices. We implement this approach in an R function which allows researchers to quickly obtain a set of R^2 measures for their fitted model, along with associated graphics and analytic decompositions of explained variance. In doing so, we hope to encourage researchers not only to utilize these measures, but also to consult the framework provided and demonstrated here to understand precisely what each R^2 is measuring.

The remainder of the article proceeds as follows. We first review data models for a single-level regression mixture and a multilevel regression mixture with latent classes only at Level-2. Next, we describe the four different R^2 measures from Table 1 for these models (the computations are the same for both). We then show how these R^2's are related using simulated graphical demonstrations and explain how each can be analytically decomposed into distinct, meaningful sources of explained variance. We next review the data model for a more general multilevel regression mixture with latent classes at both Level-1 and Level-2. We then describe the seven R^2 measures from Table 2 for this model, and also discuss and graphically illustrate how these can be related to one another and analytically decomposed. We

[1] Note that, in multilevel contexts, there are different ways of defining the variance of the predicted scores (e.g., conditional on random effects or not; see Vonesh & Chinchilli, 1997; Vonesh, Chinchilli, & Pu, 1996).

[2] See Nakagawa and Schielzeth (2013) for a more in-depth discussion of the latter point.

[3] In nonmixture modeling contexts, researchers have similarly already expressed interest in substantively determining which sources they want to contribute to explained variance in R^2's (e.g., Nakagawa & Schielzeth, 2013; Vonesh & Chinchilli, 1997). For instance, total R^2's for *nonmixture* multilevel regression are increasingly distinguished by whether they consider outcome variance attributable to random effect variance to be explained variance, termed conditional R^2's, versus whether they consider it to be unexplained variance, termed marginal R^2's (e.g., Edwards, Muller, Wolfinger, Qaqish, & Schabenberger, 2008; Nakagawa & Schielzeth, 2013; Orelien & Edwards, 2008; Vonesh et al., 1996; Wang & Schaalje, 2009); these distinctions have conceptual parallels to distinctions among our total R^2's.

Table 2
Measures for Multilevel Regression Mixtures With Classes at Both Level-1 and Level-2

Source of explained variance	Outcome variance: Total	L2 class specific	L1L2 class specific
• By predictors via marginal slopes • By predictors via class variation in slopes • By means via class variation	$R_T^{2(fvm)}$	$R_h^{2(fvm)}$	
• By predictors via marginal slopes • By predictors via class variation in slopes	$R_T^{2(fv)}$	$R_h^{2(fv)}$	
• By predictors via marginal slopes	$R_T^{2(f)}$	$R_h^{2(f)}$	$R_{kh}^{2(f)}$

Note. L1 = Level-1; L2 = Level-2.

present two empirical examples wherein we compute all relevant R^2 measures and discuss their substantive interpretations. Tables 1 and 2 can be used as a taxonomy and a reference for the R^2's that are more thoroughly defined and explained throughout the article. We conclude by describing software implementation, limitations, and future directions.

Before continuing, a few clarifications about the scope of our article should be noted. First, we are proposing that our R^2 measures be used as an assessment tool after researchers have already defined a set of p predictors of substantive interest and after they have already selected the number of latent classes. In regression mixture applications in psychology, the p predictors are typically chosen based on substantive theory (Ding, 2006; Dyer et al., 2012; Fagan et al., 2013; Karakos, 2015; Montgomery et al., 2013; Muthén & Asparouhov, 2009; Schmiege et al., 2009; Van Horn et al., 2009, 2016; Vermunt, 2010; Vermunt & Magidson, 2005; Wong & Maffini, 2011; Wong, Owen, & Shea, 2012). In such applications, the number of classes is chosen by fitting the regression mixture model with p predictors using alternative numbers of classes and selecting the best-fitting model using, for instance, information criteria. In contrast, when there is no theoretical basis for choosing a set of p predictors (e.g., exploratory data mining) there are data-driven algorithms under development in other disciplines that perform automated iterative searches for selecting predictor variables simultaneously with selecting the number of classes in regression mixtures (e.g., Khalili & Chen, 2012; Raftery & Dean, 2006). In this latter context where such algorithms are employed, our R^2 measures can still be used after a final model is selected in order to assess the overall proportion of variance explained.

As a second clarification, our R^2 measures provide meaningful information about a regression mixture model even when prediction is not the primary objective of the analysis. Though mixture models are increasingly being used for individual prediction (e.g., Cole & Bauer, 2016; Cudeck & Henly, 2003; de Kort, Dolan, Lubke, & Molenaar, in press; Sterba & Bauer, 2014) it is still more common for analysts' primary objective to involve description and/or explanation of theorized data-generating processes. In the latter context, our R^2's serve as relatable, quantitative effect size measures that aid substantive interpretation, which has heretofore usually been qualitative and informal in mixture contexts. It is also important to underscore that—just as in nonmixture regression contexts—a high R^2 does not imply that a regression mixture model is necessarily useful or accurately reflects reality; conversely, a low R^2 does not imply that a regression mixture model is useless or incorrect (for further discussion, see King, 1986 or Cohen, Cohen, West, & Aiken, 2003). Hence, we propose that our measures supplement, but not replace, existing procedures for model selection and overall model evaluation for mixtures.

As a third clarification, R^2 measures provided in this article are for the objective of assessing the correspondence between predicted scores on an outcome variable and observed scores on an outcome variable. Other R^2's, called entropy-based R^2's, have been developed in mixture contexts for the different objective of quantifying how well class memberships are predicted from observed responses (e.g., Lukočienė, Varriale, & Vermunt, 2010; Wedel & Kamakura, 1998), which is not our focus here.

As a fourth clarification, in this article all predictors are exogenous, which is definitional of regression mixtures more generally (see Sterba, 2014 for a detailed discussion of this point) and is the software default employed in empirical applications. Departing from a mixture regression approach by using endogenous predictors whose distributions depend on class is possible (e.g., Ingrassia, Minotti, & Vittadini, 2012; Lamont, Vermunt, & Van Horn, 2016) but implies a different generating process (Sterba, 2014). Our R^2 measures could be extended to apply to the latter situation but this is not addressed in the scope of the present article.

As a final clarification, we do not address R^2 measures for hybrid multilevel mixtures that combine latent classes with random effects that are continuously distributed across clusters (see, e.g., Muthén & Asparouhov, 2009). In other words, all models considered in this article have residuals only at Level-1. This is a popular approach that can greatly save in computational time (e.g., Vermunt, 2004, 2008, 2010) and allows cluster-level dependencies to be accommodated by discrete class-variation in regression coefficients (for details see Nagin, 2005; Rights & Sterba, 2016; Vermunt, 2010). R^2 measures for hybrid multilevel mixtures that include continuously distributed random effects, however, are an extension for future research. Such hybrid multilevel mixtures allow inclusion of continuous random effects in conjunction with classes at one level (i.e., at Level-1 or Level-2) or in conjunction with classes at multiple levels. In such models, random effect distributions can be specific to a given class or class-combination, or can be constrained equal across classes.

Data Model for a Single-Level Regression Mixture

To begin, we consider a regression mixture model for single-level (unclustered) data (e.g., Ding, 2006; Dyer et al., 2012; Wedel & DeSarbo, 1994). With i denoting observation ($i = 1 \ldots N$) and c_i denoting latent class membership for observation i (with classes ranging from $k = 1$ to K), we can model a univariate outcome y_i conditional on class k membership, as follows:

$$\begin{aligned} y_{i|c_i=k} &= \mathbf{x}_i'\boldsymbol{\gamma}^k + \varepsilon_i \\ \varepsilon_i &\sim N(0, \theta^k) \\ p(c_i = k) &= \pi^k = \frac{\exp(\omega^k)}{\sum_{k=1}^{K}\exp(\omega^k)} \end{aligned} \tag{1}$$

Here, $\mathbf{x}_i'$ denotes a row vector consisting of 1 and all p predictors for observation i (i.e., $x_{1i}, \ldots, x_{pi}$). This vector is multiplied by the column vector $\boldsymbol{\gamma}^k$ containing all regression coefficients specific to class k (i.e., $\gamma_0^k, \gamma_1^k, \ldots, \gamma_p^k$), with γ_0^k denoting the class-specific intercept and each of $\gamma_1^k, \ldots, \gamma_p^k$ denoting class-specific slopes. The residual, ε_i, is normally distributed with a class-specific variance, θ^k. Note that in the current article we focus on within-class models which are linear with normally distributed within-class residuals; such normal mixtures are the most common kind of empirical mixture applied (e.g., Sterba, Baldasaro, & Bauer, 2012). The probability of membership in class k, $p(c_i = k)$, is denoted π^k and is modeled with class-specific multinomial intercepts, $\omega^1 \ldots \omega^K$, with ω^K constrained to 0 for identification. This model in Equation (1) can be useful in modeling heterogeneity of regression coefficients across latent classes in single-level contexts (e.g., differing effects of parental sensitivity on child social competence across latent classes; Van Horn et al., 2015).

Data Model for a Multilevel Regression Mixture With Classes Only at Level-2

Next, we consider a regression mixture for multilevel (clustered) data which involves classes only at Level-2 (cluster-level; e.g., Nagin, 2005; Vermunt & van Dijk, 2001). In this model, the across-cluster distributions of slopes and intercepts are not assumed to be normal (as in traditional multilevel models); rather, they are distributed discretely across Level-2 classes.

Let i denote an observation within cluster j (where $i = 1 \ldots N_j$ and $j = 1 \ldots J$). Let d_j denote the Level-2 (i.e., cluster-level) latent class membership for cluster j, with classes ranging from $h = 1$ to H. We can then model a univariate y_{ij} conditional on Level-2 latent class h membership as follows:

$$\begin{aligned} y_{ij|d_j=h} &= \mathbf{x}_{ij}'\boldsymbol{\gamma}^h + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim N(0, \theta^h) \\ p(d_j = h) &= \pi^h = \frac{\exp(\varpi^h)}{\sum_{h=1}^{H} \exp(\varpi^h)} \end{aligned} \tag{2}$$

Note that this model expression differs from Equation (1) in only two ways: there is now a j subscript denoting cluster, and Level-2 classes are now denoted by h. Here, $\mathbf{x}_{ij}'$ denotes a row vector consisting of 1 and all predictors for observation i within cluster j. This vector is multiplied by the column vector $\boldsymbol{\gamma}^h$ containing all regression coefficients specific to Level-2 class h. The residual, ε_{ij}, is normally distributed with a Level-2-class-specific variance, θ^h. The probability of a cluster belonging to class h, $p(d_j = h)$, is denoted π^h and is modeled by a Level-2 class specific multinomial intercept, ϖ^h. For identification, $\varpi^H = 0$.

Equation (2) can be useful in modeling heterogeneity of regression coefficients across Level-2 latent classes (e.g., school-level latent classes, in an analysis where students are nested within schools). This model is also widely applied by psychologists in longitudinal contexts where the Level-1 unit is time and the Level-2 unit is person (e.g., Broidy et al., 2003; Mulvey et al., 2010). In this context, psychologists are interested in variation in growth trajectory coefficients across discrete classes of persons.

R-Squared Measures for Regression Mixtures With Classes at Only One Level

Having defined two kinds of regression mixtures with classes at only one level (Equations 1 and 2), we now proceed to motivate and then delineate R^2 measures from Table 1 that apply to *either* of these model specifications. That is, even though classes in Equation (1) have very different substantive meanings than classes in Equation (2), the same R^2 measures from this section can be applied in both modeling contexts. For simplicity, in this section we will express and discuss these measures with reference to the single-level regression mixture (i.e., Equation 1), but the formulas can be modified for the context of only Level-2 classes by changing all k subscripts/superscripts to h subscripts/superscripts.

Total R-Squared Measures

First we consider: "Why would researchers be interested in a total R^2 for a regression mixture?" A total R^2 is useful because it provides a measure of the *overall* practical significance of a model. When *directly interpreting* classes—wherein latent classes are thought to represent distinct subpopulations (Titterington, Smith, & Makov, 1985)—it may seem intuitive that only class-specific R^2's would be of interest; however, it is still important to compute a total R^2 to obtain an omnibus understanding of how much variance as a whole (i.e., across all latent classes) can be explained by the model. Furthermore, when directly interpreting classes it can also be useful to compare a total R^2 with class-specific R^2's (discussed later). For example, even if a model explains a large proportion of variance for a given class, this does not imply that the model explains a large proportion of the total variance. A total R^2 would naturally also be useful for researchers who are *indirectly interpreting* latent classes. With an indirect interpretation, classes are thought not to represent literal subpopulations; rather, taken together, they are used to approximate some underlying continuous distribution of effects (Titterington et al., 1985). In this case, each class in isolation is not substantively meaningful and thus a class-specific R^2 may not be useful. Lastly, total R^2's can be useful for comparison across studies with similar predictor sets, even when the number of classes differs across studies. Comparing class-specific R^2's would be more difficult in that this would require the same class structure across studies.

Next we consider: "What are potential sources of explained variance for a total R^2 for a regression mixture?" Substantively, researchers fitting mixtures often informally characterize results in terms of one or more of three potential sources of explained variance, but they describe these sources only heuristically and qualitatively, not quantitatively (e.g., Halliday-Boykins, Henggeler, Rowland, & Delucia, 2004; Morin & Marsh, 2015; Sher, Jackson, & Steinley, 2011; Sterba & Bauer, 2014). These three potential sources of explained variance (given in Table 1) are: variance explained (a) by predictors via their marginal (i.e., weighted across-class average) slopes; (b) by predictors via class variation in slopes; and (c) by outcome means via class variation. To better understand these three sources of explained variance, consider a regression mixture analysis in which our outcome is children's social competence (e.g., Van Horn et al., 2015) and, for ease of graphical representation, suppose we have only one covariate, parental sensitivity. Now consider three hypothetical sets of results shown in Figure 1 Panels A–C. In each of these, explained variance stems primarily from only one out of the three potential sources. This is, of course, an overly simplistic situation, but nevertheless is useful in conceptualizing and visualizing these three potential sources of explained variance. In the first hypothetical set of results, shown in Figure 1 Panel A, classes have virtually identical intercepts and slopes; here, outcome variance can be explained primarily by parental sensitivity via the *marginal slope*. Substantively, one might infer there to be an overall positive effect of parental sensitivity, with little heterogeneity across class. In the second hypothetical set of results, in Figure 1 Panel B, the marginal slope is near-zero and across-class outcome mean variation is near-zero; here, outcome variance can be explained primarily by parental sensitivity via *across-class slope variation*. One might now infer that the effect of parental sensitivity depends heavily on class membership. In contrast to both of these, in Figure 1 Panel C, the marginal slope is near-zero and across-class slope variation is near-zero; here, explained variance can be attributable primarily to *across-class mean variation*[4] in social competence. In this situation, one might conclude there is no effect of parental sensitivity for any of the classes, but that classes differ on the degree of child social competence. Comparing these three conditions, each situation corresponds with a unique interpretation, reflecting the fact that each source represents something substantively unique. Lastly, consider a more nuanced and realistic situation, depicted in Figure 1 Panel D. Here, explained variance can be attributable in part to all three of the aforementioned components.

Now having understood these potential sources of explained variance, we can use this information to choose a total R^2 measure for a given study. In particular, an appropriate total R^2 measure would only count as explained variance those sources which the researcher deems to be substantively meaningful. In realistic situations, all three sources will likely be nonzero (as in Figure 1 Panel D) but, importantly, a researcher can choose whether to count all or some of them as explained variance. The following total R^2 measures will be distinguished by whether three, two, or one of these sources are counted toward explained variance, as shown in Table 1.

Total $R^{2(fvm)}$ measure. Suppose a researcher wanted to answer the question: "What proportion of variance is explained by predictors via their marginal slopes, by predictors via across-class slope variation, and by variation in class means of the outcome?" In other words, here all three sources depicted in Figure 1 would be considered to explain variance. To address this research question, we can compute $R^{2(fvm)}$ as:[5]

$$R^{2(fvm)} = \frac{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\prime}\,\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet}}{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\prime}\,\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet} + \theta^{\bullet}} \tag{3}$$

The denominator is an expression for the model-implied total variance of y_i (i.e., computed as a function of model parameter estimates).[6] A derivation for this denominator expression from the data model is given in Appendix A.[7] Each symbol in this denominator expression is defined individually in Table 3. Here, a brief summary of the terms in the denominator expression is provided. The first term in the denominator, $\mathbf{s}'\boldsymbol{\psi}$, reflects outcome variance attributable to $\mathbf{x}_i$ via coefficient variation across class; $2\mathbf{p}'\boldsymbol{\kappa}$ reflects outcome variance attributable to $\mathbf{x}_i$ via coefficient covariation across class; and $\boldsymbol{\gamma}^{\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet}$ represents outcome variance attributable to $\mathbf{x}_i$ via marginal regression coefficients.[8] These marginalized regression coefficients, in $\boldsymbol{\gamma}^{\bullet}$, are averages of the class-specific coefficients, $\boldsymbol{\gamma}^k$, each weighted by class-probability π^k (i.e., $\sum_{k=1}^{K}\pi^k\boldsymbol{\gamma}^k$). The term $\theta^{\bullet}$ reflects marginalized residual variance (a weighted average of the class-specific residual variances). The numerator contains only $\mathbf{s}'\boldsymbol{\psi}$, $2\mathbf{p}'\boldsymbol{\kappa}$, and $\boldsymbol{\gamma}^{\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet}$, leaving the marginalized residual variance as the unexplained portion of variance.

Total $R^{2(fv)}$ measure. Suppose instead that a researcher wanted to answer the question: "What proportion of variance is explained by predictors via their marginal slopes and across-class slope variation?" In other words, here only the first two sources depicted in Figure 1 (Panels A and B) would be considered to explain variance. Recall that the $R^{2(fvm)}$ would consider mean outcome variation across classes to be meaningful, explained variance. The difference between, for instance, a low-mean class and a high-mean class could thus make the $R^{2(fvm)}$ fairly large. To answer the above research question, however, this mean variation needs to be partialed out, thus isolating the influence of predictors via their slopes. To do so, we compute $R^{2(fv)}$ by subtracting from $R^{2(fvm)}$ the proportion of variance explained by class mean separation alone:

$$\begin{aligned} R^{2(fv)} &= \frac{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\prime}\,\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet}}{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\prime}\,\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet} + \theta^{\bullet}} - \frac{\mathrm{var}\big(E\big[\mathbf{x}_i'\boldsymbol{\gamma}^k + \varepsilon_i \,|\, c_i = k\big]\big)}{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\prime}\,\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet} + \theta^{\bullet}} \\ &= \frac{\mathbf{v}'\boldsymbol{\psi} + 2\mathbf{r}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\prime}\,\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet}}{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\prime}\,\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet} + \theta^{\bullet}} \end{aligned} \tag{4}$$

The numerator of the proportion of variance that is subtracted in Equation (4) is equal to the variance of model-implied class means from the fitted model. A detailed derivation of the numerator expression in Equation (4) is given in Appendix B.[9] The denominator of Equation (4) is the model-implied total variance of y_i, unchanged from the calculation for $R^{2(fvm)}$ in Equation (3). In other words, the proportion of variance that is subtracted in Equation (4) is an analytically derived R^2 for an intercept-only regression mixture with class-varying intercepts fixed to the model-implied means of each class in the fitted model. This subtraction results in an equation with several new terms: $\mathbf{v}'$ is a row vector of variances of each element of $\mathbf{x}_i'$ across observations and $\mathbf{r}'$ is a row vector of pairwise covariances of all nonredundant elements of $\mathbf{x}_i'$ across observations.[10] Thus, the $R^{2(fv)}$ expression is similar to $R^{2(fvm)}$, only replacing $\mathbf{s}'$ with $\mathbf{v}'$ and replacing $\mathbf{p}'$ with $\mathbf{r}'$.

[4] If all predictors are centered (grand mean centered for Level-1 or Level-2 predictors, or cluster-mean centered for Level-1 predictors) what we are referring to as across-class mean separation on the outcome can also be interpreted as across-class intercept separation on the outcome.

[5] In the superscript of $R^{2(fvm)}$, "m" reflects across-class mean variation, "v" reflects across-class slope variation, and "f" reflects marginal slopes (which is somewhat similar conceptually to a fixed component of the slopes).

[6] In other modeling contexts there is a precedent for computing R^2 measures using a model-implied total outcome variance in the denominator (e.g., Lahuis et al., 2014; Snijders & Bosker, 2012).

[7] The Appendix A derivation is provided in the context of a more general regression mixture model that allows classes at both Level-1 and Level-2, presented later in Equation (7). The present model is a constrained special case.

[8] For details about how these definitions apply to the first element in $\mathbf{x}_i$, see examples in Table 3.

[9] The Appendix B derivation is provided in the context of a more general regression mixture model (allowing classes at both Level-1 and Level-2), presented later in Equation (7). The present model is a constrained special case.

[10] For instance, if $\mathbf{x}_i' = [1\ x_{1i}\ x_{2i}]$ then $\mathbf{v}' = [0\ \mathrm{var}(x_{1i})\ \mathrm{var}(x_{2i})]$ and $\mathbf{r}' = [0\ 0\ \mathrm{cov}(x_{1i}, x_{2i})]$.

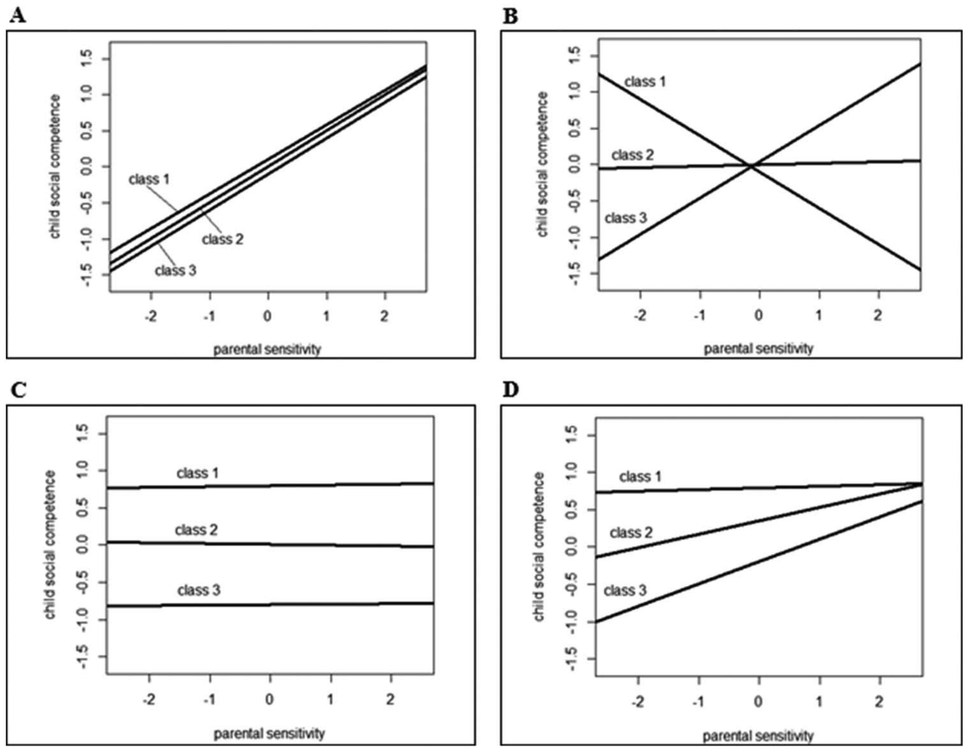

Figure 1. Hypothetical regression mixture wherein parental sensitivity predicts child social competence: Illustrating potential sources of explained variance. Each panel depicts a hypothetical set of results from a three-class regression mixture model with equal class proportions. Outcome variance is explained primarily by the following sources: by parental sensitivity via the marginal slope (Panel A), by parental sensitivity via across-class slope variance (Panel B), by social competence means via across-class variation (Panel C), and by the combination thereof (Panel D).

Total $R^{2(f)}$ measure. Now suppose a researcher wanted to answer the question: "What proportion of variance is explained by predictors via their marginal slopes?" In other words, here only the first source depicted in Figure 1 (Panel A) would be considered to explain variance. To address this research question, we compute $R^{2(f)}$. It could be the case, for instance, that on average (i.e., pooling across-class), the predictors have a large impact, leading to a large $R^{2(f)}$. Alternatively, a low $R^{2(f)}$ may indicate that meaningful relationships between the outcome and the predictors are apparent only when considering heterogeneity in the regression coefficients. We represent $R^{2(f)}$ as:

$$R^{2(f)} = \frac{\boldsymbol{\gamma}^{\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet}}{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet} + \theta^{\bullet}} \quad (5)$$

The numerator of Equation (5) indicates that only the variance explained via the marginalized regression coefficients counts toward predicted variance. This $R^{2(f)}$ is akin to the proportion of variance that would be explained had a traditional (nonmixture) linear regression model been fit.

Class-Specific R-Squared Measure

In addition to these total R^2's, we next consider: "Why would researchers be interested in a class-specific R^2 for a regression mixture?" A class-specific R^2 could be particularly useful when directly interpreting classes. With a class-specific R^2, one can determine the proportion of variance explained for each subpopulation, and differences between these class-specific R^2 measures may reflect substantively meaningful group differences. Note that, unlike for a total R^2, there is no option to choose among alternate sources of explained variance for a class-specific R^2 (given that there is no across-class heterogeneity in either means or slopes within a single class). Instead, for a class-specific R^2, there is only one source of explained variance, as described below.

Class-specific $R_k^{2(f)}$ measure. Suppose a researcher wanted to answer the question: "What proportion of variance in class k is explained by the predictors?" To address this research question, we compute $R_k^{2(f)}$. Differences in $R_k^{2(f)}$'s across $k = 1 \ldots K$ could provide evidence to support the idea that the predictors better explain the outcome in some classes compared to others. $R_k^{2(f)}$ is given as:

$$R_k^{2(f)} = \frac{\boldsymbol{\gamma}^{k\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{k}}{\boldsymbol{\gamma}^{k\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{k} + \theta^{k}} \quad (6)$$

The denominator of Equation (6) is an expression for the model-implied total variance of y_i within class k. Because the denominator expression is specific to class k, there are no terms reflecting across-class variation (in means or in slopes). As mentioned

Table 3
Definitions of Terms in Expression for Model-Implied Total Outcome Variance

$\mathbf{s}'$	Definition: Row vector of means of squares of each element of $\mathbf{x}_i'$. Example: if $\mathbf{x}_i' = [1 \quad x_{1i} \quad x_{2i}]$ then $\mathbf{s}' = [1 \quad \overline{x_{1i}^2} \quad \overline{x_{2i}^2}]$
$\boldsymbol{\psi}$	Definition: Column vector of weighted variances (across class) of regression coefficients. Example: if $\boldsymbol{\gamma}^k = \begin{bmatrix}\gamma_0^k\\ \gamma_1^k\\ \gamma_2^k\end{bmatrix}$ then $\boldsymbol{\psi} = \begin{bmatrix}\text{var}_k(\gamma_0^k)\\ \text{var}_k(\gamma_1^k)\\ \text{var}_k(\gamma_2^k)\end{bmatrix} = \begin{bmatrix}\sum_{k=1}^K \pi^k(\gamma_0^k)^2 - \left(\sum_{k=1}^K \pi^k\gamma_0^k\right)^2\\ \sum_{k=1}^K \pi^k(\gamma_1^k)^2 - \left(\sum_{k=1}^K \pi^k\gamma_1^k\right)^2\\ \sum_{k=1}^K \pi^k(\gamma_2^k)^2 - \left(\sum_{k=1}^K \pi^k\gamma_2^k\right)^2\end{bmatrix}$
$\mathbf{p}'$	Definition: Row vector of means of pairwise products of all nonredundant elements of $\mathbf{x}_i'$. Example: if $\mathbf{x}_i' = [1 \quad x_{1i} \quad x_{2i}]$ then $\mathbf{p}' = [\overline{x_{1i}} \quad \overline{x_{2i}} \quad \overline{x_{1i}x_{2i}}]$
$\boldsymbol{\kappa}$	Definition: Column vector of weighted covariances (across class) of regression coefficients corresponding to the elements in $\mathbf{p}'$. Example: if $\boldsymbol{\gamma}^k = \begin{bmatrix}\gamma_0^k\\ \gamma_1^k\\ \gamma_2^k\end{bmatrix}$ then $\boldsymbol{\kappa} = \begin{bmatrix}\text{cov}_k(\gamma_0^k, \gamma_1^k)\\ \text{cov}_k(\gamma_0^k, \gamma_2^k)\\ \text{cov}_k(\gamma_1^k, \gamma_2^k)\end{bmatrix} = \begin{bmatrix}\sum_{k=1}^K \pi^k\gamma_0^k\gamma_1^k - \sum_{k=1}^K \pi^k\gamma_0^k\sum_{k=1}^K \pi^k\gamma_1^k\\ \sum_{k=1}^K \pi^k\gamma_0^k\gamma_2^k - \sum_{k=1}^K \pi^k\gamma_0^k\sum_{k=1}^K \pi^k\gamma_2^k\\ \sum_{k=1}^K \pi^k\gamma_1^k\gamma_2^k - \sum_{k=1}^K \pi^k\gamma_1^k\sum_{k=1}^K \pi^k\gamma_2^k\end{bmatrix}$
$\theta^{\bullet}$	Definition: Marginalized residual variance (weighted average of class-specific residual variances). Example: $\theta^{\bullet} = \sum_{k=1}^K \pi^k\theta^k$
$\boldsymbol{\gamma}^{\bullet}$	Definition: Marginalized regression coefficients (weighted average of class-specific coefficients). Example: $\boldsymbol{\gamma}^{\bullet} = \sum_{k=1}^K \pi^k\boldsymbol{\gamma}^k$
$\boldsymbol{\Phi}$	Definition: Covariance matrix of all elements of $\mathbf{x}_i'$ across observations i. Example: if $\mathbf{x}_i' = [1 \quad x_{1i} \quad x_{2i}]$ then $\boldsymbol{\Phi} = \begin{bmatrix}0 & & \\ 0 & \text{var}_i(x_{1i}) & \\ 0 & \text{cov}_i(x_{1i}, x_{2i}) & \text{var}_i(x_{2i})\end{bmatrix}$

Note. Examples are written specifically for a single-level mixture. However, definitions also apply to a multilevel mixture with classes only at Level-2 if $\mathbf{x}_{ij}'$ is substituted for $\mathbf{x}_i'$. Definitions apply to a multilevel mixture with classes at both levels if $\mathbf{x}_{ij}'$, $\theta^{\bullet\bullet}$, and $\boldsymbol{\gamma}^{\bullet\bullet}$ are substituted for $\mathbf{x}_i'$, $\theta^{\bullet}$, and $\boldsymbol{\gamma}^{\bullet}$.

earlier, $\boldsymbol{\gamma}^k$ is a vector of regression coefficients specific to class k, and θ^k is the residual variance in class k. The $\boldsymbol{\Phi}$ matrix retains the same definition from Table 3 because, as in nonmixture regression models, predictors are by-definition exogenous in mixture regression models (see Sterba, 2014 for a detailed review); hence their distributions are not class-specific. Finally, the numerator of Equation (6) is the variance in class k explained by predictors, thus excluding the unexplained, residual variance in class k, θ^k.

Simulated Demonstrations of Relationships Among R-Squareds for Regression Mixtures With Classes at Only One Level

With these R^2's now delineated, we graphically demonstrate some key points regarding how parameter values differentially influence each measure. For simplicity of visualization, these illustrations involves a single-level regression mixture with one centered predictor (x_i) with unit variance, $K = 2$ latent classes, and unit class-specific residual variances. In Figures 2, 3, and 4, each panel shows two class-specific regression lines, with class labels beside each line.

Manipulating the three potential sources of explained variance. We first illustrate how increasing the *marginal slope* influences the suite of R^2 measures. Going from Figure 2 Panel A to Panel B, the marginal slope is increased (by increasing each class-specific slope by the same amount, holding all else constant). This increased marginal slope leads to an increase for all of the total R^2 measures: $R^{2(fvm)}$, $R^{2(fv)}$, and $R^{2(f)}$. This reflects the property that variance explained by the predictor via the marginal slope is considered *explained variance* in each of the three measures.

We next illustrate how increasing *across-class slope variation* influences the suite of R^2's. Going from Figure 2 Panel A to Panel C, the across-class slope variance increases, with all else held constant. Because across-class slope variation is considered *explained variance* in both $R^{2(fvm)}$ and $R^{2(fv)}$, these values increase from Panel A to C. However, because across-class slope variation is considered *residual variance* in $R^{2(f)}$, this value decreases.

Lastly, we illustrate how increasing *across-class mean outcome variation* influences the suite of R^2's. Going from Figure 2 Panels A to D, the across-class mean separation increases while all else is held constant. Because class mean variation is considered *explained variance* in $R^{2(fvm)}$ but *residual variance* in $R^{2(fv)}$ and $R^{2(f)}$, this increased class mean variation leads to an increased $R^{2(fvm)}$ but a decreased $R^{2(fv)}$ and $R^{2(f)}$.

Relating class-specific R-squareds to total R-squareds. Next, we consider the relationship between the total R^2's and the class-specific R^2's. It is important to understand that if a researcher knows only the class-specific R^2 values, $R^2_{k=1}$ and $R^2_{k=2}$, there is not a way to know what the total R^2 values ($R^{2(fvm)}$, $R^{2(fv)}$, and $R^{2(f)}$) are; hence, it is useful to consider the

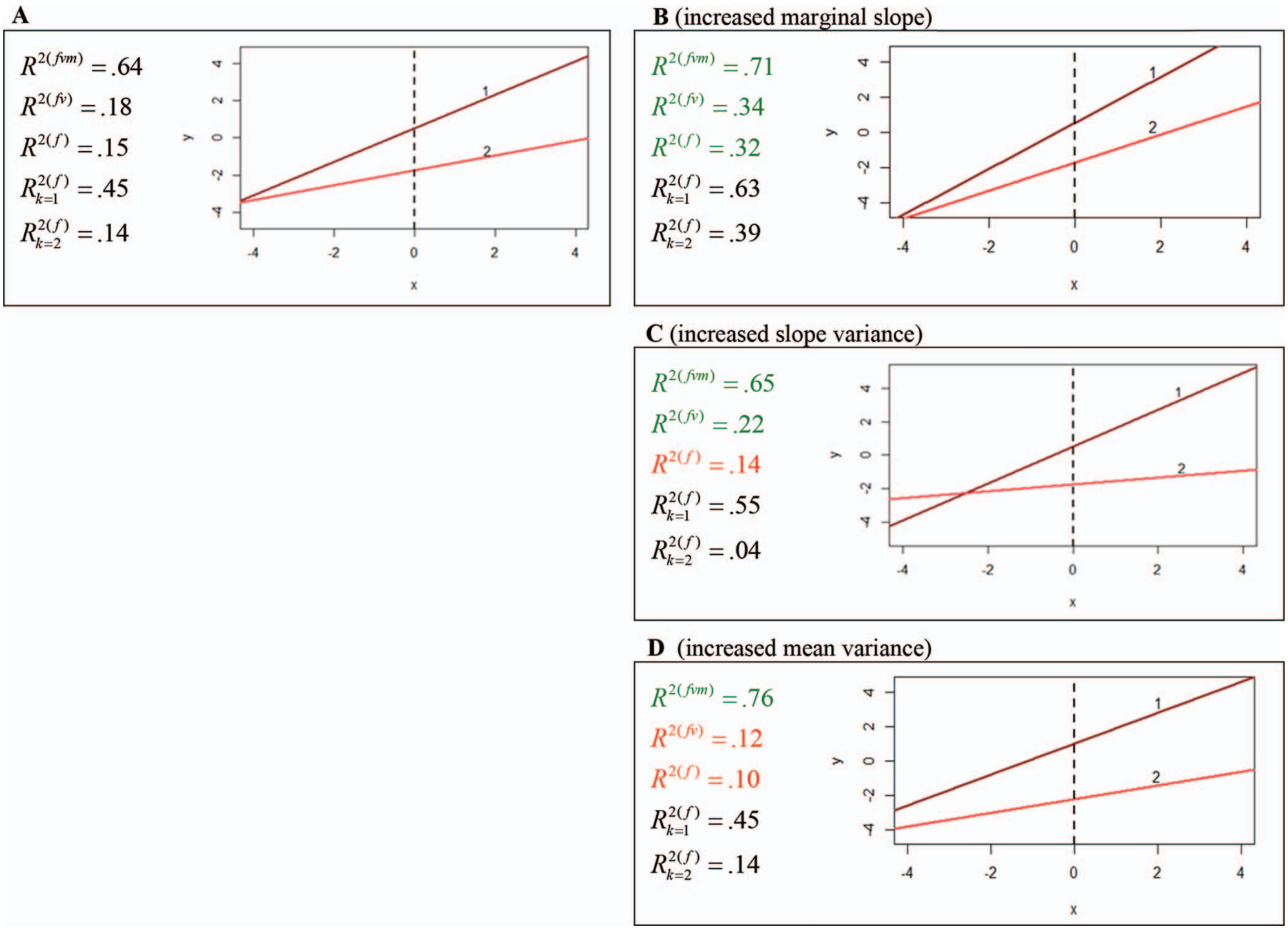

Figure 2. Simulated demonstration: Consequences for total R^2's of manipulating the three sources of explained variance in a regression mixture with classes at only one level. Compare Panel A with each of Panels B–D. Each line denotes a class-specific regression line for a $K = 2$ single-level regression mixture. Comparing Panel A with Panels B, C, and D demonstrates, respectively, the effects of increased marginal slope, increased across-class slope variance, and increased across-class mean outcome variance on the total $R^{2(fvm)}$, $R^{2(fv)}$, $R^{2(f)}$. (green = R^2 increased, red = R^2 decreased, compared with Panel A). Note that for simplicity class proportions are held equal (here and in Figures 3, 4, 5, and 7), but the same patterns would hold regardless. See the online article for the color version of this figure.

set of R^2's together. We provide demonstrations of several points related to this concept.

A first point is that all total R^2's can differ even when the class-specific R^2's are the same; this point is illustrated by comparing Figure 2 Panels A and D. A second point is that an increased class-specific R^2 need not imply that every kind of total R^2 is also larger. This point is illustrated in Figure 3, wherein increasing $R^2_{k=1}$ (by increasing the absolute value of the slope in Class 1, holding all else equal), can either increase $R^{2(f)}$ (comparing Figure 3 Panel A vs. B) or decrease $R^{2(f)}$ (comparing Figure 3 Panel A vs. C). The $R^2_{k=2}$ remains constant, however, because manipulating the model-implied variance of predicted scores or model-implied total variance within $k = 1$ does not affect those quantities for $k = 2$.

Different patterns of class-specific parameters can yield identical total R-squareds. A final, important point is illustrated in Figure 4: that substantively different patterns of class-specific parameters can yield the same total R^2. In Figure 4, Panels A and B yield the same $R^{2(fvm)}$, whereas Panels A and C yield the same $R^{2(f)}$. This is despite the fact that the conditions in Figure 4 Panels B or C correspond with very different substantive interpretations of class-specific parameters than Figure 4 Panel A. In Panel A, for instance, an interpretation of class specific parameters might be that "there is heterogeneity across classes in terms of the slope of y_i on x_i, but no heterogeneity in means of y_i across class." In contrast, Panel B would yield an interpretation of "there is *no* heterogeneity in the effect of x_i, but there is heterogeneity in the means of y_i." Looking at only a single total R^2, it would not be possible to distinguish these interpretations. For this reason, we encourage researchers to consider the entire suite of R^2 measures together. Furthermore, to have a more precise understanding of the sources of variance, a researcher can examine the decompositions of explained variance, described in the next section.

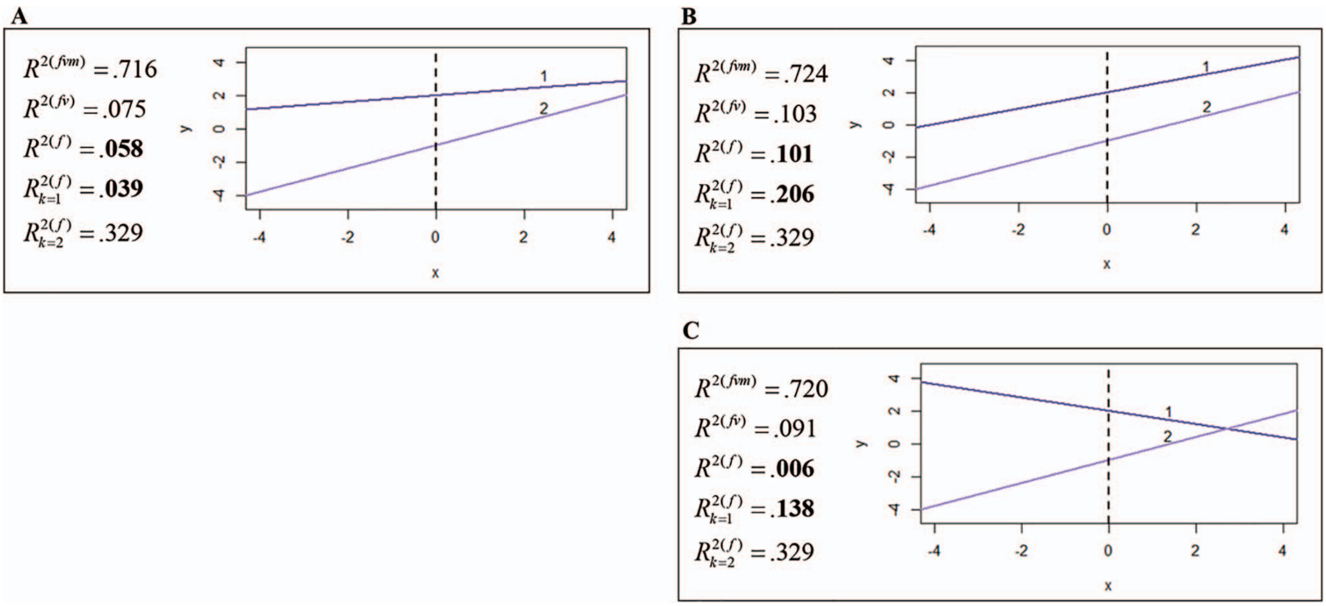

Figure 3. Simulated demonstration: Relating class-specific R^2's to total R^2's in a regression mixture with classes at only one level. Compare Panel A with each of Panels B and C. Each line denotes a class-specific regression line for a $K = 2$ single-level regression mixture. **Bold** R^2 results are compared and discussed in the article. See the online article for the color version of this figure.

Decomposing Explained Variance in R-Squareds for Regression Mixtures With Classes at Only One Level

In this section we more explicitly consider how the three total R^2's are related by showing how they can be analytically decomposed into one, two, or three components of explained variance (all of which were depicted in Figure 1). Researchers can use these relations to inform *what* they wish to interpret as substantively meaningful explained variance, and thus decide which R^2 to implement. Furthermore, each distinct component may have its own useful substantive interpretation.

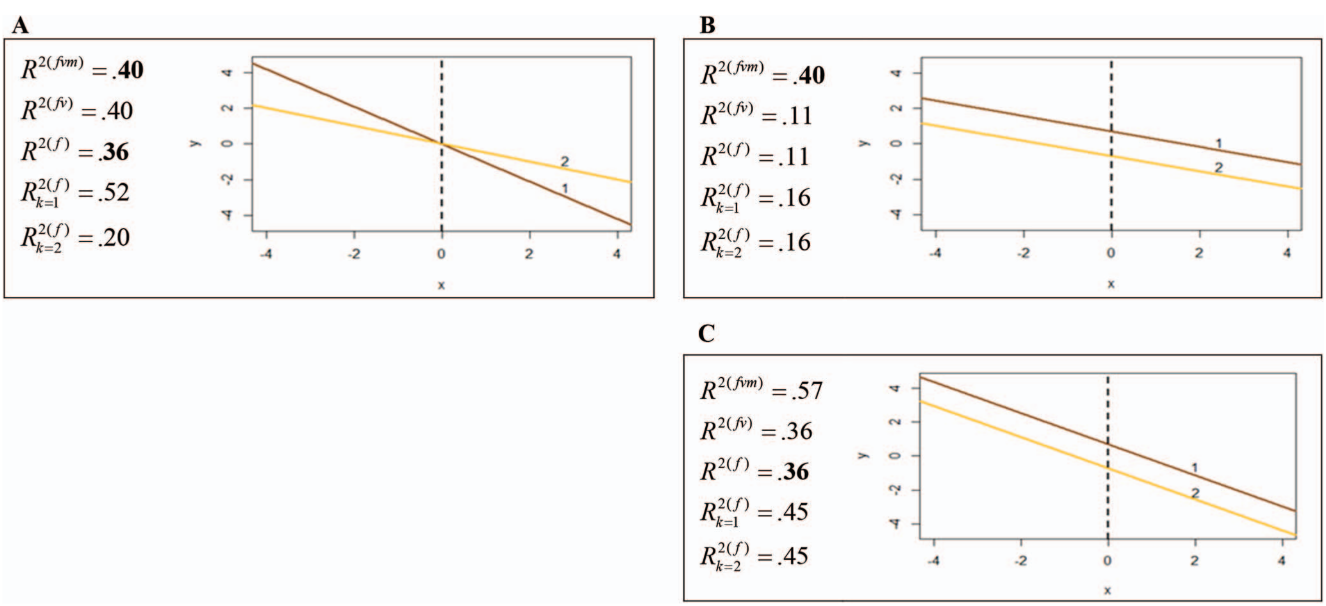

Figure 4. Simulated demonstration: Substantively different patterns of class-specific regression equations can yield the same total R^2 in a regression mixture with classes at only one level. Compare Panel A with each of Panels B and C. Each line denotes a class-specific regression line for a $K = 2$ single-level regression mixture. **Bold** R^2 results are compared and discussed in the article. See the online article for the color version of this figure.

Table 4
Decomposing Variance Explained in Total R^2 for Single-Level Regression Mixtures or Multilevel Regression Mixtures With Classes Only at Level-2

$$\begin{aligned} R^{2(fvm)} &= \begin{Bmatrix}\text{contribution of predictors}\\ \text{via marginal slopes}\end{Bmatrix} + \begin{Bmatrix}\text{contribution of predictors}\\ \text{via class variation in slopes}\end{Bmatrix} + \begin{Bmatrix}\text{contribution of means}\\ \text{via class variation}\end{Bmatrix} \\ &= \{\ \text{Equation(5)}\ \} + \{\ \text{Eq(4)} - \text{Eq(5)}\ \} + \{\ \text{Eq(3)} - \text{Eq(4)}\ \} \\ R^{2(fv)} &= \begin{Bmatrix}\text{contribution of predictors}\\ \text{via marginal slopes}\end{Bmatrix} + \begin{Bmatrix}\text{contribution of predictors}\\ \text{via class variation in slopes}\end{Bmatrix} \\ &= \{\ \text{Equation(5)}\ \} + \{\ \text{Eq(4)} - \text{Eq(5)}\ \} \\ R^{2(f)} &= \begin{Bmatrix}\text{contribution of predictors}\\ \text{via marginal slopes}\end{Bmatrix} \\ &= \{\ \text{Equation(5)}\ \} \end{aligned}$$

First consider $R^{2(fvm)}$. Table 4 shows how it can be decomposed into three distinct components: (a) the contribution of predictors via their marginal slopes; (b) contribution of predictors via across-class variation in slopes,[11] and (c) the contribution of outcome means via across-class variation. In contrast, Table 4 shows that $R^{2(fv)}$ can be decomposed into only the first two of these components—reflecting the fact that $R^{2(fv)}$ involves partialing out the contribution of across-class variation in means. Finally, Table 4 shows that $R^{2(f)}$ consists of only the first component because it involves partialing out both across-class variation in means and across-class variation in slopes. Note that a decomposition is not relevant for the class-specific R^2, R_k^2, because there is no across-class variation to partial out, leaving only marginal slopes to consider (similar to the $R^{2(f)}$).

Table 4 shows how the components of each R^2 can be separately obtained from manipulations of the formulas described earlier. Our R (R Core Team, 2016) software function (described later) computes these quantities and produces bar charts that allow a visual illustration of these decompositions. These bar charts are provided for an empirical example in the next section, in which we demonstrate the practical utility of obtaining such a decomposition.

Empirical Example 1

Here we present an empirical example involving a single-level regression mixture predicting average professor salary for private colleges in the United States (i.e., each datapoint is a college). Our data set consists of $N = 693$ private colleges compiled for the 1995 Data Analysis Exposition of the American Statistical Association.[12] In this regression mixture model, we are substantively interested in exploring the heterogeneity of effects of three predictors that have been related to faculty compensation in prior nonmixture modeling studies (e.g., Fairweather, 1996; Langton & Pfeffer, 1994). These predictors are: the number of students accepted divided by the number of applications received (*admit*), student-to-faculty ratio (*sfratio*), and percentage of faculty with a Ph.D. (*perPhD*). We allowed all slopes and intercepts to vary by latent class and, for parsimony, constrained the residual variances equal.

Using M*plus* to estimate the model and the Bayesian Information Criterion (BIC) to select the number of classes (see, e.g., Nylund, Asparouhov, & Muthén, 2007), we found evidence for $K = 3$ latent classes as best fitting. Parameter estimates and standard errors for the $K = 3$ solution are provided in online supplemental Appendix A and patterns of significant results are summarized here. The first class (25% of colleges) consisted of schools with low mean salaries wherein admit and sfratio *positively* predicted salary. The second class (60% of colleges) consisted of schools with medium mean salary wherein admit and sfratio *negatively* predicted salary and perPhD *positively* predicted salary. The third class (15% of schools) consisted of schools with high mean salaries wherein admit *negatively* predicted salary and perPhD *positively* predicted salary. Comparing the three classes, $k = 1$ might reflect schools with limited endowments (thus lower average salary) which allocate more funds toward professor salary as more money is obtained via increasing admissions and class size. Regional colleges (e.g., Albertus Magnus College) often had highest posterior probabilities of assignment to class $k = 1$. In contrast, national liberal arts colleges (e.g., Swarthmore College) often had highest posterior probabilities of assignment to class $k = 2$. Class $k = 2$ might reflect schools that strongly value teaching quality and compensate commensurate with teaching. Schools in this class that have better teaching atmospheres (via smaller class sizes, more educated professors, and a more selective student body) compensate their professors accordingly. R1 research universities (e.g., California Institute of Technology) had highest posterior probabilities of assignment to the final class, $k = 3$. Class $k = 3$ schools may similarly value quality teaching, but may have less focus on teaching atmosphere and compensate professors commensurate with say, research output. Though not hypothesized a priori, these classes might be conceptually labeled: *regional*, *liberal arts*, and *research institution* classes.

Researchers often summarize the "distinctness" of classes using qualitative descriptions of heterogeneity, such as "there is some across-class heterogeneity in the effects of the predictors," and "there is some across-class heterogeneity in means of salary."

[11] Note that, in the context of uncentered predictors, across-class slope variation can influence model-implied across-class mean variation. Hence the need to clarify that this second component, class-specific slope variation, accounts for explained variance due to variation in class-specific slopes *above and beyond* the effect that this slope variation has on model-implied mean variation.

[12] This data set was compiled using information from the 1995 U.S. News & World Report's Guide to America's Best Colleges and the American Association of University Professors (AAUP) annual faculty salary survey of American colleges and universities. For more detail see https://ww2.amstat.org/publications/jse/datasets/usnews.txt.

However, these descriptions are not too informative in the absence of a quantitative indication of *how much* variance in salary can be explained by each source. Inputting the estimates obtained from M*plus* into our R function *regMixR2* (described in the Discussion section) yields all relevant R^2 measures and decompositions. These decompositions (see Table 4) supplement the qualitative descriptions by quantifying the proportion of variance explained by each distinct component. Hence, we can more precisely state that the smallest proportion of the total variance in college-provided salary is attributable to predictors via between-class differences in their slopes (15%), followed by that attributable to between-class differences in the means of salary (21%); the largest proportion (47%) is explained by the predictors via their marginal slopes. This quantitative decomposition is illustrated in a bar chart in Figure 5, which communicates the importance of considering heterogeneity across subpopulations of schools when assessing effects on faculty salary.

Researchers interested in all three components of explained variance may report $R^{2(fvm)} = .83$ as their primary summary measure. However, if they are more interested in the first and third components, they may report $R^{2(fv)} = .62$. If they are, instead, solely interested in the third component, across-class average effects, they could report $R^{2(f)} = .47$.

Lastly, we consider the class-specific R^2s. For the three classes, these values are $R^{2(f)}_{k=1} = .26$, $R^{2(f)}_{k=2} = .82$, and $R^{2(f)}_{k=3} = .86$. These measures provide evidence that the three predictors are more important sources of explained variance for the liberal arts college class and the research university class, as compared with the regional college class.

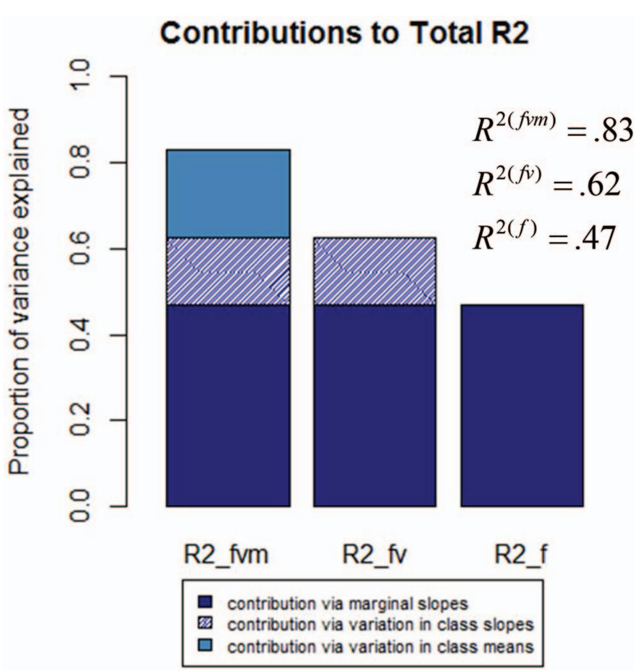

Figure 5. Empirical Example 1 results: Decompositions of total R^2's for a fitted single-level regression mixture. Equations corresponding to these decompositions are given in Table 4. See the online article for the color version of this figure.

Data Model for Multilevel Regression Mixture With Classes at Both Level-1 and Level-2

We now consider a more general multilevel regression mixture that allows latent classes at both Level-1 and Level-2 (e.g., Vermunt, 2010; Vermunt & Magidson, 2005). Let d_j denote Level-2 (i.e., cluster-level) latent class membership for cluster j, and let c_{ij} denote Level-1 latent class membership for observation i in cluster j. Level-2 classes range from $h = 1$ to $h = H$, whereas Level-1 classes range from $k = 1$ to $k = K$. This yields class-combinations ranging from $kh = 11$ to $kh = KH$. Each class-combination can be seen as a Level-1 class that is nested within a given Level-2 class.[13] We can then model a univariate y_{ij} conditional on class-combination membership.

$$
\begin{aligned}
&y_{ij|c_{ij}=k,d_j=h} = \mathbf{x}'_{ij}\boldsymbol{\gamma}^{kh} + \varepsilon_{ij}\\
&\varepsilon_{ij} \sim N(0, \theta^{kh})\\
&p(d_j = h) = \pi^h = \frac{\exp(\varpi^h)}{\sum_{h=1}^{H}\exp(\varpi^h)}\\
&p(c_{ij} = k \mid d_j = h) = \pi^{k|h} = \frac{\exp(\omega^k + \delta^{kh})}{\sum_{k=1}^{K}\exp(\omega^k + \delta^{kh})}
\end{aligned}
\tag{7}
$$

Here, $\mathbf{x}'_{ij}$ denotes a row vector consisting of 1 and all predictors for observation *ij*. This vector is multiplied by the column vector $\boldsymbol{\gamma}^{kh}$ containing all regression coefficients specific to class-combination *kh*. The residual, ε_{ij}, is normally distributed with a class-combination-specific variance, θ^{kh}. The probability of cluster j belonging to Level-2 class h, $p(d_j = h)$, is denoted π^h and is modeled by a Level-2 class-specific multinomial intercept, ϖ^h. The conditional probability of observation i in cluster j belonging to Level-1 class k, given that their cluster is a member of Level-2 class h, is denoted $\pi^{k|h}$ and is modeled by a multinomial intercept, ω^k, and multinomial slope, δ^{kh}. The following are constrained to 0 for identification: ϖ^H, ω^K, δ^{kH} for all k, and δ^{Kh} for all h.

Note that, in this general specification, a given k need not be comparable across Level-2 class (e.g., $kh = 11$ may not be similar to $kh = 12$). It is, however, possible to place constraints on the model to make each k comparable (i.e., have equal parameter estimates) across h (i.e., $\gamma_q^{k1} = \ldots = \gamma_q^{kH}$ and $\theta^{k1} = \ldots = \theta^{kH}$ for each k and q; see, e.g., Lukočienė et al., 2010). In this more constrained specification, Level-2 classes are distinguished solely by varying Level-1 class proportions within Level-2 class. Additionally, it is possible to constrain particular Level-1 class parameters equal within a given Level-2 class (i.e., $\gamma_q^{1h} = \ldots = \gamma_q^{Kh}$). The computation of R^2 measures we describe in subsequent sections applies to any of these constrained or unconstrained specifications. Though for more constrained specifications computations could be slightly simplified (e.g., computing weighted averages across Level-2 classes rather than across all class-combinations), we continue to reference the Equation (7) specification for generality.

[13] Notation used here corresponds to the common specification wherein each Level-2 class has the same number of Level-1 classes nested within it. If this is not the case, K would be replaced by K_h. Note that our software tool allows for the latter scenario where K differs across h.

R-Squared Measures for Multilevel Regression Mixtures With Classes at Both Level-1 and Level-2

We now proceed to delineate seven R^2 measures that apply to the model in Equation (7) with classes at both Level-1 and Level-2. An overview of these measures is given in Table 2. Note that some measures in Table 2—the total R^2's and class-combination R^2's—have analogues in the earlier-presented Table 1 for simpler mixtures, but, as we will show, both require modified computation and, furthermore, the total R^2's can now be decomposed in new ways. The other kind of measure in Table 2—Level-2 class-specific R^2's—have no exact analog in Table 1.

Total R-Squared Measures

For the multilevel regression mixture with classes at *both* levels, the expressions for total R^2's parallel the earlier-presented Equations (3), (4), and (5) for simpler regression mixtures with classes at *only one* level—with two exceptions. These exceptions are that marginalized regression coefficients are now denoted $\boldsymbol{\gamma}^{\bullet\bullet}$ and the marginalized residual variance is now denoted $\theta^{\bullet\bullet}$. These are both weighted averages across all $K \times H$ class-combinations. For instance, $\boldsymbol{\gamma}^{\bullet\bullet} = \sum_{h=1}^{H}\sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{kh}$, where $\pi^{kh} = \pi^{k|h}\pi^{h}$ (i.e., class-combination probability). Other terms in Equations (8), (9), and (10) retain the same definitions as in Table 3. As before, the following total R^2's are distinguished by the sources in Table 2.

Total $R_T^{2(fvm)}$ measure. To answer the question: "What proportion of variance is explained by class-combination variation in outcome means as well as by predictors via their marginal slopes and their across-class-combination variation in slopes?" a researcher could compute $R_T^{2(fvm)}$:

$$R_T^{2(fvm)} = \frac{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}}{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet} + \theta^{\bullet\bullet}} \tag{8}$$

Note that, comparing Equations (8) and (3), the difference is that Equation (8) contains double-dot superscripts rather than single-dot superscripts, reflecting the fact that Equation (8) involves marginalization across *both* Level-1 and Level-2 class. The full derivation of the total model-implied variance in the denominator of Equation (8) is given in Appendix A.

Total $R_T^{2(fv)}$ measure. To instead answer the question: "What proportion of variance is explained by predictors via their marginal slopes and their across-class-combination variation in slopes?" a researcher could compute $R_T^{2(fv)}$:

$$\begin{aligned} R_T^{2(fv)} &= \frac{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}}{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet} + \theta^{\bullet\bullet}} \\ &\quad - \frac{\operatorname{var}\left(E\left[\mathbf{x}_{ij}'\boldsymbol{\gamma}^{kh} + \varepsilon_{ij} \mid c_{ij} = k, d_j = h\right]\right)}{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet} + \theta^{\bullet\bullet}} \\ &= \frac{\mathbf{v}'\boldsymbol{\psi} + 2\mathbf{r}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}}{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet} + \theta^{\bullet\bullet}} \end{aligned} \tag{9}$$

$R_T^{2(fv)}$ is obtained analogously to Equation (4). A full derivation of the numerator of this Equation (9) expression is provided in Appendix B.

Total $R_T^{2(f)}$ measure. A researcher now posing the question: "What proportion of variance is explained by predictors via their marginal slopes?" could compute $R_T^{2(f)}$:

$$R_T^{2(f)} = \frac{\boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}}{\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet} + \theta^{\bullet\bullet}} \tag{10}$$

Level-2 Class-Specific R-Squared Measures

A researcher can ask the three research questions described above but isolate the proportion of variance within a given Level-2 class h. To do so, the researcher would compute three separate R^2 measures for a given Level-2 class h; these measures differ in what types of heterogeneity across the K Level-1 classes nested within h are considered explained variance. The expressions for these three Level-2 class-specific R^2's, in Equations (11)–(13) below, parallel the earlier-presented total R^2's in Equations (8)–(10)—with three exceptions. A first exception is that the denominator of Equations (11)–(13) now contains the model-implied total variance of y_{ij} *within Level-2 class h.* A second exception is that the new superscript $\bullet h$ indicates *marginalized across all Level-1 classes within h.* As such, the $\boldsymbol{\gamma}^{\bullet h}$ vector in Equations (11)–(13) contains *marginal Level-2 class* regression coefficients (i.e., $\boldsymbol{\gamma}^{\bullet h} = \sum_{k=1}^{K}\pi^{k|h}\boldsymbol{\gamma}^{kh}$) and $\theta^{\bullet h}$ denotes the *marginal Level-2 class* residual variance. A final exception is that the $\boldsymbol{\psi}^{h}$ vector contains variances of all regression coefficients *across k within h*, whereas $\boldsymbol{\kappa}^{h}$ contains their covariances *across k within h*. Other terms in Equations (11)–(13) retain the same definitions as in Table 3.

Level-2 class-specific $R_h^{2(fvm)}$ measure. Suppose a researcher asked: "For Level-2 class h, what proportion of variance is explained by across-k variation in outcome means as well as by predictors via their marginal h-class slopes and their across-k variation in slopes?" This research question is addressed by computing $R_h^{2(fvm)}$, defined as the proportion of variance explained within Level-2 class h, *without* partialing out any Level-1 class mean or slope variation within class h. In Equation (11), $R_h^{2(fvm)}$ is computed as:

$$R_h^{2(fvm)} = \frac{\mathbf{s}'\boldsymbol{\psi}^{h} + 2\mathbf{p}'\boldsymbol{\kappa}^{h} + \boldsymbol{\gamma}^{\bullet h\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet h}}{\mathbf{s}'\boldsymbol{\psi}^{h} + 2\mathbf{p}'\boldsymbol{\kappa}^{h} + \boldsymbol{\gamma}^{\bullet h\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet h} + \theta^{\bullet h}} \tag{11}$$

Level-2 class-specific $R_h^{2(fv)}$ measure. A researcher may also want to know: "For Level-2 class h, what proportion of variance is explained by predictors via their marginal h-class slopes and their across-k variation in slopes?" To address this research question, we compute $R_h^{2(fv)}$, which, unlike the $R_h^{2(fvm)}$ described above, does not allow for mean variation across Level-1 class within Level-2 class to contribute to explained variance. Similar to Equation (9), we subtract from $R_h^{2(fvm)}$ the variance of model-implied means of each class-combination nested within h, divided by the total model-implied variance of y_{ij} within class h.[14]

$$\begin{aligned} R_h^{2(fv)} &= \frac{\mathbf{s}'\boldsymbol{\psi}^{h} + 2\mathbf{p}'\boldsymbol{\kappa}^{h} + \boldsymbol{\gamma}^{\bullet h\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet h}}{\mathbf{s}'\boldsymbol{\psi}^{h} + 2\mathbf{p}'\boldsymbol{\kappa}^{h} + \boldsymbol{\gamma}^{\bullet h\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet h} + \theta^{\bullet h}} \\ &\quad - \frac{\operatorname{var}_{k|h}\left(E\left[\mathbf{x}_{ij}'\boldsymbol{\gamma}^{kh} + \varepsilon_{ij} \mid c_{ij} = k, d_j = h\right]\right)}{\mathbf{s}'\boldsymbol{\psi}^{h} + 2\mathbf{p}'\boldsymbol{\kappa}^{h} + \boldsymbol{\gamma}^{\bullet h\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet h} + \theta^{\bullet h}} \\ &= \frac{\mathbf{v}'\boldsymbol{\psi}^{h} + 2\mathbf{r}'\boldsymbol{\kappa}^{h} + \boldsymbol{\gamma}^{\bullet h\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet h}}{\mathbf{s}'\boldsymbol{\psi}^{h} + 2\mathbf{p}'\boldsymbol{\kappa}^{h} + \boldsymbol{\gamma}^{\bullet h\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet h} + \theta^{\bullet h}} \end{aligned} \tag{12}$$

[14] The notation $\operatorname{var}_{k|h}$ indicates taking the variance across Level-1 classes within Level-2 class h.

Level-2 class specific $R_h^{2(f)}$ measure. To address the question: "For Level-2 class h, what proportion of variance is explained by predictors via the marginal h-class slopes?" one can compute $R_h^{2(f)}$. This $R_h^{2(f)}$ measures the proportion of variance explained within Level-2 class h, partialing out all mean and slope variation across Level-1 classes within class h.

$$R_h^{2(f)} = \frac{\boldsymbol{\gamma}^{\bullet h'}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet h}}{\mathbf{s}'\boldsymbol{\psi}^h + 2\mathbf{p}'\boldsymbol{\kappa}^h + \boldsymbol{\gamma}^{\bullet h'}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet h} + \theta^{\bullet h}} \tag{13}$$

Class-Combination-Specific R-Squared Measure

Earlier, in Equation (6), we presented a class-specific R^2 for a mixture with classes at a single level. Its analog, when classes exist at two levels and each Level-2 class has a distinct set of Level-1 classes nested within it, is a class-combination-specific R^2. If, as mentioned previously, researchers want to constrain all within-Level-1 class coefficients equal across Level-2 classes (for parsimony or to ensure comparability of interpretation across Level-2 classes) the class-combination-specific R^2 would instead be a *Level-1 class-specific* R^2 and would be interpreted in the manner of Equation (6). Note that with this more constrained specification, the interpretation of the total R^2's and the Level-2 class-specific R^2's would be unchanged.

Lastly, a researcher might pose the question: "For a particular class-combination kh, what proportion of variance is explained by predictors?" To address this, we compute $R_{kh}^{2(f)}$. Its computation involves dividing the model-implied variance of predicted scores within a given kh by the model-implied total variance of y_{ij} within kh.

$$R_{kh}^{2(f)} = \frac{\boldsymbol{\gamma}^{kh'}\boldsymbol{\Phi}\boldsymbol{\gamma}^{kh}}{\boldsymbol{\gamma}^{kh'}\boldsymbol{\Phi}\boldsymbol{\gamma}^{kh} + \theta^{kh}} \tag{14}$$

This $R_{kh}^{2(f)}$ can be useful to assess how the proportion of variance explained by the predictors differs across the latent class combinations.

Simulated Demonstrations of Relationships Among R-Squareds for Multilevel Regression Mixtures With Classes at Level-1 and Level-2

We have now covered seven R^2 measures that are applicable to multilevel regression mixtures with classes at both levels. It is important to understand the relationships among these R^2's, rather than simply understanding the meaning of each in isolation. We now illustrate these relationships by showing how they are affected by different patterns of parameter values, as we similarly demonstrated for single-level regression mixtures. For simplicity of visualization, these illustrations involve a multilevel regression mixture with one (grand or cluster-mean) centered predictor, x_{ij}, and $K = 2$ Level-1 latent classes and $H = 2$ Level-2 latent classes. As before, all class-combination residual variances are equal to 1. In Figures 6–7, each panel shows $K \times H = 4$ class-combination-specific regression lines, with class-combination labels beside each line.

Manipulating potential sources of explained variance. In Figure 6 we manipulate each of the components, one at a time, that could contribute to a total R^2; the consequences for the suite of R^2 measures are illustrated by comparing Panel A to each of Panels B–F. We first consider how increasing the *marginal slope* influences the suite of R^2 measures. Comparing Figure 6 Panel A to Panel B, we increased the marginal slope by increasing each class-combination slope by the same amount, holding all else constant. Each total R^2 increases because variance attributable to predictors via marginal slopes is *explained variance* in all measures.

We next consider the effect of increasing *across-class-combination slope variance*. There are now two ways of accomplishing this. First, as shown in Figure 6 Panel C, we can increase *between Level-2 class slope variance*, holding within Level-2 class slope variance constant. In other words, when comparing Figure 6 Panels A and C, we changed the heterogeneity between only Level-2 classes—differences between Level-1 classes within each Level-2 class remain the same. Alternatively, as shown when comparing Figure 6 Panels A and D, we can increase *within Level-2 class slope variance* (for simplicity, just done for $h = 2$) holding between Level-2 class slope variance constant. Either manipulation has the same effect of increasing $R_T^{2(fvm)}$ and $R_T^{2(fv)}$ while decreasing $R_T^{2(f)}$—reflecting the fact that this source is *explained variance* in the former two measures and *residual variance* in the latter measure.

Lastly, we consider the effect of increasing *across-class-combination mean outcome variation*. There are now two ways of increasing mean variation. We can increase *between* Level-2 class mean variance holding all else constant, as demonstrated by comparing Figure 6 Panels A and E. We can instead increase *within* Level-2 class mean variance (here, only in $h = 1$), as demonstrated by comparing Figure 6 Panels A and F. Either manipulation serves to increase the $R_T^{2(fvm)}$ but decrease $R_T^{2(fv)}$ and $R_T^{2(f)}$, reflecting the property that this source is *explained variance* in $R_T^{2(fvm)}$ but *residual variance* in $R_T^{2(fv)}$ and $R_T^{2(f)}$.

Different patterns of class-combination parameters can yield identical total (or Level-2) R-squareds. As a separate demonstration, in Figure 7, we show how very different patterns of class-specific parameters can still yield the same total R^2. We provide this demonstration for one total R^2, $R_T^{2(fvm)}$, but this principle could also be demonstrated for $R_T^{2(fv)}$ or $R_T^{2(f)}$. In Figure 7, Panels A–C have the exact same total $R_T^{2(fvm)}$ but their regression coefficients correspond to different substantive interpretations. For instance, Figure 7 Panel A corresponds to an interpretation of "heterogeneity of the effect of x_{ij} across class-combinations, but no heterogeneity in means of y_{ij}." In contrast, Panel B corresponds to an interpretation of "*no* heterogeneity of the effect of x_{ij} across class-combinations, but heterogeneity in means of y_{ij}." A more complex situation is illustrated in Panel C, wherein within $h = 1$ there is heterogeneity in means only, but within $h = 2$ there is heterogeneity in slopes only.

Likewise, it is also the case that substantively different patterns of Level-1 parameters within a given Level-2 class can yield the same Level-2 class-specific R^2. We demonstrate this for one Level-2 class-specific R^2, $R_{h=1}^{2(fvm)}$, but note that it could be demonstrated for the $R_{h=1}^{2(fv)}$ and $R_{h=1}^{2(f)}$. Figure 7 Panels A versus C have the same $R_{h=1}^{2(fvm)}$ for Level-2 class $h = 1$ (despite heterogeneity in slopes but not means in Panel A and heterogeneity in means but not slopes in Panel C).

Overall, Figure 7 illustrates that a total R^2 and a Level-2 class-specific R^2 are omnibus statistics that can reflect many different patterns of class-combination effects. Thus, it is our recommendation that researchers examine all types of R^2's in juxtaposition to

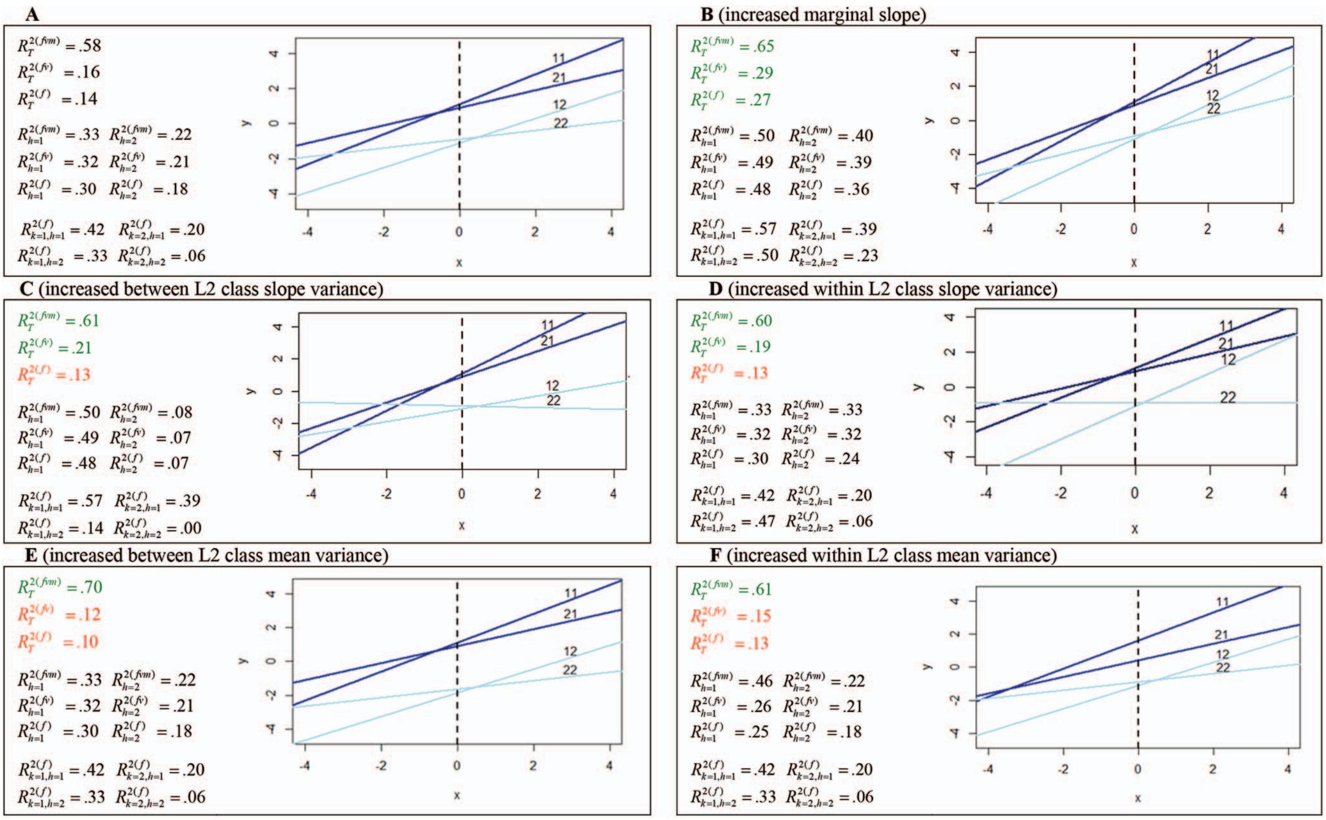

Figure 6. Simulated demonstration: Consequences for total R^2's of manipulating the five sources of explained variance for multilevel regression mixture with classes at both levels. Compare Panel A with each of Panels B–F. Each line denotes a class-combination-specific regression line; the two dark lines are for Level-2 Class 1 and the two light lines are for Level-2 Class 2. (green = R^2 increased, red = R^2 decreased, compared with Panel A; L2 = Level 2). See the online article for the color version of this figure.

each other to understand what type of variance one's model explains. Furthermore, researchers can again compute and report a decomposition of explained variance—described in the next section—to get a more precise understanding of the sources of explained variance.

Decomposing Explained Variance in R-Squareds for Multilevel Regression Mixtures With Classes at Level-1 and Level-2

Earlier, for regression mixtures with classes at *only one* level, we analytically decomposed total R^2's into three meaningful components that were separately interpreted and portrayed using bar charts. Now, in the context of mixtures with classes at *both* levels, we extend these developments in two new ways: We decompose both Level-2 class-specific R^2's as well as total R^2's and, furthermore, we decompose total R^2's into *five* components of explained variance.

The Level-2 class R^2's can be decomposed into three, two, or one distinct components of explained variance. This decomposition distinguishes the R^2's based on what specific factors constitute explained variance versus residual variance. Table 5 shows how the $R_h^{2(fvm)}$ is decomposed into three distinct components: (a) contribution of predictors via their marginal Level-2 slopes; (b) contribution of predictors via variation in Level-1 class slopes within Level-2 class h; and (c) the variation in Level-1 class means within Level-2 class h. Table 5 shows how the $R_h^{2(fv)}$ can be decomposed into the first two of these components, and how $R_h^{2(f)}$ consists of only the first component. Furthermore, Table 5 shows how these components can be obtained from earlier-presented formulas. Table 5 clarifies that, for instance, any variation in Level-1 class means within Level-2 class h contributes to *explained variance* in $R_h^{2(fvm)}$ but contributes to *unexplained, residual variance* in $R_h^{2(fv)}$ and $R_h^{2(f)}$. Table 5 also clarifies that variation in Level-1 class slopes within Level-2 class h contributes to *explained variance* in $R_h^{2(fvm)}$ and $R_h^{2(fv)}$ but not $R_h^{2(f)}$. As was the case for single-level regression mixtures, the following relationship will always hold: $R_h^{2(f)} \leq R_h^{2(fv)} \leq R_h^{2(fvm)}$.

The total R^2's for the multilevel mixture with classes at both levels can also be decomposed, but into *five* distinct components. Unlike previous decompositions, we can now consider both *within Level-2 class* mean and slope variance as well as *between Level-2 class* mean and slope variance, as demonstrated in the previous section. Specifically, Table 6 shows how $R_T^{2(fvm)}$ can be decomposed into the: (a) contribution of predictors via marginal slopes; (b) contribution of predictors via between-h slope variance (i.e., variance in marginal Level-2 class slopes); (c) contribution of

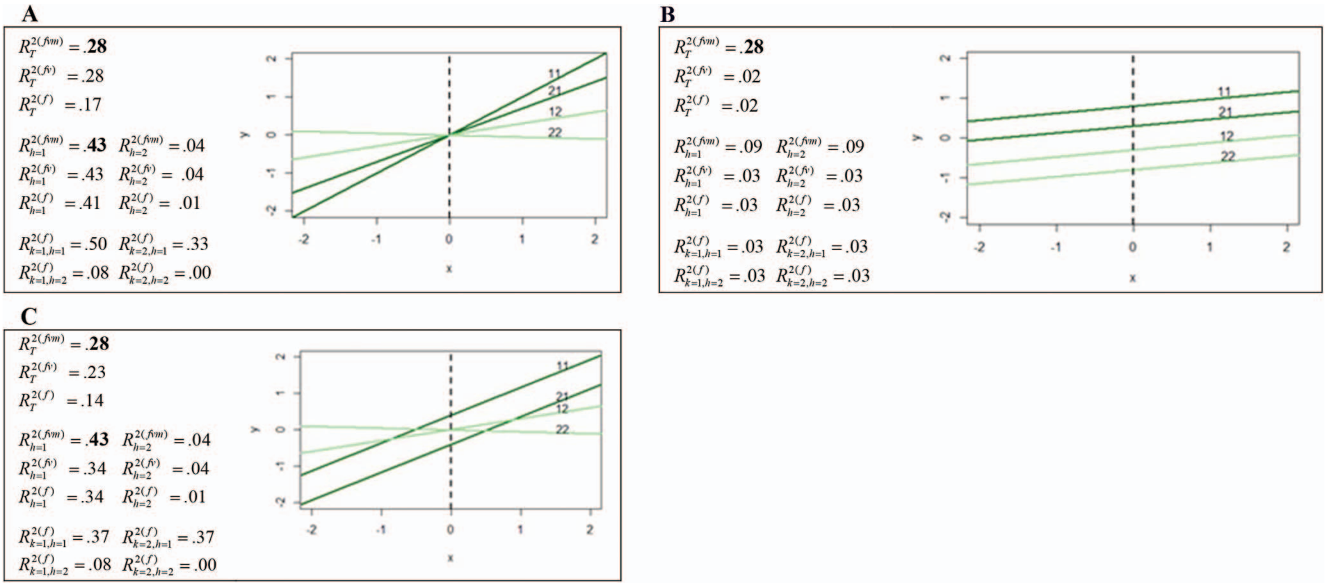

Figure 7. Simulated demonstration: Substantively different patterns of class-combination-specific regression equations can yield the same total R^2 (and/or same Level-2 class-specific R^2) in a regression mixture with classes at both levels. Compare Panel A with each of Panels B and C. Each line denotes a class-combination-specific regression line. The two dark lines are for Level-2 Class 1 and the two light lines are for Level-2 Class 2. **Bold** R^2 results are compared and discussed in the article. See the online article for the color version of this figure.

predictors via within-*h* slope variance (i.e., variance in Level-1 class slopes within Level-2 class); (d) contribution of outcome means via between-*h* variation; and (e) contribution of outcome means via within-*h* variation. In contrast, Table 6 shows that $R_T^{2(fv)}$ can be decomposed into only the first three of these components, whereas $R_T^{2(f)}$ consists only of the first component.

As shown in Table 6, these components can be obtained using formulas described earlier, together with the following two additional formulas. These two additional formulas are needed to distinguish within Level-2 class variation from between Level-2 class variation. Specifically, the Table 6 decompositions require knowing the proportion of variance explained when using marginal Level-2 class regression coefficients to compute predicted scores, which can be denoted using previously defined terms as:

$$\frac{\mathbf{s}'\left(\boldsymbol{\psi}-\sum_{h=1}^{H}\pi^h\boldsymbol{\psi}^{\bullet h}\right)+2\mathbf{p}'\left(\boldsymbol{\kappa}-\sum_{h=1}^{H}\pi^h\boldsymbol{\kappa}^{\bullet h}\right)+\boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}}{\mathbf{s}'\boldsymbol{\psi}+2\mathbf{p}'\boldsymbol{\kappa}+\boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}+\theta^{\bullet\bullet}} \quad (15)$$

The Table 6 decompositions also require knowing the proportion of variance explained when using marginal Level-2 class regression slopes to compute predicted scores, partialing out marginal Level-2 class mean separation:

$$\frac{\mathbf{v}'\left(\boldsymbol{\psi}-\sum_{h=1}^{H}\pi^h\boldsymbol{\psi}^{\bullet h}\right)+2\mathbf{r}'\left(\boldsymbol{\kappa}-\sum_{h=1}^{H}\pi^h\boldsymbol{\kappa}^{\bullet h}\right)+\boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}}{\mathbf{s}'\boldsymbol{\psi}+2\mathbf{p}'\boldsymbol{\kappa}+\boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}+\theta^{\bullet\bullet}} \quad (16)$$

Table 5
Decomposing Variance Explained in Level-2-Class-Specific R^2's for Multilevel Regression Mixtures With Classes at Level-1 and Level-2

$$R_h^{2(fvm)} = \left\{\begin{array}{l}\text{contribution of predictors}\\ \text{via marginal L2 class}\\ \text{slopes for } h\end{array}\right\} + \left\{\begin{array}{l}\text{contribution of predictors}\\ \text{via variation in L1 class}\\ \text{slopes within L2 class } h\end{array}\right\} + \left\{\begin{array}{l}\text{contribution of means}\\ \text{via variation in L1 class}\\ \text{means within L2 class } h\end{array}\right\}$$
$$= \{\ \text{Equation(13)}\ \} + \{\ \text{Eqn(12)} - \text{Eqn(13)}\ \} + \{\ \text{Eq(11)} - \text{Eq(12)}\ \}$$

$$R_h^{2(fv)} = \left\{\begin{array}{l}\text{contribution of predictors}\\ \text{via marginal L2 class}\\ \text{slopes for } h\end{array}\right\} + \left\{\begin{array}{l}\text{contribution of predictors}\\ \text{via variation in L1 class}\\ \text{slopes within L2 class } h\end{array}\right\}$$
$$= \{\ \text{Equation(13)}\ \} + \{\ \text{Eq(12)} - \text{Eq(13)}\ \}$$

$$R_h^{2(f)} = \left\{\begin{array}{l}\text{contribution of predictors}\\ \text{via marginal L2 class}\\ \text{slopes for } h\end{array}\right\}$$
$$= \{\ \text{Equation(13)}\ \}$$

Table 6
Decomposing Variance Explained in Total R^2's for Multilevel Regression Mixtures With Classes at Level-1 and Level-2

$$R_T^{2(fvm)} = \left\{\begin{array}{l}\text{contribution of}\\ \text{predictors via}\\ \text{marginal slopes}\end{array}\right\} + \left\{\begin{array}{l}\text{contribution of}\\ \text{predictors via}\\ \text{variation in}\\ \text{L2 class slopes}\end{array}\right\} + \left\{\begin{array}{l}\text{contribution of}\\ \text{predictors via}\\ \text{variation in}\\ \text{L1 class slopes}\\ \text{within L2 class}\end{array}\right\} + \left\{\begin{array}{l}\text{contribution of}\\ \text{means via}\\ \text{variation in}\\ \text{L2 class means}\end{array}\right\} + \left\{\begin{array}{l}\text{contribution of}\\ \text{means via}\\ \text{variation in}\\ \text{L1 class means}\\ \text{within L2 class}\end{array}\right\}$$

$$= \{\ \text{Equation}(10)\ \} + \{\text{Eq}(16) - \text{Eq}(10)\} + \{\text{Eq}(9) - \text{Eq}(16)\} + \{\text{Eq}(15) - \text{Eq}(16)\} + \left\{\begin{array}{l}\text{Eq}(8) - \text{Eq}(9) -\\ \text{Eq}(15) + \text{Eq}(16)\end{array}\right\}$$

$$R_T^{2(fv)} = \left\{\begin{array}{l}\text{contribution of}\\ \text{predictors via}\\ \text{marginal slopes}\end{array}\right\} + \left\{\begin{array}{l}\text{contribution of}\\ \text{predictors via}\\ \text{variation in}\\ \text{L2 class slopes}\end{array}\right\} + \left\{\begin{array}{l}\text{contribution of}\\ \text{predictors via}\\ \text{variation in}\\ \text{L1 class slopes}\\ \text{within L2 class}\end{array}\right\}$$

$$= \{\ \text{Equation}(10)\ \} + \{\text{Eq}(16) - \text{Eq}(10)\} + \{\text{Eq}(9) - \text{Eq}(16)\}$$

$$R_T^{2(f)} = \left\{\begin{array}{l}\text{contribution of}\\ \text{predictors via}\\ \text{marginal slopes}\end{array}\right\}$$

$$= \{\ \text{Equation}(10)\ \}$$

Our R program (described later) automatically computes the quantities in Tables 5 and 6 and produces bar charts that allow a visual illustration of these decompositions. Such bar charts are utilized in the presentation of the empirical example results in the next section.

Empirical Example 2

In this section we consider another empirical example involving a multilevel regression mixture application of the Equation (7) model using a subset of the High School and Beyond dataset (National Center for Educational Statistics, 1986)[15] with $N = 2{,}640$ high school teachers nested within $J = 100$ high schools. In our model, we predict teacher satisfaction from teacher-reported control over curriculum (*control*), teacher-reported quality of principal's leadership (*leadership*), and principal-reported school-level delinquency (*delinquency*). Though all three variables had been found predictive of teacher satisfaction in prior nonmixture studies (Bogler, 2001; LeBlanc, Swisher, Vitaro, & Tremblay, 2007; Pearson & Moomaw, 2005), our interest was in exploring how determinants of satisfaction could have differing effects across types of teachers and schools. Control and leadership were cluster-mean-centered whereas delinquency was grand-mean-centered. All intercepts and slopes were allowed to vary across class-combinations; for parsimony, we constrained residual variances equal.

Using M*plus* to estimate the models and BIC as a selection criterion, we compared the fit of all combinations of $K = 1$ with 6 and $H = 1$ with 6. BIC preferred three teacher-level classes nested within each of two school-level classes (i.e., $K = 3$, $H = 2$). The parameter estimates and standard errors for this best-fitting solution are shown in online supplemental Appendix B and patterns of significant effects are summarized here. The two school-level classes (i.e., the two sets of three Level-1 classes nested within each h) were distinguished by both mean levels of teacher satisfaction and the effect of delinquency. Class $h = 1$ reflects schools with more-satisfied teachers where there is no effect of delinquency, whereas $h = 2$ reflects schools with dissatisfied teachers who are even less satisfied as delinquency increases. Above and beyond these Level-2 class differences, the pattern of teacher-level (Level-1) classes was quite consistent within each of these Level-2 classes. Class $k = 1$ was most populous and reflects highly satisfied teachers, regardless of levels of the predictors. Class $k = 2$ reflects less satisfied teachers, whose satisfaction improves with better principal leadership and more personal control. Class $k = 3$ was least populous and reflects less satisfied teachers, whose satisfaction improves with better leadership (but not personal control).

These summary descriptions indicate that there is some heterogeneity across and within school-level class in terms of the effects of the predictors and in terms of the mean of job satisfaction. Using the decomposition obtained from *regMixR2* (see Table 6), we can now supplement these descriptions by clarifying that the total variance in job satisfaction can primarily be explained by two things: (a) the predictors via their marginal slopes (18%); and (b) the heterogeneity in means between (7%) and within (44%) Level-2 class. As shown in the bar chart in Figure 8, comparatively little total variance was attributable to predictors via slope differences—either between Level-2 class (1%) or within Level-2 class (3%). Thus, we may conclude that the across-school-level class differences in the predictors' effects are, in fact, not practically significant, whereas there might be some substantively meaningful differences in means. Such a conclusion would have been less clear from just examining class-specific regression coefficients. Furthermore, for this example, it could be misleading to report $R_T^{2(fvm)}$ (.73) as a summary measure if one were primarily interested in the potential effects of predictors and not in outcome mean differences across class. In this case, it would be more useful to report $R_T^{2(fv)}$ (.22) or $R_T^{2(f)}$ (.18).

Results for the Level-2 R^2's and decompositions are shown in the bar charts of Figure 9. These results were consistent for the two

[15] For more information see: https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/106.

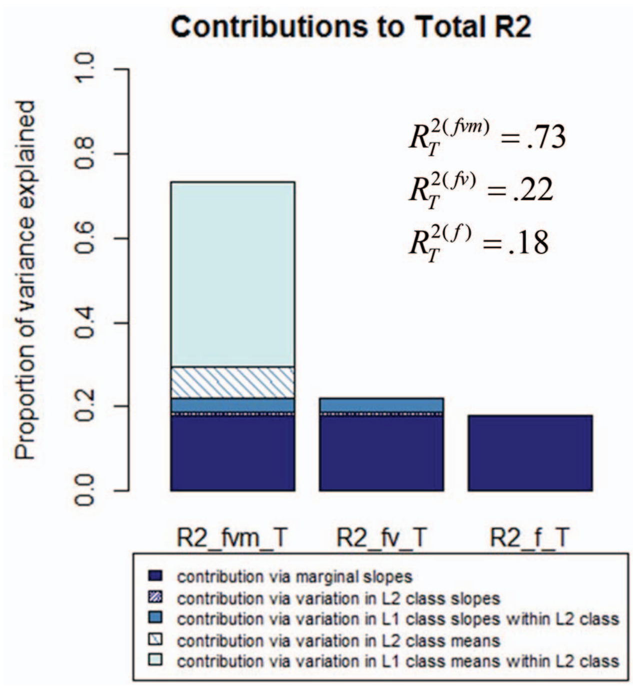

Figure 8. Empirical Example 2 results: Decompositions of total R^2's for a fitted multilevel regression mixture with classes at Level-1 and Level-2. Equations corresponding to these decompositions are given in Table 6. See the online article for the color version of this figure.

Level-2 latent classes and mirrored the general patterns found for the total R^2's; thus, they are not discussed further here.

Lastly, we consider the class-combination R^2's. From our above qualitative description of the teacher-level latent classes, we would expect that the "high-satisfaction regardless of predictors" teacher-level classes would have less variance explained in satisfaction compared to the other two teacher-level classes. This is confirmed and quantified by comparing the two high-satisfaction class-combination R^2's ($R^{2(f)}_{k=1,h=1}$ = .15 and $R^{2(f)}_{k=1,h=2}$ = .13) to the other four class-combination R^2's ($R^{2(f)}_{k=2,h=1}$ = .52 and $R^{2(f)}_{k=2,h=2}$ = .59 for k = 2; $R^{2(f)}_{k=3,h=1}$ = .45 and $R^{2(f)}_{k=3,h=2}$ = .59 for k = 3); the predictors better explain teacher satisfaction within the latter four class-combinations.

Discussion

Regression mixture models are often applied in single-level (i.e., unclustered) and multilevel (i.e., clustered) data analysis contexts in the social sciences. Though researchers applying nonmixture regression models widely report R^2 measures of explained variance as effect sizes, there has been no general treatment of R^2 measures for regression mixtures with classes at one or two levels. Consequently, it is common for applied researchers to summarize results of a fitted regression mixture by simply reporting significant class-specific regression coefficients and providing qualitative class-label interpretations, rather than considering measures of effect size. In this article, we have filled this gap by providing an integrative framework of R^2 measures for single-level regression mixture models and multilevel regression mixture models (with classes at Level-2 or both levels). Specifically, we described 11 R^2 measures, most of which were newly developed here. Using two empirical examples, we showed how each measure can help researchers answer distinct substantive questions. We related these measures analytically and through graphical illustrations. Further, we newly demonstrated how these R^2's can be decomposed into substantively meaningful sources of explained variance. In the remainder of the Discussion, we address software implementation, limitations, and future directions.

Software Implementation

To aid researchers in utilizing the R^2 measures in practice, we developed an R function, *regMixR2*, that computes all 11 R^2 measures delineated in our framework. This function is found in online supplemental Appendix C and is briefly described here.

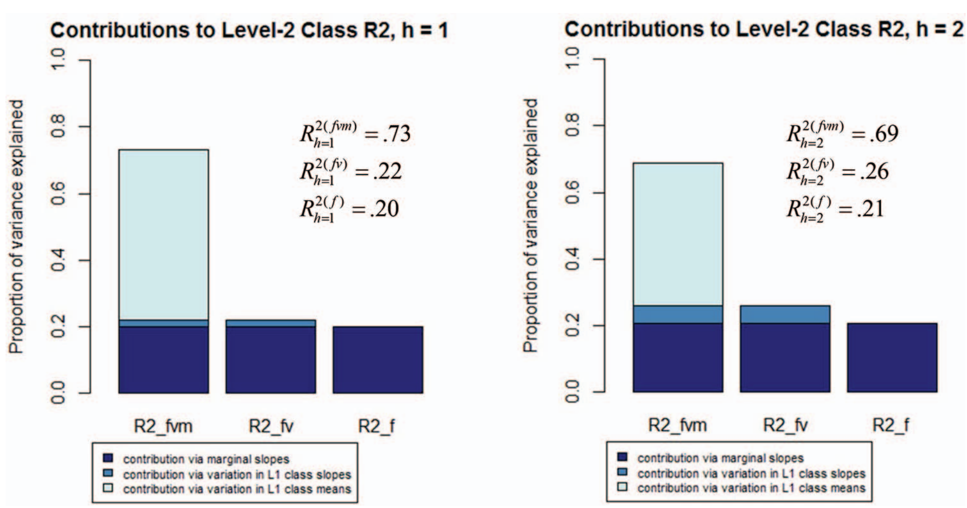

Figure 9. Empirical Example 2 results: Decompositions of Level-2 class-specific R^2's for a fitted multilevel regression mixture with classes at Level-1 and Level-2. Equations corresponding to these decompositions are given in Table 5. See the online article for the color version of this figure.

With this function, a user inputs raw data, the number of Level-1 and/or Level-2 classes from their regression mixture model, and the parameter estimates from their regression mixture model. The function then outputs the following: (a) all relevant R^2 measures for that fitted model; (b) all relevant decompositions of explained variance; and (c) bar chart illustrations of these decompositions (as shown in Figures 5, 8, and 9). Note that obtained estimates from any mixture modeling software package can serve as input to our function so long as all of the relevant estimates (detailed in online supplemental Appendix C) are provided. Though we used M*plus* for the empirical examples in this article (see online supplemental Appendix D for illustrative M*plus* syntax), other software can fit single-level regression mixtures (e.g., *FlexMix*, Leisch, 2004; Latent GOLD, Vermunt & Magidson, 2016) as well as multilevel regression mixtures (e.g., Latent GOLD).

As mentioned in the introduction, two commercial software programs, M*plus* (Muthén & Muthén, 1998–2016) and Latent GOLD (Vermunt & Magidson, 2016), output certain R^2 measures for regression mixtures, but with some limitations. Specifically, M*plus* does not provide quantities analogous to our total or Level-2 class-specific R^2's ($R^{2(fvm)}$, $R^{2(fv)}$, $R^{2(f)}$, $R_T^{2(fvm)}$, $R_T^{2(fv)}$, $R_T^{2(f)}$, $R_h^{2(fvm)}$, $R_h^{2(fv)}$, $R_h^{2(f)}$) nor does it provide decompositions as in Tables 4–6. M*plus* does provide the option of outputting class-specific R^2's that are analogous to $R_k^{2(f)}$ (when fitting Equations 1 or 2) and $R_{kh}^{2(f)}$ (when fitting Equation 7). Nonetheless, as discussed earlier, just reporting a class-specific R^2 provides an incomplete picture because it can be large even when the total R^2 is small, and vice versa. Additionally, note that, when fitting Equation (2) or (7) in long-format in M*plus* (but not in wide-format), M*plus*' class-specific or class-combination-specific measure is more restrictive than ours (or Latent GOLD's; see below). Specifically, in this situation their measures allow only Level-1 but not Level-2 covariates to contribute to explained variance.[16]

Latent GOLD does not provide quantities analogous to our $R^{2(fv)}$, $R^{2(f)}$, $R_T^{2(fv)}$, $R_T^{2(f)}$, $R_h^{2(fvm)}$, $R_h^{2(fv)}$, $R_h^{2(f)}$, nor does it provide decompositions as in Tables 4–6. Latent GOLD does provide measures analogous to the class-specific R^2's, $R_k^{2(f)}$ and $R_{kh}^{2(f)}$, as well as measures analogous to one kind of total R^2: $R^{2(fvm)}$ and $R_T^{2(fvm)}$. However, a key contrast is that Latent GOLD's total R^2's are computed in a different manner that does not lead to their decomposition into distinct components of explained variance, as we did in Tables 4 and 6 and Figures 5 and 8. Researchers reporting only a single number $R_T^{2(fvm)}$ would not be able to identify where the explained variance is coming from—rather, it would be a blend of the components we have described throughout this article. For instance, without these decompositions, we would not know that the $R^{2(fvm)}$ from empirical example 1 was dominated by the contribution via the marginal component of slopes (see Figure 5) but that the $R_T^{2(fvm)}$ from empirical example 2 was dominated by the contribution via variation in Level-1 class means within Level-2 class (see Figure 8). Moreover, as we mentioned in the introduction, the R^2's implemented in existing software and the R^2's newly developed here had not before been systematically described or interrelated in the literature.

Limitations and Future Directions

Several limitations of this article can be noted to serve as future research directions. First, in our two empirical examples we focused on the utility of our measures with regards to the *direct* interpretation of classes (Titterington et al., 1985), wherein classes are thought to represent distinct subpopulations. Nonetheless, under an *indirect* interpretation (wherein classes are instead thought to approximate underlying continua) our measures—particularly the total R^2's and decompositions thereof—would still provide useful information in terms of variance explained. Future research can demonstrate the application of the R^2 measures in this research context. Second, our empirical examples focused on cross-sectional data. Our measures could also prove useful in longitudinal mixture modeling contexts. Third, for the multilevel regression mixture specifications, we focused on two-level models. The general R^2 framework could in theory be extended to models with any number of levels, though current software options limit what models can be specified. Fourth, we have not discussed predictor-specific contributions to explained variance, such as a semipartial R^2 (see, e.g., Edwards et al., 2008 or Jaeger, Edwards, Das, & Sen, 2016 in a nonmixture modeling context). This would be a useful extension that could help further decompose the explained variance into meaningful sources.

Lastly, for the Equation (7) regression mixture with classes at both levels, we combined a set of five distinct sources of explained variance from Table 6 in creating three total R^2 measures that yielded substantively meaningful interpretations. As an extension, a researcher could combine our decompositions of explained variance from Table 6 in a different way to compute still other effect size measures. For instance, for this regression mixture, one might be interested in the proportion of outcome variance explained by predictors via *exclusively within Level-2 class* slope and mean heterogeneity, or might be interested in the outcome variance explained by predictors via *exclusively between Level-2 class* slope and mean heterogeneity. An applied researcher can use quantities in Table 6 to quickly create such novel measures and, further, could compare their magnitudes. In sum, in future work researchers can combine our decompositions in different ways to cater to their analytic goals.

Conclusions

In light of widespread recommendations to report effect size measures to convey practical significance (e.g., American Psychological Association, 2009; Harlow et al., 1997; Panter & Sterba, 2011; Wilkinson, 1999) we suggest that researchers consult our framework to understand options for quantifying the proportion of variance explained for a given regression mixture. Importantly, researchers are advised to consider all R^2 measures simultaneously, as well as their decompositions, to fully understand what sources of variance are being explained and for whom. It is our hope that these quantitative effect size measures will be substantively useful complements to qualitative interpretations of latent classes.

[16] This restriction is not imposed in our Equation (14) because a Level-2 predictor is not constant for members of the same class-combination (even though by definition it is constant for a given cluster). Members of a given class-combination can belong to different clusters, each having different values of the Level-2 predictor. As such, a Level-2 predictor can explain variance within a class-combination (relatedly, see Rights & Sterba, 2016).

References

American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Bogler, R. (2001). The influence of leadership style on teacher job satisfaction. *Educational Administration Quarterly, 37,* 662–683. http://dx.doi.org/10.1177/00131610121969460

Bowsher, C. G., & Swain, P. S. (2012). Identifying sources of variation and the flow of information in biochemical networks. *Proceedings of the National Academy of Sciences of the United States of America, 109,* E1320–E1328. http://dx.doi.org/10.1073/pnas.1119407109

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., . . . Vitaro, F. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology, 39,* 222–245. http://dx.doi.org/10.1037/0012-1649.39.2.222

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cole, V. T., & Bauer, D. J. (2016). A note on the use of mixture models for individual prediction. *Structural Equation Modeling, 23,* 615–631. http://dx.doi.org/10.1080/10705511.2016.1168266

Cudeck, R., & Henly, S. J. (2003). A realistic perspective on pattern representation in growth data: Comment on Bauer and Curran (2003). *Psychological Methods, 8,* 378–383. http://dx.doi.org/10.1037/1082-989X.8.3.378

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

de Kort, J. M., Dolan, C. V., Lubke, G. H., & Molenaar, D. (in press). Studying the strength of prediction using indirect mixture modeling: Nonlinear latent regression with heteroskedastic residuals. *Structural Equation Modeling*.

DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification, 5,* 249–282. http://dx.doi.org/10.1007/BF01897167

DeSarbo, W. S., Jedidi, K., & Sinha, I. (2001). Customer value analysis in a heterogeneous market. *Strategic Management Journal, 22,* 845–857. http://dx.doi.org/10.1002/smj.191

Ding, C. (2006). Using regression mixture analysis in educational research. *Practical Assessment, Research & Evaluation, 11,* 1–11.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York, NY: Wiley. http://dx.doi.org/10.1002/9781118625590

Dyer, W. J., Pleck, J., & McBride, B. (2012). Using mixture regression to identify varying effects: A demonstration with paternal incarceration. *Journal of Marriage and Family, 74,* 1129–1148. http://dx.doi.org/10.1111/j.1741-3737.2012.01012.x

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R^2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine, 27,* 6137–6157. http://dx.doi.org/10.1002/sim.3429

Fagan, A. A., Van Horn, M. L., Hawkins, J. D., & Jaki, T. (2013). Differential effects of parental controls on adolescent substance use: For whom is the family most important? *Journal of Quantitative Criminology, 29,* 347–368. http://dx.doi.org/10.1007/s10940-012-9183-9

Fairweather, J. (1996). *Faculty work and public trust: Restoring the value of teaching and public service in American academic life*. Boston, MA: Allyn & Bacon.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and hierarchical/multilevel models*. New York, NY: Cambridge University Press.

George, M. R., Yang, N., Jaki, T., Feaster, D. J., Lamont, A. E., Wilson, D. K., & Van Horn, M. L. (2013). Finite mixtures for simultaneously modeling differential effects and nonnormal distributions. *Multivariate Behavioral Research, 48,* 816–844. http://dx.doi.org/10.1080/00273171.2013.830065

Grewal, R., Chandrashekaran, M., Johnson, J. L., & Mallapragada, G. (2013). Environments, unobserved heterogeneity, and the effect of market orientation on outcomes for high-tech firms. *Journal of the Academy of Marketing Science, 41,* 206–233. http://dx.doi.org/10.1007/s11747-011-0295-9

Halliday-Boykins, C. A., Henggeler, S. W., Rowland, M. D., & Delucia, C. (2004). Heterogeneity in youth symptom trajectories following psychiatric crisis: Predictors and placement outcome. *Journal of Consulting and Clinical Psychology, 72,* 993–1003. http://dx.doi.org/10.1037/0022-006X.72.6.993

Harlow, L. L., Muliak, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Ingrassia, S., Minotti, S. C., & Vittadini, G. (2012). Local statistical modeling via a cluster-weighted approach with elliptical distributions. *Journal of Classification, 29,* 363–401. http://dx.doi.org/10.1007/s00357-012-9114-3

Jaeger, B. C., Edwards, L. J., Das, K., & Sen, P. K. (2016). An R2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics*. Advance online publication. http://dx.doi.org/10.1080/02664763.2016.1193725

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

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17,* 137–152. http://dx.doi.org/10.1037/a0028086

Khalili, A., & Chen, J. (2012). Variable selection in finite mixture of regression models. *Journal of the American Statistical Association, 102,* 1025–1038. http://dx.doi.org/10.1198/016214507000000590

King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science, 30,* 666–687. http://dx.doi.org/10.2307/2111095

Kliegel, M., & Zimprich, D. (2005). Predictors of cognitive complaints in older adults: A mixture regression approach. *European Journal of Ageing, 2,* 13–23. http://dx.doi.org/10.1007/s10433-005-0017-6

LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods, 17,* 433–451. http://dx.doi.org/10.1177/1094428114541701

Lamont, A. E., Vermunt, J. K., & Van Horn, M. L. (2016). Regression mixture models: Does modeling the covariance between independent variables and latent classes improve the results? *Multivariate Behavioral Research, 51,* 35–52. http://dx.doi.org/10.1080/00273171.2015.1095063

Langton, N., & Pfeffer, J. (1994). Paying the professor: Sources of salary variation in academic labor markets. *American Sociological Review, 59,* 236–256. http://dx.doi.org/10.2307/2096229

LeBlanc, L., Swisher, R., Vitaro, F., & Tremblay, R. E. (2007). School social climate and teachers' perceptions of classroom behavior problems: A 10 year longitudinal and multilevel study. *Social Psychology of Education, 10,* 429–442. http://dx.doi.org/10.1007/s11218-007-9027-x

Leisch, F. (2004). FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software, 11,* 1–18. http://dx.doi.org/10.18637/jss.v011.i08

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. http://dx.doi.org/10.1111/j.1467-9531.2010.01231.x

Magee, L. (1990). R^2 measures based on Wald and likelihood ratio joint significance tests. *The American Statistician, 44,* 250–253.

Manchia, M., Zai, C. C., Squassina, A., Vincent, J. B., De Luca, V., & Kennedy, J. L. (2010). Mixture regression analysis on age at onset in bipolar disorder patients: Investigation of the role of serotonergic genes. *European Neuropsychopharmacology, 20,* 663–670. http://dx.doi.org/10.1016/j.euroneuro.2010.04.001

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley. http://dx.doi.org/10.1002/0471721182

Montgomery, K. L., Vaughn, M. G., Thompson, S. J., & Howard, M. O. (2013). Heterogeneity in drug abuse among juvenile offenders: Is mixture regression more informative than standard regression? *International Journal of Offender Therapy and Comparative Criminology, 57,* 1326–1346. http://dx.doi.org/10.1177/0306624X12459185

Morin, A. J. S., & Marsh, H. W. (2015). Person-centered analyses: An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling, 22,* 39–59. http://dx.doi.org/10.1080/10705511.2014.919825

Mulvey, E. P., Steinberg, L., Piquero, A. R., Besana, M., Fagan, J., Schubert, C., & Cauffman, E. (2010). Trajectories of desistance and continuity in antisocial behavior following court adjudication among serious adolescent offenders. *Development and Psychopathology, 22,* 453–475. http://dx.doi.org/10.1017/S0954579410000179

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

Muthén, L. K., & Muthén, B. O. (1998–2016). *Mplus user's guide: Version 7.4*. Los Angeles, CA: Author.

Nagin, D. (2005). *Group-based modeling of development*. Boston, MA: Harvard University Press. http://dx.doi.org/10.4159/9780674041318

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4,* 133–142. http://dx.doi.org/10.1111/j.2041-210x.2012.00261.x

National Center for Educational Statistics. (1986). *High school and beyond, 1980: Sophomore cohort second follow-up (1984). Data file user's manual*. Ann Arbor, MI: Interuniversity Consortium for Political and Social Research.

Nowrouzi, B., Souza, R. P., Zai, C., Shinkai, T., Monda, M., Lieberman, J., . . . De Luca, V. (2013). Finite mixture regression model analysis on antipsychotics induced weight gain: Investigation of the role of the serotonergic genes. *European Neuropsychopharmacology, 23,* 224–228. http://dx.doi.org/10.1016/j.euroneuro.2012.05.008

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14,* 535–569. http://dx.doi.org/10.1080/10705510701575396

O'Grady, K. E. (1982). Measures of explained variance: Cautions and limitations. *Psychological Bulletin, 92,* 766–777. http://dx.doi.org/10.1037/0033-2909.92.3.766

Orelien, J. G., & Edwards, L. J. (2008). Fixed-effect variable selection in linear mixed models using R^2 statistics. *Computational Statistics & Data Analysis, 52,* 1896–1907. http://dx.doi.org/10.1016/j.csda.2007.06.006

Panter, A. T., & Sterba, S. K. (2011). *Handbook of ethics in quantitative methodology*. New York, NY: Routledge.

Pearson, L. C., & Moomaw, W. (2005). The relationship between teacher autonomy and stress, work satisfaction, empowerment, and professionalism. *Educational Research Quarterly, 29,* 37–53.

Raftery, A. E., & Dean, N. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association, 101,* 168–178. http://dx.doi.org/10.1198/016214506000000113

R Core Team. (2016). *R: A language and environment for statistical computing*. Retrieved from http://www.R-project.org

Rights, J. D., & Sterba, S. K. (2016). The relationship between multilevel models and non-parametric multilevel mixture models: Discrete approximation of intraclass correlation, random coefficient distributions, and residual heteroscedasticity. *British Journal of Mathematical & Statistical Psychology, 69,* 316–343. http://dx.doi.org/10.1111/bmsp.12073

Schmiege, S. J., Levin, M. E., & Bryan, A. D. (2009). Regression mixture models of alcohol use and risky sexual behavior among criminally-involved adolescents. *Prevention Science, 10,* 335–344. http://dx.doi.org/10.1007/s11121-009-0135-z

Sher, K. J., Jackson, K. M., & Steinley, D. (2011). Alcohol use trajectories and the ubiquitous cat's cradle: Cause for concern? *Journal of Abnormal Psychology, 120,* 322–335. http://dx.doi.org/10.1037/a0021813

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. (2014). Handling missing covariates in conditional mixture models under missing at random assumptions. *Multivariate Behavioral Research, 49,* 614–632. http://dx.doi.org/10.1080/00273171.2014.950719

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. http://dx.doi.org/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. http://dx.doi.org/10.1080/10705511.2014.915189

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 models to identify Level-1 heterogeneity in Level-2 effects. *Structural Equation Modeling, 23,* 259–269. http://dx.doi.org/10.1080/10705511.2015.1035437

Van Horn, M. L., Jaki, T., Masyn, K., Howe, G., Feaster, D. J., Lamont, A. E., . . . Kim, M. (2015). Evaluating differential effects using regression interactions and regression mixture models. *Educational and Psychological Measurement, 75,* 677–714. http://dx.doi.org/10.1177/0013164414554931

Van Horn, M. L., Jaki, T., Masyn, K., Ramey, S. L., Smith, J. A., & Antaramian, S. (2009). Assessing differential effects: Applying regression mixture models to identify variations in the influence of family resources on academic achievement. *Developmental Psychology, 45,* 1298–1313. http://dx.doi.org/10.1037/a0016427

Vermunt, J. K. (2004). An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. *Statistica Neerlandica, 58,* 220–233. http://dx.doi.org/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. http://dx.doi.org/10.1177/0962280207081238

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

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. http://dx.doi.org/10.1007/3-540-28084-7_26

Vermunt, J. K., & Magidson, J. (2016). *Technical guide for Latent GOLD 5.1: Basic, advanced, and syntax*. Belmont, MA: Statistical Innovations Inc.

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

Vonesh, E. F., & Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. New York, NY: Marcel Dekker.

Vonesh, E. F., Chinchilli, V. M., & Pu, K. (1996). Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics, 52,* 572–587. http://dx.doi.org/10.2307/2532896

Wang, J., & Schaalje, G. B. (2009). Model selection for linear mixed models using predictive criteria. *Communications in Statistics Simulation and Computation, 38,* 788–801. http://dx.doi.org/10.1080/03610910802645362

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.
Wedel, M., & Kamakura, W. A. (1998). *Market segmentation: Concepts and methodological foundations*. Boston, MA: Kluwer Academic Publishers.
Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanation. *American Psychologist, 54,* 594–604. http://dx.doi.org/10.1037/0003-066X.54.8.594
Wong, Y. J., & Maffini, C. S. (2011). Predictors of Asian American adolescents' suicide attempts: A latent class regression analysis. *Journal of Youth and Adolescence, 40,* 1453–1464. http://dx.doi.org/10.1007/s10964-011-9701-3
Wong, Y. J., Owen, J., & Shea, M. (2012). A latent class regression analysis of men's conformity to masculine norms and psychological distress. *Journal of Counseling Psychology, 59,* 176–183. http://dx.doi.org/10.1037/a0026206
Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine, 22,* 3527–3541. http://dx.doi.org/10.1002/sim.1572

Appendix A

Derivation of Model-Implied Total Outcome Variance

Appendix A provides the derivation for the denominator expression of article Equations (8), (9), and (10), i.e., the model-implied total variance of y_{ij} using the general mixture model in Equation (7). Slight modifications of this derivation yield the denominator expression in Equations (3), (4), and (5), i.e., the model-implied total variance of y_i using the simpler special case mixture model in Equation (1). Specifically, the below derivation also applies to the Equations (3), (4), and (5) expressions after the following replacements are made: replace all ij subscripts with i, replace all kh superscripts with k, replace all $^{\bullet\bullet}$ superscripts with $^{\bullet}$, and remove $d_j = h$ from all operations.

As defined in the article: i = individual, j = cluster, c_{ij} = Level-1 latent classification variable with classes $k = 1 \ldots K$, d_j = Level-2 latent classification variable with classes $h = 1 \ldots H$, $\mathbf{x}'_{ij}$ = a vector of 1 and exogenous predictors, and $\boldsymbol{\gamma}^{kh}$ = a vector of regression coefficients (intercepts and slopes) specific to class-combination kh. We denote the total model-implied variance of y_{ij} as $\text{var}_{ij}(y_{ij})$. In taking the variance across ij, as in $\text{var}_{ij}(.)$, we are equivalently taking the variance across all kh, (i.e., $\text{var}_{kh}(.)$) because each individual i within cluster j is a member of a class-combination kh. We will use the ij subscript throughout the derivation for simplicity. Relatedly, let $E_{ij}(.)$ denote the expectation across i and j.

To begin, using an extended application of the law of total variance (see, e.g., Bowsher & Swain, 2012's Equation 13) $\text{var}_{ij}(y_{ij})$ can be expressed as Equation (A.1):

$$\begin{aligned}\text{var}_{ij}(y_{ij}) = & E_{ij}\left[\text{var}_{ij}\left(y_{ij}\,|\,\mathbf{x}'_{ij}, (c_{ij}=k, d_j=h)\right)\right] \\ & + E_{ij}\left[\text{var}_{ij}\left(E_{ij}\left[y_{ij}\,|\,\mathbf{x}'_{ij}, (c_{ij}=k, d_j=h)\right]\,|\,\mathbf{x}'_{ij}\right)\right] \\ & + \text{var}_{ij}\left(E_{ij}\left[y_{ij}\,|\,\mathbf{x}'_{ij}\right]\right)\end{aligned} \tag{A.1}$$

First we will show how Equation (A.1) can be written as a function of quantities in the data model of manuscript Equation (7), yielding Equation (A.2):

$$\begin{aligned}\text{var}_{ij}(y_{ij}) = & E_{ij}\left[\mathbf{x}'_{ij}E_{ij}[(\boldsymbol{\gamma}^{kh} - \boldsymbol{\gamma}^{\bullet\bullet})(\boldsymbol{\gamma}^{kh} - \boldsymbol{\gamma}^{\bullet\bullet})']\mathbf{x}_{ij}\right] + E_{ij}\left[\left(\mathbf{x}'_{ij}\boldsymbol{\gamma}^{\bullet\bullet}\right)^2\right] \\ & - E_{ij}\left[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{\bullet\bullet}\right]^2 + \theta^{\bullet\bullet}\end{aligned} \tag{A.2}$$

Subsequently we will show how Equation (A.2) can be re-expressed as the denominator of Equations (8)–(10), i.e., Equation (A.3).

$$\text{var}_{ij}(y_{ij}) = \mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\,\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet} + \theta^{\bullet\bullet}. \tag{A.3}$$

Steps involved in re-expressing Equation (A.1) as Equation (A.2) are as follows:

The first term, $E_{ij}[\text{var}_{ij}(y_{ij}\,|\,\mathbf{x}'_{ij}, (c_{ij} = k, d_j = h))]$, is simply the expected value of the residual variance.

$$\begin{aligned}& E_{ij}\left[\text{var}_{ij}\left(y_{ij}\,|\,\mathbf{x}'_{ij}, (c_{ij}=k, d_j=h)\right)\right] \\ & = E_{ij}[\theta^{kh}] \\ & = \theta^{\bullet\bullet}\end{aligned}$$

The second term, $E_{ij}[\text{var}_{ij}\ (E_{ij}[y_{ij}\,|\,\mathbf{x}'_{ij}, (c_{ij} = k, d_j = h)]\,|\,\mathbf{x}'_{ij})]$, simplifies as follows:

$$\begin{aligned}& E_{ij}[\text{var}_{ij}(E_{ij}[y_{ij}\,|\,\mathbf{x}'_{ij}, (c_{ij}=k, d_j=h)]\,|\,\mathbf{x}'_{ij})] \\ & = E_{ij}[\text{var}_{ij}(\mathbf{x}'_{ij}\boldsymbol{\gamma}^{kh}\,|\,\mathbf{x}'_{ij})] \\ & = E_{ij}[\mathbf{x}'_{ij}\text{var}_{ij}(\boldsymbol{\gamma}^{kh})\mathbf{x}_{ij}] \\ & = E_{ij}[\mathbf{x}'_{ij}E_{ij}[(\boldsymbol{\gamma}^{kh} - \boldsymbol{\gamma}^{\bullet\bullet})(\boldsymbol{\gamma}^{kh} - \boldsymbol{\gamma}^{\bullet\bullet})']\mathbf{x}_{ij}]\end{aligned}$$

(*Appendices continue*)

The third term, $\text{var}_{ij}(E_{ij}[y_{ij} \mid \mathbf{x}'_{ij}])$, simplifies as follows:

$$\begin{aligned}
\text{var}_{ij}(E_{ij}[y_{ij} \mid \mathbf{x}'_{ij}]) &= \text{var}_{ij}(E_{ij}[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{kh} + \varepsilon_{ij} \mid \mathbf{x}'_{ij}]) \\
&= \text{var}_{ij}(E_{ij}[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{kh} \mid \mathbf{x}'_{ij}] + \cancel{E_{ij}[\varepsilon_{ij} \mid \mathbf{x}'_{ij}]}) \\
&= \text{var}_{ij}(E_{ij}[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{kh} \mid \mathbf{x}'_{ij}]) \\
&= \text{var}_{ij}(\mathbf{x}'_{ij}E_{ij}[\boldsymbol{\gamma}^{kh}]) \\
&= \text{var}_{ij}(\mathbf{x}'_{ij}\boldsymbol{\gamma}^{\bullet\bullet}) \\
&= E_{ij}[(\mathbf{x}'_{ij}\boldsymbol{\gamma}^{\bullet\bullet})^2] - E_{ij}[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{\bullet\bullet}]^2
\end{aligned}$$

Note that, above, the exogenous $\mathbf{x}'_{ij}$ is a constant with respect to the conditional expectation but not with respect to the variance. Thus, the total variance of y_{ij} can be expressed as Equation (A.2), i.e.:

$$\text{var}_{ij}(y_{ij}) = E_{ij}[\mathbf{x}'_{ij}E_{ij}[(\boldsymbol{\gamma}^{kh} - \boldsymbol{\gamma}^{\bullet\bullet})(\boldsymbol{\gamma}^{kh} - \boldsymbol{\gamma}^{\bullet\bullet})']\mathbf{x}_{ij}] + E_{ij}[(\mathbf{x}'_{ij}\boldsymbol{\gamma}^{\bullet\bullet})^2] - E_{ij}[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{\bullet\bullet}]^2 + \theta^{\bullet\bullet}$$

In the remainder of Appendix A we show how this Equation (A.2) is equal to the denominator of Equations (8)–(10) (i.e., $\mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet} + \theta^{\bullet\bullet}$ from Equation A.3 above).

We first show that $E_{ij}[(\mathbf{x}'_{ij}\boldsymbol{\gamma}^{\bullet\bullet})^2] - E_{ij}[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{\bullet\bullet}]^2 = \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}$, as follows:

$$\begin{aligned}
E_{ij}[(\mathbf{x}'_{ij}\boldsymbol{\gamma}^{\bullet\bullet})^2] - E_{ij}[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{\bullet\bullet}]^2 &= \text{var}_{ij}(\mathbf{x}'_{ij}\boldsymbol{\gamma}^{\bullet\bullet}) \\
&= \text{var}_{ij}(\boldsymbol{\gamma}^{\bullet\bullet\prime}\,\mathbf{x}_{ij}) \\
&= \boldsymbol{\gamma}^{\bullet\bullet\prime}\,\text{var}_{ij}(\mathbf{x}_{ij})\boldsymbol{\gamma}^{\bullet\bullet} \\
&= \boldsymbol{\gamma}^{\bullet\bullet\prime}\,\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}
\end{aligned}$$

We next show that $E_{ij}[\mathbf{x}'_{ij}E_{ij}[(\boldsymbol{\gamma}^{kh} - \boldsymbol{\gamma}^{\bullet\bullet})(\boldsymbol{\gamma}^{kh} - \boldsymbol{\gamma}^{\bullet\bullet})']\mathbf{x}_{ij}] = \mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa}$ as follows:

$$\begin{aligned}
E_{ij}[\mathbf{x}'_{ij}E_{ij}[(\boldsymbol{\gamma}^{kh} - \boldsymbol{\gamma}^{\bullet\bullet})(\boldsymbol{\gamma}^{kh} - \boldsymbol{\gamma}^{\bullet\bullet})']\mathbf{x}_{ij}] &= E_{ij}[\mathbf{x}'_{ij}\text{var}_{ij}(\boldsymbol{\gamma}^{kh})\mathbf{x}_{ij}] \\
&= E_{ij}[\mathbf{x}'_{ij}\underbrace{(\mathbf{V} + \mathbf{C})}_{\substack{\mathbf{V} = \text{diag matrix of } \gamma^{kh} \text{ variances} \\ \mathbf{C} = \text{cov matrix of } \gamma^{kh}\text{; 0s on diag}}}\mathbf{x}_{ij}] \\
&= E_{ij}[\mathbf{x}'_{ij}\mathbf{V}\mathbf{x}_{ij} + \mathbf{x}'_{ij}\mathbf{C}\mathbf{x}_{ij}] \\
&= \underbrace{E_{ij}[\mathbf{x}'_{ij}\mathbf{V}\mathbf{x}_{ij}]}_{\text{see below: \#1}} + \underbrace{E_{ij}[\mathbf{x}'_{ij}\mathbf{C}\mathbf{x}_{ij}]}_{\text{see below: \#2}}
\end{aligned}$$

#1

$$\begin{aligned}
E_{ij}[\mathbf{x}'_{ij}\mathbf{V}\mathbf{x}_{ij}] &= E_{ij}[tr(\mathbf{x}'_{ij}\mathbf{V}\mathbf{x}_{ij})] \\
&= E_{ij}[tr(\mathbf{x}_{ij}\mathbf{x}'_{ij}\mathbf{V})] \\
&= E_{ij}[tr((\mathbf{x}_{ij}\mathbf{x}'_{ij})_D\mathbf{V} + \cancel{(\mathbf{x}_{ij}\mathbf{x}'_{ij})_{OD}\mathbf{V}})]
\end{aligned}$$

Where $\mathbf{x}_{ij}\mathbf{x}'_{ij}$ is equal to its diagonal elements, $(\mathbf{x}_{ij}\mathbf{x}'_{ij})_D$, plus its off diagonal elements, $(\mathbf{x}_{ij}\mathbf{x}'_{ij})_{OD}$. Note that the term $(\mathbf{x}_{ij}\mathbf{x}'_{ij})_{OD}\,\mathbf{V}$ cancels because $\mathbf{V}$ is diagonal.

$$\begin{aligned}
&= E_{ij}[tr((\mathbf{x}_{ij}\mathbf{x}'_{ij})_D\mathbf{V})] \\
&= tr(E_{ij}[(\mathbf{x}_{ij}\mathbf{x}'_{ij})_D]\mathbf{V}) \\
&= \sum_t [E_{ij}[(\mathbf{x}_{ij}\mathbf{x}'_{ij})_D]]_{tt}[\mathbf{V}]_{tt} \\
&= \sum_t [\mathbf{s}']_t[\boldsymbol{\psi}]_t \\
&= \mathbf{s}'\boldsymbol{\psi}
\end{aligned}$$

That is, because both $E_{ij}[(\mathbf{x}_{ij}\mathbf{x}'_{ij})_D]$ and $\mathbf{V}$ are diagonal, the sum of products of diagonal elements is the same as the inner product of two vectors. Thus, $\mathbf{s}'$ is a row vector of means of squared elements of $\mathbf{x}_{ij}$ and $\boldsymbol{\psi}$ is a column vector containing diagonal elements of $\mathbf{V}$.

#2

$$\begin{aligned}
E_{ij}[\mathbf{x}'_{ij}\mathbf{C}\mathbf{x}_{ij}] &= E_{ij}[tr(\mathbf{x}'_{ij}\mathbf{C}\mathbf{x}_{ij})] \\
&= E_{ij}[tr(\mathbf{x}_{ij}\mathbf{x}'_{ij}\mathbf{C})] \\
&= E_{ij}[tr(\cancel{(\mathbf{x}_{ij}\mathbf{x}'_{ij})_D\mathbf{C}} + (\mathbf{x}_{ij}\mathbf{x}'_{ij})_{OD}\mathbf{C})]
\end{aligned}$$

Note that $(\mathbf{x}_{ij}\mathbf{x}'_{ij})_D\mathbf{C}$ cancels because $\mathbf{C}$ has 0's on diagonal.

$$\begin{aligned}
&= E_{ij}[tr((\mathbf{x}_{ij}\mathbf{x}'_{ij})_{OD}\mathbf{C})] \\
&= tr(E_{ij}[(\mathbf{x}_{ij}\mathbf{x}'_{ij})_{OD}\mathbf{C}]) \\
&= \sum_{t,u}[E_{ij}[(\mathbf{x}_{ij}\mathbf{x}'_{ij})_{OD}]]_{tu,t\neq u}[\mathbf{C}]_{ut,t\neq u}
\end{aligned}$$

Because the diagonals of $\mathbf{C}$ are 0, it follows that only off-diagonal elements of $\mathbf{x}_{ij}\mathbf{x}'_{ij}$ and $\mathbf{C}$ need to be considered (i.e., $u \neq t$).

$$= \sum_{t,u}[E_{ij}[(\mathbf{x}_{ij}\mathbf{x}'_{ij})_{OD}]]_{tu,t\neq u}[\mathbf{C}]_{tu,t\neq u}$$

Because $\mathbf{C}$ is symmetric, it follows that $[\mathbf{C}]_{ut} = [\mathbf{C}]_{tu}$.

$$= 2\sum_{t,u}[E_{ij}[(\mathbf{x}_{ij}\mathbf{x}'_{ij})_{OD}]]_{tu,u<t}[\mathbf{C}]_{tu,u<t}$$

Because $\mathbf{x}_{ij}\mathbf{x}'_{ij}$ and $\mathbf{C}$ are both symmetric, only the lower-triangular elements of $\mathbf{x}_{ij}\mathbf{x}'_{ij}$ and $\mathbf{C}$ need to be considered ($u < t$), twice.

$$\begin{aligned}
&= 2\sum_g [\mathbf{p}']_g[\boldsymbol{\kappa}]_g \\
&= 2\mathbf{p}'\boldsymbol{\kappa}
\end{aligned}$$

That is, we are renaming the typical lower-triangular elements of matrix $E_{ij}[(\mathbf{x}_{ij}\mathbf{x}'_{ij})_{OD}]$ and of matrix $\mathbf{C}$ to be typical elements of vector $\mathbf{p}'$ and of vector $\boldsymbol{\kappa}$, respectively. Thus, $\mathbf{p}'$ is a row vector of means for the pairwise products of all nonredundant elements of $\mathbf{x}_{ij}$ and $\boldsymbol{\kappa}$ is a column vector of lower-triangular elements of $\mathbf{C}$.

(*Appendices continue*)

Appendix B

Derivation of Equation 9 Numerator

Appendix B provides the derivation for the numerator of article Equation (9), i.e., $\mathbf{v}'\boldsymbol{\psi} + 2\mathbf{r}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}$. Slight modifications of this derivation yield the numerator of article Equation (4), as a special case. Specifically, the below derivation also applies to Equation (4) after the following replacements are made: replace all ij subscripts with i, replace all kh superscripts with k, replace all $^{\bullet\bullet}$ superscripts with $^{\bullet}$, and remove $d_j = h$ from all operations.

Please see Appendix A for definitions of these terms: i, j, c_{ij}, k, d_j, h, $\mathbf{x}'_{ij}$, $\boldsymbol{\gamma}^{kh}$, $\mathbf{s}'\boldsymbol{\psi}$, $2\mathbf{p}'\boldsymbol{\kappa}$, $\boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}$, and $\theta^{\bullet\bullet}$. Note that $\text{var}_{ij}(.)$ indicates taking the variance across ij, or equivalently across all kh, (i.e., $\text{var}_{kh}(.)$) because each individual i within cluster j is a member of a class-combination kh. We will use the ij subscript throughout the derivation for simplicity. Relatedly, let $E_{ij}(.)$ denote the expectation across i and j.

The numerator of $R_T^{2(fv)}$ (from Equation 9), is here newly denoted $\text{var}_{ij}(\hat{y}_T^{fv})$. As stated in the manuscript, it involves subtracting the variance of the model-implied class-combination means of y_{ij} from the explained portion of variance from $R_T^{2(fvm)}$.[17]

$$\text{var}_{ij}(\hat{y}_T^{fv}) = \mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet} - \text{var}_{ij}\left(E_{ij}\left[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{kh} + \varepsilon_{ij} \mid (c_{ij} = k, d_j = h)\right]\right) \tag{B.1}$$

First we will show that equation (B.1) can be expressed as equation (B.2):

$$\text{var}_{ij}(\hat{y}_T^{fv}) = \mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet} - (\mathbf{a}'\boldsymbol{\psi} + 2\mathbf{q}'\boldsymbol{\kappa}) \tag{B.2}$$

Then we will show that Equation (B.2) can be re-expressed as Equation (B.3), which is the numerator of article Equation (9).

$$\text{var}_{ij}(\hat{y}_T^{fv}) = \mathbf{v}'\boldsymbol{\psi} + 2\mathbf{r}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet} \tag{B.3}$$

To begin, we show how $\text{var}_{ij}(E_{ij}[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{kh} + \varepsilon_{ij} \mid (c_{ij} = k, d_j = h)])$ from Equation (B.1) $= \mathbf{a}'\boldsymbol{\psi} + 2\mathbf{q}'\boldsymbol{\kappa}$ from Equation (B.2). The former can be simplified as follows:

$$\begin{aligned}
&\text{var}_{ij}(E_{ij}[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{kh} + \varepsilon_{ij} \mid (c_{ij} = k, d_j = h)]) \\
&\quad = \text{var}_{ij}\left(E_{ij}\left[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{kh} \mid (c_{ij} = k, d_j = h)\right] + \cancel{E_{ij}[\varepsilon_{ij} \mid (c_{ij} = k, d_j = h)]}\right) \\
&\quad = \text{var}_{ij}\left(E_{ij}\left[\mathbf{x}'_{ij}\boldsymbol{\gamma}^{kh} \mid (c_{ij} = k, d_j = h)\right]\right) \\
&\quad = \text{var}_{ij}\left(E_{ij}\left[\mathbf{x}'_{ij}\right]\boldsymbol{\gamma}^{kh}\right)
\end{aligned}$$

Note that, above, $\boldsymbol{\gamma}^{kh}$ is a constant with respect to the conditional expectation but not with respect to the variance. Now designate $\mathbf{m}$ as a vector of means of the exogenous $\mathbf{x}_{ij}$, i.e., $\mathbf{m} = E_{ij}[\mathbf{x}'_{ij}]$.

$$\begin{aligned}
&= \mathbf{m}'\,\text{var}_{ij}(\boldsymbol{\gamma}^{kh})\mathbf{m} \\
&= \mathbf{m}'\underbrace{(\mathbf{V} + \mathbf{C})}_{\substack{\mathbf{V} = \text{diag matrix of } \boldsymbol{\gamma}^{kh} \text{ variances} \\ \mathbf{C} = \text{cov matrix of } \boldsymbol{\gamma}^{kh}\text{; 0s on diag}}}\mathbf{m} \\
&= \underbrace{\mathbf{m}'\mathbf{V}\mathbf{m}}_{\text{see below: \#1}} + \underbrace{\mathbf{m}'\mathbf{C}\mathbf{m}}_{\text{see below: \#2}}
\end{aligned}$$

#1

$$\begin{aligned}
\mathbf{m}'\mathbf{V}\mathbf{m} &= tr(\mathbf{m}'\mathbf{V}\mathbf{m}) \\
&= tr(\mathbf{m}\mathbf{m}'\mathbf{V}) \\
&= tr((\mathbf{m}\mathbf{m}')_D\mathbf{V} + \cancel{(\mathbf{m}\mathbf{m}')_{OD}\mathbf{V}})
\end{aligned}$$

Where $\mathbf{mm}'$ is equal to its diagonal elements, $(\mathbf{mm}')_D$, plus its off diagonal elements, $(\mathbf{mm}')_{OD}$. Note that the term $(\mathbf{mm}')_{OD}\,\mathbf{V}$ cancels because $\mathbf{V}$ is diagonal.

$$\begin{aligned}
&= tr((\mathbf{m}\mathbf{m}')_D\mathbf{V}) \\
&= \sum_t [(\mathbf{m}\mathbf{m}')_D]_{tt}[\mathbf{V}]_{tt} \\
&= \sum_t [\mathbf{a}']_t[\boldsymbol{\psi}]_t \\
&= \mathbf{a}'\boldsymbol{\psi}
\end{aligned}$$

That is, because both $(\mathbf{mm}')_D$ and $\mathbf{V}$ are diagonal, the sum of products of diagonal elements is the same as the inner product of two vectors. Thus, $\mathbf{a}'$ is a row vector of squared means of $\mathbf{x}_{ij}$ and $\boldsymbol{\psi}$ is a column vector containing diagonal elements of $\mathbf{V}$.

#2

$$\begin{aligned}
\mathbf{m}'\mathbf{C}\mathbf{m} &= tr(\mathbf{m}'\mathbf{C}\mathbf{m}) \\
&= tr(\mathbf{m}\mathbf{m}'\mathbf{C}) \\
&= tr(\cancel{(\mathbf{m}\mathbf{m}')_D\mathbf{C}} + (\mathbf{m}\mathbf{m}')_{OD}\mathbf{C})
\end{aligned}$$

Note that $(\mathbf{mm}')_D\,\mathbf{C}$ cancels because $\mathbf{C}$ has 0's on diagonal.

$$\begin{aligned}
&= tr((\mathbf{m}\mathbf{m}')_{OD}\mathbf{C}) \\
&= \sum_{t,u} [(\mathbf{m}\mathbf{m}')_{OD}]_{tu,t\neq u}[\mathbf{C}]_{ut,t\neq u}
\end{aligned}$$

[17] This is analytically derived but is akin to literally fitting a null model with the same number of classes and with intercepts fixed to model-implied means from the full model, and subtracting its explained variance from the explained portion of variance from $R_T^{2(fvm)}$ (i.e., Equation 8).

(*Appendices continue*)

Because the diagonals of **C** are 0, it follows that only off-diagonal elements of $\mathbf{mm}'$ and **C** need to be considered (i.e., $u \neq t$).

$$= \sum_{t,u} [(\mathbf{mm}')_{OD}]_{tu,t\neq u}[\mathbf{C}]_{tu,t\neq u}$$

Because **C** is symmetric, it follows that $[\mathbf{C}]_{ut} = [\mathbf{C}]_{tu}$.

$$=2\sum_{t,u} [(\mathbf{mm}')_{OD}]_{tu,u<t}[\mathbf{C}]_{tu,u<t}$$

Because $\mathbf{mm}'$ and **C** are both symmetric, only the lower-triangular elements of $\mathbf{mm}'$ and **C** need to be considered ($u <$ t), twice.

$$= 2\sum_{g} [\mathbf{q}']_g[\boldsymbol{\kappa}]_g$$
$$= 2\mathbf{q}'\boldsymbol{\kappa}$$

That is, we are renaming the typical lower-triangular elements of matrix $(\mathbf{mm}')_{OD}$ and of matrix **C** to be typical elements of vector $\mathbf{q}'$ and of vector $\boldsymbol{\kappa}$, respectively. Thus, $\mathbf{q}'$ is a row vector of the pairwise products of means of all nonredundant elements of $\mathbf{x}_{ij}$ and $\boldsymbol{\kappa}$ is a column vector of lower-triangular elements of **C**.

Next we show how Equation (B.2) can be re-expressed as Equation (B.3), which is the numerator of article Equation (9). Rearranging terms yields:

$$\begin{aligned}\mathrm{var}_{ij}(\hat{y}_T^{fv}) &= \mathbf{s}'\boldsymbol{\psi} + 2\mathbf{p}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet} - (\mathbf{a}'\boldsymbol{\psi} + 2\mathbf{q}'\boldsymbol{\kappa})\\ &= (\mathbf{s}' - \mathbf{a}')\boldsymbol{\psi} + 2(\mathbf{p}' - \mathbf{q}')\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}\end{aligned}$$

Because, as mentioned above,

s = vector of means (across *i* and *j*) of squared elements of $\mathbf{x}_{ij}$

a = vector of squared means (across *i* and *j*) of $\mathbf{x}_{ij}$

p = vector of means (across *i* and *j*) of the pairwise products of all nonredundant elements of $\mathbf{x}_{ij}$

q = vector of pairwise products of means (across *i* and *j*) of all nonredundant elements of $\mathbf{x}_{ij}$

We can substitute

$\mathbf{v}' = (\mathbf{s}' - \mathbf{a}')$ which are variances (across *i* and *j*) of $\mathbf{x}_{ij}$

$\mathbf{r}' = (\mathbf{p}' - \mathbf{q}')$ which are covariances (across *i* and *j*) of all nonredundant elements of $\mathbf{x}_{ij}$

thus yielding the numerator of article Equation (9):

$$\mathrm{var}_{ij}(\hat{y}_T^{fv}) = \mathbf{v}'\boldsymbol{\psi} + 2\mathbf{r}'\boldsymbol{\kappa} + \boldsymbol{\gamma}^{\bullet\bullet\prime}\boldsymbol{\Phi}\boldsymbol{\gamma}^{\bullet\bullet}.$$

Received June 28, 2016
Revision received November 22, 2016
Accepted December 18, 2016 ■