Published in final edited form as:

J Stat Comput Simul. 2013 January 1; 83(4): . doi:10.1080/00949655.2011.636363.

Using regression mixture models with non-normal data: Examining an ordered polytomous approach

Melissa R. W. George^a, Na Yang^b, M. Lee Van Horn^a, Jessalyn Smith^c, Thomas Jaki^d, Dan Feaster^e, Katherine Masyn^f, and George Howe^g

^aDepartment of Psychology, University of South Carolina, Columbia, South Carolina, USA

^bAdvanceMed Corporation, Nashville, TN, USA

^cPsychometric Services, CTB/McGraw Hill, Monterey, CA, USA

^dDepartment of Mathematics and Statistics, Lancaster University, Lancaster, UK

^eDepartment of Epidemiology and Public Health, University of Miami, Coral Gables, FL, USA

^fGraduate School of Education, Harvard University, Larsen 409, Cambridge, MA, USA

^gDepartment of Psychology, George Washington University, Washington DC, USA

Abstract

Mild to moderate skew in errors can substantially impact regression mixture model results; one approach for overcoming this includes transforming the outcome into an ordered categorical variable and using a polytomous regression mixture model. This is effective for retaining differential effects in the population; however, bias in parameter estimates and model fit warrant further examination of this approach at higher levels of skew. The current study used Monte Carlo simulations; three thousand observations were drawn from each of two subpopulations differing in the effect of X on Y. Five hundred simulations were performed in each of the ten scenarios varying in levels of skew in one or both classes. Model comparison criteria supported the accurate two class model, preserving the differential effects, while parameter estimates were notably biased. The appropriate number of effects can be captured with this approach but we suggest caution when interpreting the magnitude of the effects.

Keywords

Regression mixture models; non-normal errors; differential effects

1. Introduction

Regression mixture models are a class of finite mixture model in which heterogeneity in the effects of covariates on an outcome is allowed across latent classes [1–4]. While they have been shown to perform well in under ideal conditions, regression mixture models rely on strong assumptions about the normality of within-class distributions that impact the results drastically. When subpopulations exist the presence of even mildly skewed errors may significantly impact both the latent class enumeration and parameter estimates [5]. One approach to dealing with violating these distributional assumptions is to categorize the outcome variable and use an ordered polytomous (categorical) regression model which has relaxed (less rigorous) assumptions. While previous research examining this categorization

approach has shown promise for identifying differences in the effects of a predictor on an outcome in the presence of relatively mild to moderate skew in errors [5], there was evidence of some bias in parameter estimates and model fit leaving the utility of this approach in conditions with more skewed errors unknown. Extending the previous work that investigated the sensitivity of an ordered categorical regression mixture model in conditions that have non-normal distributions of error variance, the current study examines the impact of highly skewed within-class residuals on parameter estimates and model fit. The study utilizes simulations to examine the performance of this categorical transformation approach to deal with violations of distributional assumptions of regression mixture models when within-class residuals are highly skewed.

This paper is organized in seven additional sections. In Section 2, we introduce the statistical model for regression mixtures; we discuss the assumption of normally distributed errors and the ordered polytomous approach for handling violating this assumption in Section 3. Section 4 contains the aims of the current paper. In Section 5, we discuss the method and details of the simulation settings. Section 6 reports the results of the simulations which examine the usefulness of the approach for managing violating assumptions of within-class non-normal errors. We end with a summary of our findings and give some conclusions in Section 7.

2. Statistical Model

Regression mixture models have been used primarily in marketing research [6] and only recently have been applied in the behavioural sciences [2, 7–9] as a tool for capturing unobserved heterogeneity in the effects of predictors on outcomes. These models provide a potential solution for advancing theory about differential effects of continuous predictors on individuals, consistent with many conceptual models that expect individual or group differences within the population. Many theories in the social sciences provide rationale for expecting key effects to differ amongst subpopulations and thus methods for examining heterogeneity are needed. The typical approach for modelling heterogeneous effects is the use of interaction terms which account for group differences using observed variables to group individuals (e.g., gender, age, or ethnicity). An advantage of regression mixtures over using interactions is that they can also capture unobserved heterogeneity in effects. Moreover, many times when differential effects exist they are likely to be a function of multiple moderating variables. Modelling these types of differential effects with interactions requires many, often highly correlated, interaction terms, which can result in a failure to find differential effects [10] that are a function of complex processes [11]. The presence of a large number of interactions may also yield results that are difficult to interpret [12].

Regression mixtures are a type of finite mixture model [13] which examine unobserved heterogeneity in the population's outcome. Mixture modelling assumes that the population consists of multiple subpopulations each with a unique distribution of the outcomes; latent classes are used to identify these subpopulations and their distribution on the outcome variables. For the general form of a multivariate mixture model of observed continuous variables, consider a sample of n individuals measured on a set of m continuous variables, $\mathbf{Y}=(Y_1, Y_2, \dots, Y_m)$ where y_{im} is the observed value on variable Y_m for subject i (although this paper focuses on the univariate case, for the purpose of generalization of this model we consider the multivariate outcome). The multivariate probability density function of $\mathbf{Y}$, $f(\mathbf{y}, \phi)$, is modelled as a mixture, or weighted sum, of a finite number of probability densities, $f_k(\mathbf{y}, \theta_k)$. These correspond to the variable distributions for K subgroups, with group membership represented by a latent categorical variable, C , where $C = 1, 2, \dots, K$. The value of K is specified a priori, but the mixing weights, or class proportions, $\pi_1, \pi_2, \dots, \pi_K$, are

included in the set of model parameters to be estimated. The unconditional probability density function of $\mathbf{Y}$ is expressed by

$$f(\mathbf{y}, \phi) = \sum_{k=1}^K \pi_k f_k(\mathbf{y}_k, \boldsymbol{\theta}_k), \quad (1)$$

where $\phi = (\pi, \Theta)$ denotes the vector of all unknown parameters to be estimated; $\pi = (\pi_1, \pi_2, \dots, \pi_K)$ which represents the probability of membership in class k ; and $\Theta = (\theta_1, \theta_2, \dots, \theta_K)$ consists of a set of estimated parameters which describe the probability density for each of the k classes (the specific parameters will depend on the probability density function chosen). To clarify, this is the model in its most general form and additional specific formulations will be discussed in more detail below. If we assume that each subgroup has a multivariate normal distribution, then the outcome, $\mathbf{Y}$, conditional on membership in latent class k , can be expressed as

$$\begin{aligned} \mathbf{Y}_{i|k} &= \boldsymbol{\mu}_k + \boldsymbol{\varepsilon}_{ik}, \\ \boldsymbol{\varepsilon}_{ik} &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_k), \end{aligned} \quad (2)$$

Where $\boldsymbol{\mu}_k$ is the vector of means, $\boldsymbol{\Sigma}_k$ is the variance/covariance matrix for class k , and $\theta_k = (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$.

This model is useful for assessing heterogeneity in outcomes in the population, but does not explicitly model heterogeneity in the *effects* of predictor variables on outcomes. A common assumption of latent class analyses is that the elements of $\mathbf{Y}$ are conditionally independent given class membership, thus there are not differences in the relationship of any outcomes with each other between classes. While this assumption may be relaxed, this approach still does not allow for directional relationships among variables. Regression mixture models, however, do allow for modelling these differential effects in the population by grouping respondents based on having a similar pattern of relationships between two or more variables; therefore allowing for determining differential effects in the population models. For this, we formulate a *regression mixture* model in which the distribution of outcomes is *conditional* on a set of predictor variables where the relationship between predictors and outcomes is allowed to vary between classes. Equation 2 can be extended to include predictor variables, the outcome, $\mathbf{Y}$, conditional on membership in latent class k and on a set of P observed predictors, $\mathbf{X} = (X_1, X_2, \dots, X_P)$, expressed as

$$\begin{aligned} \mathbf{Y}_{i|X,k} &= \boldsymbol{\beta}_{0k} + \sum_{p=1}^P \boldsymbol{\beta}_{pk} x_{ip} + \boldsymbol{\varepsilon}_{ik}, \\ \boldsymbol{\varepsilon}_{ik} &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_k), \end{aligned} \quad (3)$$

Where $\boldsymbol{\beta}_{0k}$ is the vector of intercepts, $\boldsymbol{\Sigma}_k$ is the residual variance/covariance matrix for class k , and $\boldsymbol{\beta}_{pk}$ is the vector of regression coefficients for X_p in latent class k . This formulation allows the effect of the predictor on the outcome to be different for each subgroup within the population. Regression mixtures explicitly allow for directly examining differential effects in the population, potentially yielding information about unobserved groups without a priori identification.

3. Limitations of Distributional Assumptions

In finite mixture models within-class distributional assumptions are a key component in the estimation of model parameters [14, 15]. A purpose of finite mixtures is to approximate the distribution of an outcome using a finite number of classes which each follow some, typically parametric, distribution of the outcome. The model makes no parametric assumptions about the mixed distribution of the outcome, but the distribution of each mixture component (each latent class) is assumed to follow a specific form which is chosen by the analyst. The most common distributional form for each of the classes is normal distribution, when a conditional model is specified the assumption is typically that the outcome is normally distributed after conditioning on covariates. With this specification results may therefore represent an approximation of a non-normal distribution or they may represent the presence of subpopulations which each follow a particular parametric distribution. Because the within class distributional assumption is an important part of parameter estimation (equation 3) the presence of even mildly skewed error terms may dramatically impact model results, affecting latent class enumeration (typically additional classes are needed to model the non-normality) and parameter estimates [14, 16]. When using regression mixtures to find differential effects, the presence of non-normal errors has been shown to have dramatic effects on the results of the models [5].

One approach that has shown promise for addressing this issue of non-normality in errors has been to transform the outcome variable to an ordered categorical variable; then a polytomous regression model which does not make any assumptions about error distributions can be used. In the case of a single categorical outcome variable, Y , that has S different levels, the variable can be assumed to be an ordered response, which requires an additional assumption, such as the proportional odds assumption. This additional assumption specifies that one β is sufficient for describing the relationship of each predictor on the outcome across all levels of s [17, 18]. The model is thus similar to a multinomial logistic regression where intercepts are estimated for each category but with the constraint that the regression weight for the effects of predictors on the outcome is constrained to be equal across all outcomes. In this situation we can take $f_k(y | \theta_k, \mathbf{X})$ to be the probability density function of a multinomial distribution and model the cumulative probability of level s in class k , θ_s^k , as

$$\theta_s^k = P(Y_i \geq s | k, \mathbf{X}) = \frac{e^{\beta_{0,s}^k + \sum_{p=1}^P \beta_p^k X_{ip}}}{\sum_{u=1}^S e^{\beta_{0,u}^k + \sum_{p=1}^P \beta_p^k X_{ip}}} \quad s=1, \dots, S \quad (4)$$

or alternatively in the linear model form as

$$\text{logit}(\theta_s^k) = \text{logit}(P(Y_i \geq s | k, \mathbf{X})) = e^{\beta_{0,s}^k + \sum_{p=1}^P \beta_p^k X_{ip}} \quad s=1, \dots, S \quad (5)$$

Note that model identification requires constraints on the parameters. Here the last class is used as a reference group to estimate the thresholds β_{0s}^k for classes 1 through $S-1$. No specific distributional assumption is made; however, the model does assume that θ_s^k is independent conditional on $\mathbf{X}$.

Previous research has indicated that regression mixture models are sensitive to non-normal errors, even mild skew, especially in both classes, results in misinterpretation of the simulated differential effects in the population [5]. Specifically, under simulated conditions

of mild to moderate non-normality in residual distributions, ranging from .15 to .49, in one or both classes, results showed that in the presence of mild skew in even one of the classes yielded biased estimates particularly in the intercepts of the latent classes; as skew increased results were not consistent across simulations. Examination of the ordered polytomous approach (the proportional odds model) indicated that the model performed very well at recovering the correct number of differential effects; however, some bias in parameter estimates was observed. Further examination of this approach for dealing with violating distributional assumptions in the presence of greater skew in errors is needed because of the previous finding of some bias in parameter estimates at moderate levels of non-normality.

4. Study Aims

The primary aim of this study is to examine the performance of a polytomous regression mixture model for finding differential effects when within class residuals are highly skewed. Previous research has shown that when errors are mild to moderately skewed, this approach is reliable for latent class enumeration and yields the correct pattern of model results. This approach, however, also demonstrated some evidence for biased parameter estimates as skew increased even under these mild conditions. The current study examines model performance when errors are highly skewed in one or more classes. We hypothesize that, when data is drawn from two populations differing in the effects of a predictor on an outcome, the model will correctly identify the presence of differential effects as evidenced by selection of the correct model over a more parsimonious or a larger model, but that parameter estimates will be increasingly biased as non-normality increases.

5. Methods

Data generation

This study uses Monte Carlo simulations to address the current aims. Data were generated in R [19], and regression mixture models were estimated using Mplus version 6 [20] called in batch mode (see Appendix for code). Three thousand observations were drawn from each of two sub populations each with simple structure $Y_i = \beta_0 + \beta_p X_i + e_i$ in both classes. Differential effects in the two classes were evidenced by different slopes for the effects of X on Y; X is generated from a standard normal distribution with a mean of 0 and a standard deviation of 1. In all analyses there are two differential effects, or distinct groups present in the data. The first class is drawn from a population using $Y_i = \beta_0 + .20X_i + e_i$, where, the standard deviation of X was 1 and e was scaled so that the variance of Y would be 1, which results in the regression coefficient being equivalent to a correlation. The range for the error terms are presented for the two most extreme conditions; for Scenario A the variances of e were between -1.33 and 3.28 ($M = -1.13e-16$, $SD = 0.648$) and for Scenario J the variances were between -1.40 and 6.08 ($M = -7.87e-17$, $SD = 0.93$). The second class is drawn from a population with an intercept of either zero or .5, $Y_i = \beta_0 + .70X_i + e_i$ or $Y_i = .50 + .70X_i + e_i$, where, the standard deviation of X was 1 and e was scaled so that the variance of Y would be 1, which resulted in the regression coefficient being equivalent to a correlation. Fifty percent of the cases are drawn from each sub population. These differential effects equations are used as they are similar to the effects that may exist in real data, there is a presence of a strong effect of X on Y for one group, and a weaker effect for another group; when distributional assumptions are met these differential effects are detectable by regression mixture models [5] and thus conditions of no skew are not investigated in the current study. Five hundred simulations were performed in each of the scenarios.

Data were generated under ten scenarios described in Table 1; each condition varied based on the degree of non-normality in the residual distribution in one or both of the differential effects groups and the absence or presence of an intercept difference between groups.

Scenarios were comprised of different degrees of skew in the distribution such that in one differential effect group there was either no skew or a positive skew of 1.0 or 1.5, and in the second differential effect group the error distribution was positively skewed 1.0 or 1.5. Each of these conditions were also designed with an intercept of 0 in one group, and varying the intercept, either 0 or .5 in the other differential effect group, resulting in ten conditions. Non-normality in errors was created by adding a constant to the error to eliminate negative values and then using a power transformation of 2.76 or 4.0 to obtain the desired skewness for the condition. The resulting error was then rescaled so that the mean was 0 and the standard deviation of Y was constant across different conditions. This transformation also impacts higher order moments including kurtosis, but the primary impact was on skewness. The values for the power transformations were chosen as they represent major violations of distributional assumptions, greater than mild to moderate levels of skew demonstrated in previous work [5]. The presence of an intercept difference between the differential effects was also examined as it is likely with real data, and is expected to yield better model performance because there is more separation between classes in the data.

Analysis models

In the presence of within-class non-normally distributed errors, one approach to circumvent the violation of distributional assumptions is to use an ordered polytomous regression mixture model. Thus, the data was transformed into an ordered polytomous variable by dividing the continuous outcome variable into six categories with equal distance between the 5th and 95th percentiles with each respondent placed in the relevant category (the top and bottom categories go to negative and positive infinity); previous work has indicated that six is an optimal number of categories with fewer categories (i.e., two and four) being too coarse. Comparisons of models for the ordered categorical data included the examination of three models: a single class model, a two class model and a three class model. Indices of model fit and parameter estimates were examined to determine if the two class model would be selected in the presence of skewed errors across the ten scenarios. The BIC and adjusted BIC were used as criteria for latent class enumeration, the AIC is reported but was not used because previous research has shown that it selects too many classes [21, 22]. Each model was examined for accuracy in parameter estimates, including regression weights and intercepts for each differential effect, and appropriate percentages of respondents for each group based on the simulated data.

6. Simulations

To evaluate the performance of regression mixtures in the presence of non-normal within-class errors we transformed the continuous outcome variable into an ordered categorical variable. Table 2 presents summaries of the results across the ten scenarios with varying conditions of skew in errors in one or both classes; values reported for the AIC, BIC, and adjusted BIC are the proportion of simulations choosing k over k-1 classes. Solutions where the log-likelihood value was not replicated in 100 starts or where no solution was obtained occurred in less than 1% of the simulations; all of the two class model solutions converged and over 99% of the three class models converged. In all cases the smallest class was of a reasonable size (27% to 50%) indicating that none of the results would be excluded from consideration because of non-convergence or small class size. Entropy was quite low for all results, ranging from .13 to .32. As the two differential effects overlap substantially it is difficult to classify a particular individual; this does not discount the value of these models for detecting differential effects in the population, but is important to consider if the goal is to classify individuals. This finding supports the ultimate inclusion of covariates for improving the prediction of class membership[5], although we recommend including covariates in a two-step process to avoid model misspecification.

The next consideration is latent class enumeration. Across the ten scenarios the BIC selected the two class model over the single class model for 95% to 100% of the simulations. The BIC selected the two class model over the three class model ranging from 89% to 100% across scenarios. There was some evidence for the BIC selecting the three class model (for 11% of the simulations) in only the case where there were intercept differences and high skew in the large effect class. The four scenarios in which the BIC selected the two class model over the single class model less than 100% of the time did not yield any support for the selection of the three class model over the two class model. The adjusted BIC selected the two class model over the single class model for 100% of cases; the two class model was selected over the three class model 100% of the time in four conditions with some skew in both classes (i.e., Scenarios E, F, G, and H). There was some evidence of the three class model being selected by the adjusted BIC for the other conditions, especially in the presence of high skew in only one class (i.e., Scenarios C and D) and high skew in both classes (i.e., Scenarios I and J). Notably, the adjusted BIC showed support for selecting the three class model in the presence of an intercept difference in 80% of the simulations in the context of high skew (1.5) in only one class and in 49% of simulations in the context of high skew (1.5) in both classes. These four scenarios, in which the adjusted BIC showed support for the three class model over the two class model, were cases in which the two class model was selected 100% of the time over the single class model. Overall, the ordinal approach was successful in selecting the appropriate two class model for the regression mixtures in the presence of skewed errors, although some support for selecting a three class model in the presence of extreme skew in one or both classes, especially with an intercept difference among differential effects.

When examining the parameter estimates for the two class models, however, the results for the differential effects were somewhat biased in most cases. Because the model is run on simulated data which is recoded to be ordinal, the parameters from the population do not directly map onto the results of the ordinal outcome model which contains five thresholds each indicating the log odds of the probability at being at the next highest level of Y conditional on X. Therefore, the population values for each threshold level are not known; estimates are derived by estimating the ordinal outcomes model separately for each class using the data from all simulations simultaneously (1,500,000 subjects are included per class), estimated population values are reported in Table 3 along with the average parameter estimates and standard errors across simulations for the two class regression mixture models. The correct substantive conclusion is reached in all cases: one class has a small positive effect of the predictor ranging from .21 to .26, and the second class has a stronger positive effect ranging from 1.10 to 1.21. However, the estimated population values and the simulated results show clear and consistent differences across simulations. The results indicated a similar pattern consistent across the ten scenarios of skew when comparing the population estimates and the simulated results; results for six scenarios (selected to demonstrate a range of skew in one or both classes) can be seen in Figure 1. The Y-axis is the predicted probability of endorsing each category for an individual at each given level of X; the relationship between X and Y is a cumulative probability where there are thresholds for X above which individuals have a greater than 50% probability of responding at the next higher level of Y. The highest line or line furthest to the left represents the cumulative probability for the threshold at or greater than category 2; while the lowest line or line furthest to the right represents the cumulative probability for the threshold at category 6. In comparison to the population values (depicted by dashed lines), the simulated results (depicted by solid lines) for the stronger effect class are shifted to the right, while the simulated results for the class with lower effects are shifted to the left which can be seen in the cumulative probability responses depicted in Figure 1. Scenarios A and C yielded the greatest bias in estimates consistent with the presence of high skew in only one class; scenarios I and J, where highly skewed errors were present in both classes, yielded

substantial bias and fairly large standard errors in parameter estimates as well as seen in Table 3. Generally consistent across scenarios, the presence of high skew in one or both classes resulted in under estimating the lower thresholds and overestimating the higher thresholds, although the point at which the bias switched from under to over estimating varied. For example, for scenarios I and J, for Class 1 the three lower thresholds were underestimated and the two higher thresholds were overestimated, while for Class 2 the two lower thresholds were underestimated and the three higher were overestimated. Although the parameter estimates for the simulated results were biased, substantively the results supported the two differential effects. Notably, for example, when plotting the density of the betas for class 1 and class 2 in scenario A the distribution shows the two distinct effects or mixtures within the distribution (Figure 2). Additionally, the proportions for each class were not always close to the 50% simulated in each class; with proportions ranging from 33% to 68% in either class.

7. Conclusions

In this paper we have examined the utility of the ordered polytomous approach when using regression mixture models for identifying two differential effects present in a simulated data. Ten scenarios differing in the level of high skew in one or both classes were examined. Model comparisons across single, one-and two-class models were examined with model fit indices generally identifying the accurate two class model in most cases, especially the BIC, which selected the two class models at or greater than in 89% of the cases across the ten scenarios. The adjusted BIC selected the two class model generally, but in the presence of an intercept difference and extreme skew in one class, a three class model was often supported. Importantly the class enumeration preserved the pattern of the differential effects allowing for identification of the differential effects.

Findings also suggest while the accurate two class model may be recovered in the context of high skew in errors, the parameter estimates for the simulated results were notably biased. While it was expected that bias would increase as skew increased, the results indicated a similar pattern consistent across all scenarios of skew. A limitation of this transformation is in the interpretation of the magnitude of the effects in each subpopulation as direct comparison to the regression coefficients observed in the continuous outcome is difficult. While estimates were bias, the overall ordering of the magnitude of the differential effects was accurate; that is, the presence of a group with a strong association between predictor and outcome and a group with a weak but significant association. Overall, the presence of non-normally skewed within-class residuals is problematic for using regression mixture models. Transforming continuous outcome variables to categorical variables may prove useful for correctly identifying the number of differential effects present in the population and the general form of those differences. However, we suggest caution in the interpretation of the magnitude of differential effects obtained with this approach.

Detecting moderation and the specific number of classes representing the effects is problematic in the context of non-normal errors. While this first step in identifying the appropriate number of effects may be captured by the transformation of the continuous variable to a categorical variable, identifying the parameter estimates within the class is more problematic. Further examination of alternative approaches for capturing parameter estimates is an important next step. In mixture models, the presence of non-normal errors typically impacts latent class enumeration by the selection of additional latent classes; this may be an important consideration in identifying procedures for determining the accuracy of parameter estimates. Perhaps once the number of classes is known, other procedures can be used with modelling the classes using the untransformed, continuous outcome to lead to detection of accurate parameter estimates. Little is known about this or other alternative

approaches to capturing parameter estimates of differential effects in the presence of highly skewed errors when using regression mixture models. Additionally, while the ordered polytomous approach was useful for identifying the presence of the differential effects in the population in the presence of highly skewed errors, other situations or conditions should be investigated to further understand the effectiveness of using this and other approaches to handling skewed errors.

Acknowledgments

This research was supported by grant number # R01HD054736 awarded by the National Institute of Child Health and Human Development. We are grateful for the support and feedback from our colleagues in the Prevention Science and Methodology Group, supported by NIDA through grant # R01MH40859 awarded to Hendricks Brown.

References

1. Muthen BO, Asparouhov T. Multilevel regression mixture analysis. *Journal of the Royal Statistical Society, Series A*. 2009; 172:639–657.
2. Van Horn ML, et al. Assessing Differential Effects: Applying Regression Mixture Models to Identify Variations in the Influence of Family Resources on Academic Achievement. *Developmental Psychology*. 2009; 45:1298–1313. [PubMed: 19702393]
3. Wedel, M.; Desarbo, WS. A review of recent developments in latent class regression models. In: Bagozzi, RP., editor. *Advanced Methods of Marketing Research*. Blackwell Publishers; Cambridge: 1994. p. 352–388.
4. Wedel M, Desarbo WS. A mixture likelihood approach for generalized linear models. *Journal of Classification*. 1995; 12:21–55.
5. Van Horn ML, et al. Not quite normal: Consequences of violating the assumption of normality with regression mixture models. *Structural Equation Modeling*. In Press.
6. Desarbo WS, Jedidi K, Sinha I. Customer value analysis in a heterogeneous market. *Strategic Management Journal*. 2001; 22:845–857.
7. Kaplan D. Finite mixture dynamic regression modeling of panel data with implications for response analysis. *Journal of Educational and Behavioral Statistics*. 2005; 30(2):169–187.
8. Ding C. Using Regression Mixture Analysis in Educational Research. *Practical Assessment Research & Evaluation*. 2006; 11:1–11.
9. Schmeige SJ, Levin ME, Bryan AD. Regression mixture models of alcohol use and risky sexual behavior among criminally-involved adolescents. *Prevention Science*. 2009; 10:335–344. [PubMed: 19459047]
10. McClelland GH, Judd CM. Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*. 1993; 114:376–390. [PubMed: 8416037]
11. Boyce WT, et al. Social context in developmental psychopathology: Recommendations for future research from the Mac Arthur Network on Psychopathology and Development. *Development and Psychopathology*. 1998; 10:143–164. [PubMed: 9635219]
12. Cohen, J., et al. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3. Mahwah, NJ: Lawrence Erlbaum Associates; 2003.
13. McLachlan, G.; Peel, D. *Finite Mixture Models*. New York: John Wiley & Sons, Inc; 2000.
14. Bauer DJ, Curran PJ. Distributional assumptions of growth mixture models: Implications for over extraction of latent trajectory classes. *Psychological Methods*. 2003; 8:338–363. [PubMed: 14596495]
15. Bauer DJ, Curran PJ. The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*. 2004; 9:3–29. [PubMed: 15053717]
16. Tofighi, D.; Enders, CK. Identifying the correct number of classes in growth mixture models. In: Hancock, GR.; Samuelsen, KM., editors. *Advances in Latent Variable Mixture Models*. Information Age; 2008. p. 317–341.
17. Kutner, MH., et al. *Applied linear statistical models*. 5. New York: McGraw-Hill; 2005.

18. Whitehead J, Jaki T. One-and two-stage design proposals for a phase II trial comparing three active treatments with control using an ordered categorical endpoint. *Statistics in Medicine*. 2009; 28(5): 828–847. [PubMed: 19061203]
19. R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing; Vienna, Austria: 2008.
20. Muthen, LK.; Muthen, BO. *Mplus*. Muthen & Muthen; Los Angeles: 2008.
21. Nylund KL, Asparouhov T, Muthen BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*. 2007; 14(4):535–569.
22. Naik PA, Shi P, Tsai CL. Extending the Akaike information Criterion to Mixture Regression Models. *Journal of the American Statistical Association*. 2007; 102(477):244–254.

Appendix. R and MPlus Code examples for data generation and analysis

```
# Generates data for regression mixture simulations in
# 'Using regression mixture models with non-normal data: Examining an
ordered # polytomous approach'
# ng1 is the number of subjects in class 1, ng2 is the number in class 2,
# rep is the number of that replication (used for multiple simulations),
# is the power that the error term is raised to for creating non-normal
errors
datagen<-function(ng1=1000, ng2=1000, rep=1, pow=1.3, location){
  dat1<-matrix(NA,ncol=3,nrow=ng1) # Data for class 1 is in dat1
  dat2<-matrix(NA,ncol=3,nrow=ng2) # Data for class 2 is in dat2
  dat1[,3]<d1 # Indicator for the true latent class
  dat2[,3]<-2 # Indicator for the true latent class
  dat1[,1]<-rnorm(ng1) # Creates the predictor variable for class 1
  dat2[,1]<-rnorm(ng2) # Creates the predictor variable for class 2
  glerror<-rnorm(ng1,sd=0.98) # this is the residual for class 1,
  # sd of .98 is a function of the effect size of .20 in this class.
  # Variance accounted for is .20^2 = .04, total variance of the outcome
  # is 1 - .04 = .96, the sd is .96^.5 = .98
  nonnorm_error<-(rnorm(ng2, 5))^pow
  # non-normal error term has a mean of 5 so that there will be no negative
  values
  x.bar<-mean(nonnorm_error)
  s<-sd(nonnorm_error)
  g2error<-(nonnorm_error-x.bar)/s*.714
  # error term for class 2 is recentered to 0, and rescaled so that the total
  # variance of the outcome in group 2 will be 1
  dat1[,2]<-dat1[,1]*0.2+glerror # The outcome for class 1 is a function of the
  # predictor variable and the error term, effect size for the predictor is .2
  dat2[,2]<-0.5 + dat2[,1]*0.7+g2error
  # effect size for the predictor is .7 in class 2
  file.str<-paste(location, "sim",rep, ".txt",sep= "")
  # Specifies the file to write out dataset
  dat<-rbind(dat1,dat2)
  # creates data set with both classes
  low<-quantile(dat[,2],.05,na.rm=T)
  high<-quantile(dat[,2],.95,na.rm=T)
```

```

interval<-(high-low)/5
# Identifies the quartile range between .05 and .95 to determine interval
cutoffs
# for transforming into a categorical outcome variable
d<-rep(6, length(dat[,3]))
d[dat[,2]< (low+interval*5)]<-5
d[dat[,2]< (low+interval*4)]<-4
d[dat[,2]< (low+interval*3)]<-3
d[dat[,2]< (low+interval*2)]<-2
d[dat[,2]< (low+interval)]<-1
dat<-cbind(dat,d)
# Identifies the 5 thresholds for each level of the ordered categorical
variable
# with six categories
write.table(dat,file.str,row.names=F,col.names=F)
# Writes out the resulting data
}
for(i in 1:50){
datagen(ng1=3000,ng2=3000,rep=i, pow=1, location = "c://work//")
}
!MPlus Code for 2 Class Model
variable:
names are x y num y_c;
categorical is y_c;
classes=c(2);
usevariables are
x y_c;
analysis: type=mixture;
starts=500 100;
processors=8 (starts);
model:
%overall%
y_c on x;
%c#1%
y_c on x; ! models an effect of predictor on outcome for class 1
%c#2%
y_c on x; ! allows for the effect to differ for class 2

```

Figure 1(a).

Figure 1(b).

Figure 1.

Graphs depict the estimated population values (dashed lines) and simulation results (solid lines) for six simulation scenarios; lines represent the cumulative probability endorsed at each level of the transformed categorical variable.

Figure 2.

Figure depicts the density of the beta weights for class 1 and class 2 of the regression mixture model for Scenario A.

Table 1

Ten scenarios based on varying skew and intercept difference for examining differential effects groups

Simulation Condition	Effect 1 $Y = B_0 + .20 * x + e$		Effect 2 $Y = B_0 + .70 * x + e$	
	Skew of e	Intercept (B_0)	Skew of e	Intercept (B_0)
Scenario A	0	0	1	0
Scenario B	0	0	1	.5
Scenario C	0	0	1.5	0
Scenario D	0	0	1.5	.5
Scenario E	1	0	1	0
Scenario F	1	0	1	.5
Scenario G	1	0	1.5	0
Scenario H	1	0	1.5	.5
Scenario I	1.5	0	1.5	0
Scenario J	1.5	0	1.5	.5

Table 2

Latent class enumeration for the ordered polytomous outcome in ten scenarios varying in skew in the errors

	Scenario A			Scenario C			Scenario E			Scenario G			Scenario I		
	No Skew, Skew=1			No Skew, Skew=1.5			Both Skew=1			Skew=1, Skew=1.5			Both Skew=1.5		
Classes	2	3		2	3		2	3		2	3		2	3	
Loglikelihood	-9868	-9856		-9774	-9755		-9776	-9763		-9616	-9605		-9531	-9517	
Entropy	0.17	0.27		0.28	0.28		0.15	0.28		0.25	0.32		0.20	0.30	
AIC	1.00	1.00		1.00	1.00		1.00	0.96		1.00	0.85		1.00	1.00	
BIC	0.96	0.00		1.00	0.01		0.95	0.00		1.00	0.00		1.00	0.00	
Adj BIC	1.00	0.07		1.00	0.38		1.00	0.00		1.00	0.00		1.00	0.18	
Smallest Class *	24%-46%	7%-30%		23%-41%	11%-30%		24%-49%	8%-29%		26%-38%	5%-31%		33%-49%	8%-30%	
Non-convergence	0%	1%		0%	0%		0%	0%		0%	1%		0%	1%	

	Scenario B			Scenario D			Scenario F			Scenario H			Scenario J		
	No Skew, Skew=1			No Skew, Skew=1.5			Both Skew=1			Skew=1, Skew=1.5			Both Skew=1.5		
Classes	2	3		2	3		2	3		2	3		2	3	
Loglikelihood	-9877	-9862		-9786	-9764		-9865	-9854		-9807	-9791		-9703	-9683	
Entropy	0.16	0.28		0.16	0.27		0.13	0.28		0.13	0.30		0.16	0.30	
AIC	1.00	1.00		1.00	1.00		1.00	0.98		1.00	0.97		1.00	1.00	
BIC	1.00	0.00		1.00	0.11		0.99	0.00		0.98	0.00		1.00	0.00	
Adj BIC	1.00	0.10		1.00	0.80		1.00	0.00		1.00	0.00		1.00	0.49	
Smallest Class *	26%-48%	8%-30%		34%-49%	12%-31%		33%-49%	9%-29%		35%-49%	9%-29%		32%-49%	12%-29%	
Non-convergence	0%	1%		0%	0%		0%	2%		0%	1%		0%	1%	

Note.

* the 95% range of the proportion of cases in the smallest class across all simulations; loglikelihood and entropy values are medians across all simulations

Table 3

Results for select scenarios of high skew in one or both classes for the two class model with an ordered polytomous outcome

Scenario A (Same Intercept, No Skew, Skew=1)						
	Population Values		Class 1: 38%		Class 2: 62% (23%-77%)*	
	Class 1	Class 2	B	(SE)	B	(SE)
T1	-1.09	-1.63	-2.14	0.09	-2.59	0.22
T2	-0.53	-0.72	-1.14	0.06	-0.77	0.12
T3	0.04	0.13	-0.16	0.06	0.89	0.23
T4	0.59	0.94	0.83	0.07	2.34	0.37
T5	1.15	1.67	1.90	0.08	3.68	0.44
Slope	0.22	1.11	0.51	0.13	2.30	0.30
Scenario C (Same Intercept, No Skew, Skew=1.5)						
	Population Values		Class 1: 68%		Class 2: 32% (24%-78%)*	
	Class 1	Class 2	B	(SE)	B	(SE)
T1	-1.06	-1.69	-2.11	0.06	-3.25	0.30
T2	-0.51	-0.69	-1.13	0.04	-0.64	0.14
T3	0.05	0.22	-0.16	0.05	1.69	0.35
T4	0.61	1.04	0.82	0.06	3.89	0.59
T5	1.17	1.75	1.85	0.07	5.39	0.78
Slope	0.21	1.17	0.59	0.07	3.33	0.37
Scenario I (Same Intercept Both Skew=1.5)						
	Population Values		Class 1: 44%		Class 2: 56% (33%-69%)*	
	Class 1	Class 2	B	(SE)	B	(SE)
T1	-0.88	-1.28	-2.49	0.31	-1.53	0.08
T2	-0.10	0.34	-0.74	0.11	0.08	0.11
T3	0.47	0.50	0.21	0.13	1.60	0.18
T4	0.92	1.25	1.02	0.09	3.02	0.36
T5	1.30	1.90	1.75	0.09	4.28	0.54
Slope	0.25	1.14	0.30	0.10	2.14	0.26

Scenario A (Same Intercept, No Skew, Skew=1)						
	Population Values		Class 1: 38%		Class 2: 62% (23%-77%)*	
	Class 1	Class 2	B	(SE)	B	(SE)
Scenario J (Different Intercept, Both Skew=1.5)						
	Population Values		Class 1: 58%		Class 2: 42% (30%-67%)*	
	Class 1	Class 2	B	(SE)	B	(SE)
T1	-0.53	-1.87	-1.30	0.06	-2.67	0.24
T2	0.17	-0.85	-0.12	0.06	-1.06	0.11
T3	0.70	0.08	0.64	0.05	0.77	0.15
T4	1.13	0.92	1.36	0.05	2.52	0.28
T5	1.49	1.65	2.12	0.06	3.97	0.37
Slope	0.24	1.18	0.40	0.07	2.49	0.26

Note.

* the 95% range of the proportion of cases in each class