Sankhyā : The Indian Journal of Statistics
2018, Volume 80-B, Part 1, pp. 98-122

An Alternate Approach to Pseudo-Likelihood Model Selection in the Generalized Linear Mixed Modeling Framework

Patrick Ten Eyck
The University of Iowa, Iowa City, USA
Joseph E. Cavanaugh
The University of Iowa, Iowa City, USA

Abstract

In this paper, we propose and investigate an alternate approach to pseudo-likelihood model selection in the generalized linear mixed modeling framework. The problem with the natural approach to the computation of pseudo-likelihood model selection criteria is that the pseudo-data vary for each candidate model, leading to criteria based on fundamentally different goodness-of-fit statistics, rendering them incomparable. We propose a technique that circumvents this problem. This new approach can be implemented using a SAS macro that obtains and applies the pseudo-data from the full model to fitting candidate models based on all possible subsets of predictor variables. We justify the propriety of the resulting pseudo-likelihood selection criteria through an extensive study designed as a factorial experiment. We then illustrate this new method in a modeling application pertaining to bullying in public schools. The data set for the application is taken from three waves of the Iowa Youth Survey.

AMS (2000) *subject classification.* Primary 62J12, Secondary 62F07.
Keywords and phrases. Akaike information criterion, Generalized linear mixed model, Pseudo-likelihood, Variable selection

1 Introduction

In regression frameworks, model selection encompasses a variety of approaches, including the use of diagnostics and criteria to evaluate a collection of candidate models and determine which provides the most appropriate fit to a given set of data. In the maximum likelihood framework, the use of information criteria based on the empirical likelihood is pervasive, where the empirical likelihood appears in the goodness-of-fit term of the criterion. To compare such criteria across different fitted models, the outcome data must be identical; otherwise, the likelihoods will not correspond.

The pseudo-likelihood method is a conventional fitting approach for the framework of the generalized linear mixed model (GLMM). With this method, *pseudo-data* are generated via a transformation of the outcome and used as a surrogate for the original response data. Pseudo-data are derived from a Taylor series expansion that utilizes both constructs from the candidate model and the original outcome. The purpose of this expansion is to offer a new outcome with an approximate normal distribution. In general, pseudo-data are inconsistent for different model specifications, leading to pseudo-likelihoods that are not comparable. Selection criteria based on the resulting goodness-of-fit statistics are fundamentally dissimilar, rendering comparisons invalid for evaluation.

The purpose of this paper is twofold. First, we investigate the natural approach to model selection using pseudo-likelihood based information criteria under the GLMM framework. With this approach, the criteria are constructed using the pseudo-data based on the candidate model at hand. Second, in this setting, we propose a new, improved method for the comparison of information criteria between candidate models. In SAS, the default method for the GLIMMIX (Generalized LInear MIXed Models) procedure is the natural, pseudo-likelihood based approach, leading to invalid model comparisons via information criteria. Our new method will be implemented using the GLIMMIX and MIXED (linear MIXED models) SAS procedures.

In SAS, the default fitting procedure for GLMMs utilizes the pseudo-likelihood. Under the natural approach to model selection in this setting, the use of information criteria, such as the Akaike information criterion (AIC; Akaike, 1973, 1974) and the Bayesian information criterion (BIC; Schwarz 1978), is problematic. This is due to their reliance on pseudo-data that are generated by transforming the original outcome vector into a new outcome, which is used to construct a Gaussian likelihood. Computing model selection criteria using pseudo-data from the specified model may seem defensible, but this approach is fundamentally flawed. Any link function other than the identity link will lead to different pseudo-data under different fitted candidate models, violating the assumption that models under comparison share the same outcome. This violation is ignored under the default GLIMMIX procedure, thereby rendering model selection criteria of dubious utility. For most GLMMs, the identity link is neither the canonical link nor the most appropriate link. In order to ensure that use of model selection criteria is valid, we need to verify that the same pseudo-data are being used for all models under consideration. The simplest way to accomplish this objective is to use the full model (i.e., the model featuring all predictor variables under consideration)to obtain the pseudo-data, and to subsequently fit all subset candidate models with this generated outcome.

One common approach to generalized linear mixed modeling utilizes residual pseudo-likelihood. The residual maximum likelihood (REML) method is sometimes preferred to the maximum likelihood (ML) method in that it accounts for fixed effects in the construction of the objective function, which reduces the bias in covariance parameter estimates, sometimes to zero. Both REML and ML can be implemented using the pseudo-likelihood method, but for the purposes of this paper, we will focus exclusively on ML since AIC and BIC have been developed for this framework. Pseudo-likelihood is one of several GLMM estimation methods, including Gaussian quadrature and Laplace approximation, which are often necessary since integrating random effects out of the joint likelihood is typically intractable. The latter two methods are valuable in that they can approximate the marginal likelihood through numeric integration, which leads to traditional model selection criteria, unlike pseudo-likelihood. However, a disadvantage of numerical approximation is computational expensiveness. Conversely, maximizing a Gaussian likelihood based on pseudo-data, which are created using Taylor series for each iteration, is computationally efficient.

The structure of this paper is as follows. In Section 2, we provide some background regarding generalized linear mixed models and the associated criteria used in model selection, as well as pseudo-likelihood and the default implementation of AIC and BIC under the GLIMMIX procedure in SAS. We will highlight the problem with this approach using an investigative simulation, then propose a solution. In Section 3, in the pseudo-likelihood setting, we propose a new approach for generalized linear mixed model selection using a heuristic justification. We provide a detailed account of the implementation via the MIXED and GLIMMIX procedures in SAS. In Section 4, we use a simulation study to illustrate and compare the behavior of the selection criteria found using the default GLIMMIX procedure and the newly proposed technique. In Section 5, we apply the new technique in a modeling application and examine the results. Section 6 concludes.

2 Background

In this section, we begin with a review of model selection criteria. We also examine the construction of the pseudo-likelihood and consider its impact on the formulation of model selection criteria. Under the default implementation of the GLIMMIX procedure in SAS, we show that the comparison of information criteria for non-normal models is not appropriate and can lead to misguided model selections. We provide an investigative simulation, which illustrates model selection tendencies under the default implementation, then propose a more suitable implementation approach for this framework.

2.1. Model Selection Criteria. Statistical models are used to characterize the relationship between an outcome of interest and explanatory factors. Models condense information into an interpretable form, from which investigators can draw inferential conclusions. Modeling frameworks have been developed to handle outcomes that assume distributions of all varieties. Once fit, models can be applied to new data in order to predict new outcomes.

An optimal statistical model is characterized by three features: (1) parsimony, which refers to model simplicity; (2) goodness-of-fit, which indicates the conformity of the fitted model to the data at hand; (3) generalizability, which reflects the ability of the fitted model to predict or describe new outcomes. Parsimony and goodness-of-fit tend to pull in opposing directions with regards to model complexity, so it is important to strike a suitable balance between those two attributes, while still achieving generalizability.

In a model selection problem, an investigator strives to find the "best" model from a collection of candidate models, where optimality may be defined based on adherence to the preceding principles. For theoretical and methodological developments pertaining to model selection, one needs to assume the existence of an underlying generating probabilistic mechanism. We will refer to this mechanism as a "true" model. In our development, we will assume that the true model is contained within the candidate collection. Although this is a strong assumption, it is commonly employed in model selection developments for either mathematical tractability or conceptual clarity. Here, we impose the assumption for the benefit of the latter.

Investigators frequently use model selection criteria in order to compare different candidate models and ascertain the one that best exemplifies the three optimality features. A common approach to the development of a model selection criterion is to estimate a measure that assesses the disparity between the fitted model under consideration and the true probabilistic mechanism. Such a measure is known as an *expected discrepancy.* One of the most popular and useful expected discrepancies is based on the Kullback-Leibler (K-L) information, a measure introduced by Kullback and Leibler (1951) and further investigated by Kullback (1968). This discrepancy serves as the basis for the ubiquitous Akaike (1973, 1974) information criterion (AIC) and its variants. One of the favorable properties of AIC is asymptotic efficiency, in the sense of Shibata (1980, 1981). Assuming that the generating model is of an infinite dimension and thus is not in the candidate collection, an efficient criterion will asymptotically select the fitted candidate model which minimizes the mean squared error of prediction. As outlined by Linhart and Zucchini (1986), a major deficiency of AIC arises in small to moderate sample-size applications, where AIC will often severely underestimate the K-L discrepancy and may tend to decrease as model complexity increases. Variants of AIC have been proposed to address this deficiency and to relax the stringent model specification assumption under which the criterion is derived. These include corrected AIC (AICc) (Sugiura, 1978; Hurvich and Tsai, 1989), designed for small-sample settings, the Takeuchi (1976) information criterion (TIC), which relaxes the model specification assumption, CAIC (Bozdogan, 1987), which corrects for the lack of consistency, and a quasi-likelihood based measure for generalized linear models fit via generalized estimating equations (QIC) (Pan, 2001). Additional variants of AIC have been proposed based on complexity penalizations that are evaluated using a computationally intensive algorithm, including cross-validation (Stone, 1977; Davies et al., 2005), bootstrapping (Ishiguro et al., 1997; Cavanaugh and Shumway, 1997; Shibata, 1997), and Monte Carlo simulation (Hurvich et al., 1990; Bengtsson and Cavanaugh, 2006).

Another common model selection criterion is the Bayesian information criterion, BIC. One of the favorable properties of BIC is its consistency, which is characterized as follows. Suppose that the generating model is of a finite dimension, and that this model is represented in the candidate collection under consideration. A consistent criterion will asymptotically select the fitted candidate model having the correct structure with probability one. BIC was originally derived by Schwarz "for the case of independent, identically distributed observations, and linear models." Variants of BIC that generalize the criterion for other frameworks have been developed by Stone (1979), Kashyap (1982), Leonard (1982), Haughton (1988), and Cavanaugh and Neath (1999).

An advantage that AIC and BIC offer over common model comparison inferential techniques, such as the likelihood ratio test, is that the models under consideration do not need to be nested or even follow the same distribution. As long as the models are fitted using the same outcome, the selection criteria can be compared to determine the more appropriate fit. For a collection of candidate models, the model with the minimum AIC or BIC is deemed the most favorable. However, for the sake of parsimony, a model of lower order and within two units of the minimum information criterion is also considered a suitable selection. (See Burnham and Anderson, 2002, p. 70, regarding AIC, and Kass and Raftery, 1995, p. 777, regarding BIC).

The conditional mean structure of a mixed model is characterized using a linear combination of predictor variables and random effects. For the purpose at hand, we assume that the selection of fixed effects is the focus. In order to ensure that an exhaustive model search has been conducted, investigators use the all possible subsets approach, which allows for comparison of every combination of variables. For a set of p predictor variables, the all subsets approach considers the $2^p - 1$ collections of predictor variables, excluding the null model. Once all models have been fitted, their corresponding model selection criteria are compared to determine which model is "best." For an effective criterion, in large sample settings where the underlying effects are all appreciable in magnitude, the true model will generally have the highest probability of being chosen out of the collection of candidate models. However, in such settings, model overfitting has a less detrimental impact on inferential objectives than underfitting.

2.2. Problem with Pseudo-Likelihood Criteria. As mentioned before, in order for AIC and BIC to serve as legitimate model selection criteria, the candidate class of models must share the same outcome. In fitting GLMMs, the pseudo-likelihood is determined for each model based on different pseudo-data. AIC and BIC use the maximum log pseudo-likelihood in place of the usual maximum log likelihood in the formulation of the goodness-of-fit term. If we wish to compare two candidate GLMMs using AIC or BIC, different pseudo-data will lead to criteria that cannot appropriately compare the two models. These notions will be developed more technically in the following subsections.

2.3. Generalized Linear Mixed Models. The GLMM framework is a natural extension of the linear mixed model (LMM) framework. GLMMs are the best available recourse for analyzing normal and non-normal data that involve random effects, requiring only the specification of a conditional distribution for the outcome, a link function, and a covariance structure for the random effects (Bolker et al., 2009).

Under the GLMM framework, conditional on the random effect parameter vector γ, the $t_i \times 1$ response measurement vector Y_i is assumed to have a distribution in the exponential family, for $i = 1, ..., s$. We have the expression

$$g_i(E[Y_i|\gamma]) = \eta_i = X_i\beta + Z_i\gamma,$$

where X_i is the fixed effects design matrix, β is the $p \times 1$ fixed effect parameter vector, Z_i is the random effects design matrix, and $g_i(\cdot)$ is a strictly monotonic function from $\mathbb{R}^{T_i}$ to $\mathbb{R}^{T_i}$ that maps the elements of the conditional mean vector $E[Y_i|\gamma] = \mu_i$ to the elements of linear predictor $\eta_i = X_i\beta + Z_i\gamma$. The relevant variance/covariance matrices can be expressed as follows:

$$\text{Var}[\gamma] = G$$

and

$$\text{Var}[Y_i|\gamma] = A_i^{1/2} R_i A_i^{1/2}.$$

Here, $A_i^{1/2} = \text{diag}(\sqrt{\text{Var}[Y_{it}]}) = \text{diag}(\sqrt{\upsilon(\mu_{it})})$, where $\upsilon(\mu_{it})$ is the variance of Y_{it}, so that $A_i^{1/2}A_i^{1/2}$ represents the variance/covariance matrix of Y_i under independence and $R_i = \text{Corr}(Y_i)$.

Using the preceding foundation for the GLMM framework, we can examine the methodology behind how a model is fit based on the pseudo-likelihood approach.

2.4. Pseudo-Likelihood Fitting Approach. Though GLMMs provide considerable opportunities for advancement and flexibility in modeling data, inference for these models is complicated by the integrals in the derivation of marginal likelihood estimating equations (Dean and Nielsen, 2007). Pseudo-likelihood is one of several fitting approaches in the GLMM framework. The other primary approaches include Gaussian quadrature and the Laplace approximation method. The advantage of using pseudo-likelihood over these alternate approaches is its computational efficiency. The qualifier "pseudo" is used because the likelihood is based on a linearized transformation of the data, which is assumed normal, and is not the actual likelihood based on the original data and its underlying distribution.

The following development is largely provided in the SAS/STAT User's Guide: Mixed Modeling. Under the GLMM framework, we have

$$E[Y|\gamma] = g^{-1}(X\beta + Z\gamma) = g^{-1}(\eta) = \mu,$$

where $Y = (y_1', \ldots, y_s')'$, $\gamma \sim N(0, G)$, $E[Y|\gamma] = \mu$,

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_s \end{bmatrix}, Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_s \end{bmatrix},$$

and $\eta = (\eta_1', \ldots, \eta_s')'$. Here, $g(\cdot)$, is a link function from $\mathbb{R}^N$ to $\mathbb{R}^N$, defined analogously to $g_i(\cdot)$, that maps the conditional mean vector μ to the systematic component η. We also have $\text{Var}[Y|\gamma] = A^{1/2}RA^{1/2}$, where

$$A = \text{blk diag}(A_1, \ldots, A_s) \text{ and } R = \text{blk diag}(R_1, \ldots, R_s).$$

Following Wolfinger and O'Connell (1993), a first-order Taylor series of μ about iterated parameter estimates $\widetilde{\beta}$ and $\widetilde{\gamma}$ yields

$$g^{-1}(\eta) \doteq g^{-1}(\widetilde{\eta}) + \widetilde{\Delta}X(\beta - \widetilde{\beta}) + \widetilde{\Delta}Z(\gamma - \widetilde{\gamma}),$$

where

$$\widetilde{\Delta} = \left(\frac{\partial g^{-1}(\eta)}{\partial \eta}\right)_{\widetilde{\beta},\widetilde{\gamma}}$$

is a diagonal matrix of derivatives of the conditional mean evaluated at the expansion locus. Rearranging terms yields the expression

$$\widetilde{\Delta}^{-1}(\mu - g^{-1}(\widetilde{\eta})) + X\widetilde{\beta} + Z\widetilde{\gamma} \doteq X\beta + Z\gamma. \tag{2.1}$$

With reference to the left side of the preceding approximation, we define the pseudo-data as

$$\widetilde{\Delta}^{-1}(Y - g^{-1}(\widetilde{\eta})) + X\widetilde{\beta} + Z\widetilde{\gamma} \equiv P. \tag{2.2}$$

Note that

$$\text{Var}[P|\widetilde{\gamma}] = \widetilde{\Delta}^{-1}A^{1/2}RA^{1/2}\widetilde{\Delta}^{-1}.$$

Based on Eqs. 2.1 and 2.2, one can thus consider the model

$$P = X\beta + Z\gamma + \epsilon, \tag{2.3}$$

which is a linear mixed model with pseudo-response data P, fixed effects β, random effects γ, and $\text{Var}[\epsilon] = \text{Var}[P|\gamma]$.

Now define

$$V(\theta) = ZGZ' + \widetilde{\Delta}^{-1}A^{1/2}RA^{1/2}\widetilde{\Delta}^{-1}$$

as the marginal variance in the linear mixed pseudo-model, where θ is the $(q \times 1)$ parameter vector containing all of the unknown parameters in G and R. Based on the linearized model in Eq. 2.3, an objective function can be defined, assuming that the distribution of P is known. The pseudo-likelihood fitting procedure assumes that ϵ has a normal distribution. The log pseudo-likelihood is then defined as

$$l(\theta, P) = -\frac{1}{2}\log[V(\theta)] - \frac{1}{2}R'V(\theta)^{-1}R - \frac{N}{2}\log\{2\pi\}$$

with $R = P - X(X'V^{-1}(\theta)X)^{-1}X'V^{-1}(\theta)P$. Here, N denotes the overall sample size. We assume that X is $N \times (p+1)$ of rank $(p+1)$ and that $X'V^{-1}(\theta)X$ is invertible. The objective function for minimization is $-2l(\theta, p)$. At convergence, the profiled parameters are estimated and the random effects are predicted as

$$\begin{aligned}\hat{\beta} &= (X'V(\hat{\theta})^{-1}X)^{-1}X'V(\hat{\theta})^{-1}\hat{P},\\ \hat{\gamma} &= \hat{G}Z'V(\hat{\theta})^{-1}\hat{R}.\end{aligned}$$

The default approach for the GLIMMIX procedure fits GLMMs based on linearizations, which utilize Taylor series expansions to approximate the data using pseudo-data. In the iterative fitting routine, the pseudo-data are constructed using current regression and covariance parameter estimates. The GLMM is then approximated by a LMM based on the pseudo-data. The LMM fitting is itself an iterative process, yielding new parameter estimates that update the linearization, generating a new LMM. This process is complete once the LMM fits fail to change beyond a pre-specified tolerance.

The empirical pseudo-likelihood under the GLMM framework can seemingly be used to calculate model selection criteria in a similar manner to the empirical likelihood under the GLM framework. The apparent definitions for AIC and BIC are as follows:

$$\begin{aligned}\text{AIC} &= -2\ell(\hat{\theta}, \hat{P}) + 2((p+1)+q),\\ \text{BIC} &= -2\ell(\hat{\theta}, \hat{P}) + \log(N)((p+1)+q).\end{aligned}$$

However, since model selection criteria constructed under the pseudo-likelihood approach utilize a different pseudo-data vector, P, for each model under consideration, comparisons of conformity between the models and the data are inappropriate.

2.5. Investigative Simulation. In order to illustrate the ineffectiveness of model selection via information criteria based on the default GLIMMIX fitting procedure, we consider a simulated data set in which the true, generating model is known. We can thereby illustrate the efficacy of a model selection criterion in delineating this true model from other candidates.

For a sample size of $N = 100$, let Y_{ij} represent the j^{th} observation on subject i, for $i = 1, 2, ..., 20$; $j = 1, 2, ..., 5$. Let

$$\{X_{1ij}, X_{2ij}, ..., X_{6ij}\} \stackrel{iid}{\sim} N(0, 1)$$

represent the j^{th} set of covariates on subject i, with

$$\gamma_i \stackrel{iid}{\sim} N(0, 1)$$

as the random effect for subject i. The Bernoulli outcome

$$Y_{ij} \stackrel{ind}{\sim} \text{Bernoulli}(\pi_{ij})$$

is based on the generating model

$$\text{logit}(\pi_{ij}) = x_{1ij} + x_{2ij} + x_{3ij} + \gamma_i.$$

An evaluation of all possible subset models for the fixed effects, not including the null model, involves fitting $2^6 - 1 = 63$ models. We will consider four criteria for model selection: minimum AIC (minAIC), most parsimonious model within two units of minimum AIC (minAIC2), minimum BIC (minBIC), and most parsimonious model within two units of minimum BIC (minBIC2).

The model selections according to the four criteria are summarized in Table 1. As we can see from the table, all four criteria select the same order one model, which does not include any of the predictors in the generating model. This gross confusion of the actual mean structure is concerning as it shows the futility of the selection procedure for data that follows this modeling framework. Ideally, in settings where the sample size is large and the effects are all reasonable in size, we would want a model selection procedure that identifies the correct model specification with higher probability than any other candidate model.

2.6. Proposed Solution. In order to make the comparison of information criteria between models valid, we need to use the same outcome for all candidate models. The pseudo-data are a function of the predictor variables in the model and may vary for different model specifications. In order to ensure that the outcome is the same for all models under consideration, we

Table 1: Model selections for four criteria using GLMMMIX procedure

Selection criteria	Selected variables
minAIC	X_5
minAIC2	X_5
minBIC	X_5
minBIC2	X_5

can generate the pseudo-data from the full model (i.e., the model that contains all predictors under consideration) and fit all other candidate models to that pseudo-data.

3 New Method

This section applies our proposed solution to model selection for the pseudo-likelihood setting under the GLMM framework, first by providing a heuristic justification, then by detailing the implementation in SAS via the GLIMMIX and MIXED procedures.

3.1. Heuristic Justification. In order to be able to make valid model comparisons using selection criteria such as AIC and BIC, the candidate models must be based on the same outcome data. Under the GLMM framework based on pseudo-likelihood estimation, unique pseudo-data are generated for each candidate model; the pseudo-data then serve as the basis for the construction of each model pseudo-likelihood. The unique pseudo-data lead to disparate pseudo-likelihoods, which are not commensurable, invalidating the comparison of AIC and BIC across candidate models.

Ideally, we would construct the pseudo-data based on the true model, which we do not know. However, if we assume that we have access to the predictors in the true model, we can generate the pseudo-data by using the full model, which subsumes the true model. Using this full model pseudo-data, a LMM can be fit with any subset of predictors from the full model. Since all models will share the same pseudo-data, information criteria can validly be compared for the purposes of model selection.

3.2. Implementation via SAS PROC MIXED/PROC GLIMMIX. The GLIMMIX procedure generates the pseudo-data, then fits the model with the normalized outcome, similar to the manner that the MIXED procedure would fit the model with the same transformed outcome. For an outcome of interest and a collection of candidate predictor variables, we fit the full model using GLIMMIX and output the predicted and residual components of the pseudo-data, given by $X\hat{\beta} + Z\hat{\gamma}$ and $\hat{\Delta}^{-1}(Y - g^{-1}(\hat{\eta}))$, respectively. Equation 2.2 shows that the sum of these components yields the pseudo-data,

Table 2: Model selections for four criteria using new technique

Selection criteria	Selected variables
minAIC	X_1 X_2 X_3 X_4 X_6
minAIC2	X_1 X_2 X_3 X_4
minBIC	X_1 X_2 X_3 X_4
minBIC2	X_1 X_2 X_3

P. We then use the MIXED procedure with the full model pseudo-data to fit all candidate models of interest and generate the desired information criteria, which can be compared for the purposes of model selection.

Looking back at the investigative simulation presented in the previous section, we can apply our new model selection technique, which produces the results featured in Table 2. The most noticeable difference from the default technique is that all four criteria select models that include the three predictors in the generating model. Although three of the criteria select overspecified models, this is preferable to favoring a model that omits any of the true predictors, highlighting the value of our proposed technique. These selection results are corroborated in the next section, which presents a large scale simulation study covering four common modeling distributions.

4 Simulation Study

We compile and report a comprehensive simulation study in order to assess and compare the performances of pseudo-likelihood based model selection criteria under the GLMM framework using the natural approach and our newly proposed approach. By generating numerous replicated samples, and using these samples for model fitting and selection, we can characterize the general behaviors of the criteria. This simulation study focuses on the criteria AIC and BIC for outcomes following Bernoulli, binomial ($n = 10$), Poisson, and gamma distributions.

Each set in the simulation study is based on generating 1000 data samples of size $N = 100$. A random effect intercept is included in the generating model, which effectively partitions the data set into 20 groupings (indexed by i) of 5 observations each. The regression parameters for the systematic component are set to be identical for all predictor variables, and the intercept is set to zero. The simulation study is designed as a factorial experiment, where the factors are the distribution (Bernoulli, binomial, Poisson, or gamma), selection criterion (AIC or BIC), and modeling construction method for the pseudo-data (candidate model, full model, or true model).

The construction methods, described in what follows, dictate the manner in which the goodness-of-fit terms are formulated for AIC and BIC.

For each replicated sample, we record the optimal model selected for each criterion. We summarize the model selections in a table of counts. Such tables allow us to assess the ability of each criterion to pick the correct model, as well as to see which types of incorrectly specified models each criterion tends to favor. Additionally, for every candidate model, we compute the means of AIC and BIC. For both criteria, we consider goodness-of-fit terms based on the candidate model construction (CMC), full model construction (FMC), and true model construction (TMC) of the pseudo-data. The CMC method is the natural (yet incorrect) approach that SAS uses as the default for the GLIMMIX procedure; the FMC method is the proposed approach; the TMC method is the ideal, yet inaccessible, approach that can be conceived as providing the gold standard. With the TMC approach, the pseudo-data are determined based on the structure and parameter values for the generating model; a linear mixed model is then fit based on this idealized pseudo-data. We plot the criterion means by model (grouped by specification) to provide a visual representation of the behaviors of the criteria for underspecified, correctly specified, overspecified, and mixed misspecified models. (Mixed misspecified models contain both legitimate and spurious predictors, yet do not contain all of the predictors represented in the true model.)

This simulation study follows the all subsets modeling setting. Here, we construct data sets with 6 fixed effect predictor variables

$$X_{1ij}, X_{2ij}, ..., X_{6ij} \stackrel{iid}{\sim} \mathrm{N}(0,1),$$

a random effect predictor

$$\gamma_i \stackrel{iid}{\sim} N(0,1),$$

and an outcome Y_{ij} that is generated as shown in Table 3. Let $\mu_{ij} = E[Y_{ij}|\gamma_i]$. For the gamma distribution, the scale parameter is set equal to one, leaving the shape parameter equal to $\mu_{ij} = g^{-1}(X\beta + \gamma_i)$, where $g(\cdot)$ is the canonical link function for the generating distribution.

With the generated data sets, we fit all possible subsets of fixed effect configurations, leading to $2^6 - 1 = 63$ models. The all subsets setting allows us to compare the criteria for models that are underspecified (candidate predictor set is a proper subset of the set of true predictors), correctly specified (candidate predictor set is exactly the same as the set of true predictors), overspecified (true predictor set is a proper subset of the candidate predictor set), and mixed misspecified (candidate predictor set is comprised of some

Table 3: Generating models and link functions for each distribution

Distribution	Generating model	Link
Bernoulli	$Y_{ij}\|\gamma_i \overset{ind}{\sim} \text{Bernoulli}(\mu_{ij})$	logit
Binomial ($n = 10$)	$Y_{ij}\|\gamma_i \overset{ind}{\sim} \text{Binomial}(\mu_{ij})$	logit
Poisson	$Y_{ij}\|\gamma_i \overset{ind}{\sim} \text{Poisson}(\mu_{ij})$	log
Gamma	$Y_{ij}\|\gamma_i \overset{ind}{\sim} \text{Gamma}(\mu_{ij})$	log

but not all of the predictors in the true set, as well as some predictors not in the true set). Again, the simulation study is designed as a factorial experiment, where the factors are the distribution, selection criterion, and the modeling method for the construction of the pseudo-data. As with the investigative simulation, in addition to selecting models with a minimum AIC or BIC, we will also select the most parsimonious model within two units of the minimum AIC and BIC.

Table 4 lists the ID for each simulation set, along with the associated levels of the factors. The Figs. 1, 2, 3, and 4 corresponding to sets AS1, AS3, AS5, and AS7 feature three sets of points; one for each modeling construction method. The criterion means of the 1000 replications are calculated for each model and plotted, illustrating the general behavior of the criteria under each construction method. In what follows, only the plots based on the AIC criterion means are provided; the analogous figures for BIC are omitted, since the mean patterns are similar. The minimum mean CMC, FMC, and TMC model criteria are indicated with a horizontal line. The Tables 5, 6, 7, and 8 corresponding to sets AS1-AS8 feature model specification selection counts by each construction method.

Note that for every configuration of the factorial design, the frequency of correctly specified model select is greater for the FMC criteria than for

Table 4: Model factor levels for all subsets simulation setting

Set ID	Distribution	Criterion
AS1	Bernoulli	AIC
AS2	Bernoulli	BIC
AS3	Binomial (n=10)	AIC
AS4	Binomial (n=10)	BIC
AS5	Poisson	AIC
AS6	Poisson	BIC
AS7	Gamma	AIC
AS8	Gamma	BIC

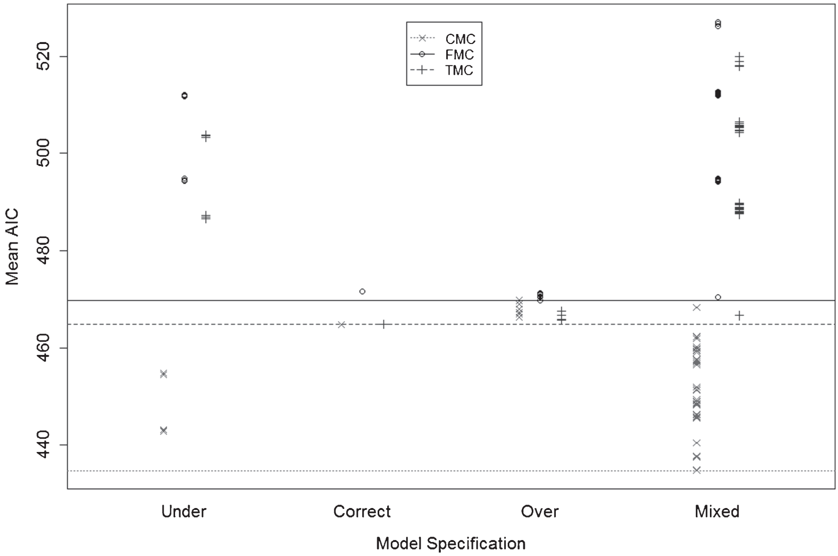

Figure 1: Set AS1—Bernoulli outcomes. Means of AIC by model specification

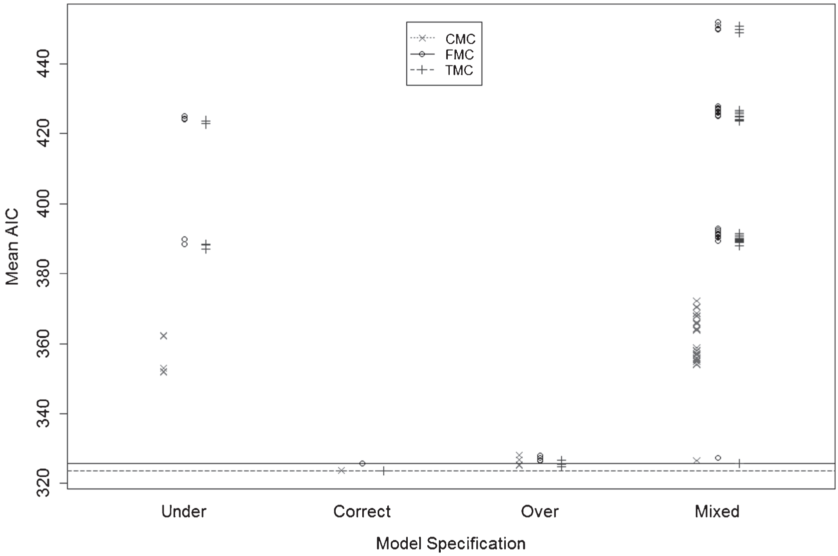

Figure 2: Set AS3—binomial outcomes. Means of AIC by model specification

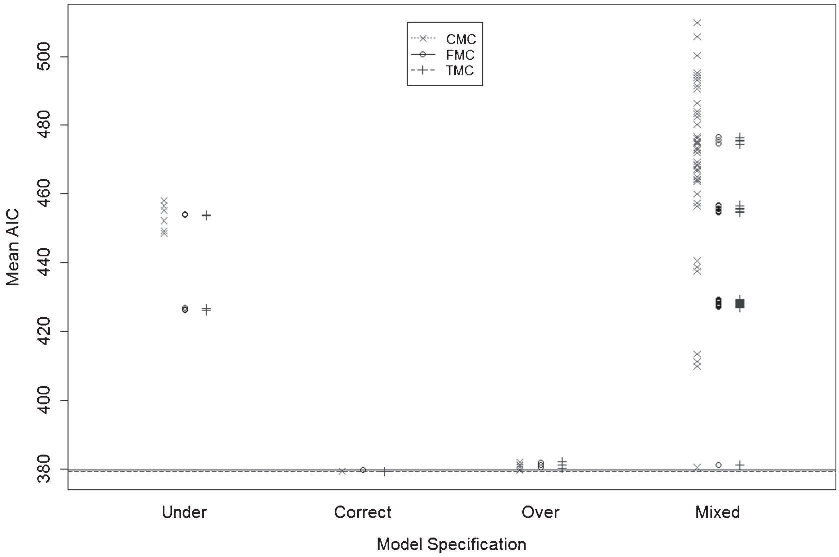

Figure 3: Set AS5—Poisson outcomes. Means of AIC by model specification

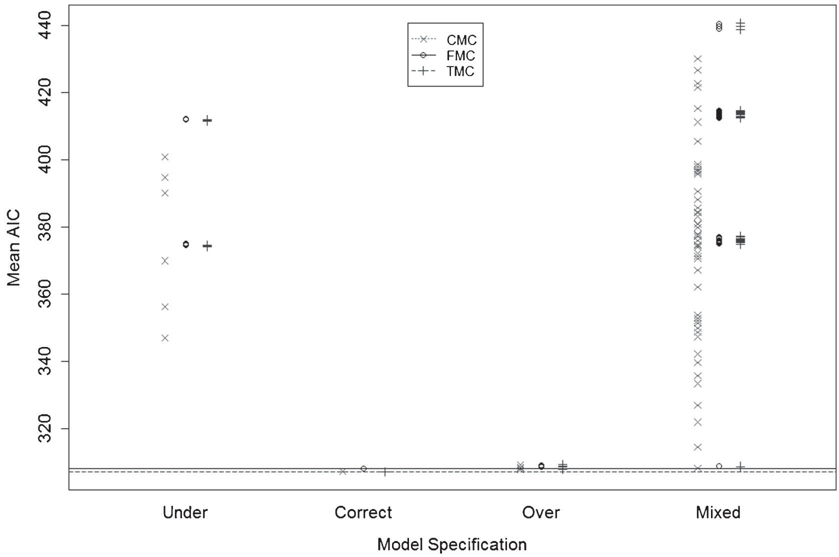

Figure 4: Set AS7—gamma outcomes. Means of AIC by model specification

Table 5: Selections by model specification for Bernoulli outcomes

	Minimum measure						Most parsimonious within 2 of minimum					
Model	minAIC (AS1)			minBIC (AS2)			minAIC2 (AS1)			minBIC2 (AS2)		
specification	CMC	FMC	TMC	CMC	FMC	TMC	CMC	FMC	TMC	CMC	FMC	TMC
Under	90	36	34	102	79	68	311	100	98	339	144	152
Correct	47	279	520	47	370	652	55	434	717	51	478	747
Over	117	658	416	90	521	251	87	442	164	58	350	85
Mixed	746	27	30	761	30	29	547	24	21	552	28	16

Table 6: Selections by model specification for binomial outcomes

	Minimum measure						Most parsimonious within 2 of minimum					
Model	minAIC (AS3)			minBIC (AS4)			minAIC2 (AS3)			minBIC2 (AS4)		
specification	CMC	FMC	TMC	CMC	FMC	TMC	CMC	FMC	TMC	CMC	FMC	TMC
Under	100	5	5	133	9	8	148	11	10	171	12	12
Correct	357	499	569	481	673	741	541	761	830	598	847	892
Over	468	492	423	336	315	248	274	224	157	206	137	93
Mixed	75	4	3	50	3	3	37	4	3	25	4	3

Table 7: Selections by model specification for Poisson outcomes

	Minimum measure						Most parsimonious within 2 of minimum					
Model	minAIC (AS5)			minBIC (AS6)			minAIC2 (AS5)			minBIC2 (AS6)		
specification	CMC	FMC	TMC	CMC	FMC	TMC	CMC	FMC	TMC	CMC	FMC	TMC
Under	206	25	27	225	42	40	252	55	59	273	75	78
Correct	138	467	528	188	635	698	224	722	778	260	788	831
Over	393	481	422	335	302	238	298	207	150	253	121	79
Mixed	263	27	23	252	21	24	226	16	13	214	16	12

Table 8: Selections by model specification for gamma outcomes

	Minimum measure						Most parsimonious within 2 of minimum					
Model	minAIC (AS7)			minBIC (AS8)			minAIC2 (AS7)			minBIC2 (AS8)		
specification	CMC	FMC	TMC	CMC	FMC	TMC	CMC	FMC	TMC	CMC	FMC	TMC
Under	113	0	0	153	0	0	184	0	0	213	0	0
Correct	281	375	516	407	556	688	483	662	790	541	766	876
Over	606	625	484	440	444	312	332	338	210	245	234	123
Mixed	0	0	0	0	0	0	1	0	0	1	0	0

the CMC criteria. As expected, correct selections are uniformly highest for the TMC criteria. The mean AIC patterns for the selection criteria under the binomial, Poisson, and gamma distributions are similar, exhibiting strong protection from choosing underspecified and mixed misspecified models, modest protection for choosing overspecified models, and minimum means corresponding to the correctly specified models.

The Bernoulli distribution produces noticeably contrasting results for the CMC and FMC criteria. The FMC criteria choose the generating model specification much more frequently than the CMC criteria, which mostly choose mixed misspecified models. The figure supports the results reflected in the model selection counts. The means for the FMC criteria favor correctly specified and overspecified models, while the means for the CMC criteria favor underspecified and mixed misspecified models. The TMC criteria perform similarly to the FMC criteria, showing a slightly higher rate of correct model selections.

The binomial ($n = 10$) distribution model selection counts support the use of FMC criteria, which rarely choose an underspecified or mixed misspecified model. The figure exhibits similar patterns for the CMC and FMC means, reflecting more protection from choosing underspecified and mixed misspecified models for the FMC criteria. Once again, the TMC criteria slightly outperform the FMC criteria in terms of correct model selections.

The Poisson distribution counts show that the CMC criteria still tend to choose underspecified and mixed misspecified models more often than the FMC criteria. The FMC criterion selection counts reflect modest protection from choosing underspecified models, but not to the degree seen with the binomial distribution. The figure illustrates good protection from choosing underspecified and mixed misspecified models for both the CMC and FMC criteria, with higher means for the CMC. The FMC criteria again perform well relative to the TMC gold standards.

The gamma distribution model counts show no selections for underspecified or mixed misspecified models for the FMC criteria and no selections for mixed misspecified models for the CMC criteria. The figure indicates a similar pattern for the CMC and FMC criterion means as that which was seen with the Poisson distribution. The FMC criteria continue to perform well relative to the ideal TMC.

5 Application

Now that we have characterized how criteria based on the CMC, FMC, and TMC approaches behave in a simulated setting, we can apply the CMC and FMC criteria in a modeling application and examine the results. The application is based on a bullying study presented in Ramirez et al. (2016).

Bullying in schools is a pervasive and long-standing problem that has recently gained the attention of the media and public nationwide. Over the past several years, many efforts have been made to curb this behavior in students, including the passing of legislation to more specifically define bullying and to specify punishments that go along with offenses. In 2005, an anti-bullying law was passed in the state of Iowa in an effort to reduce the number of student-reported incidents of bullying.

The Iowa Youth Survey (IYS) is given statewide to sixth, eighth, and eleventh grade students in Iowa Public Schools every two to three years. The IYS covers an array of topics that are part of the everyday lives of students both inside and outside of school. A section of the survey contains questions regarding bullying, which is classified as four types: psychological (`psych`; experiences involving being excluded, ignored, or made the subject of lies/rumors), verbal (`verb`; experiences involving being made fun of or teased), physical (`phys`; experiences involving being hit, kicked, or shoved) and cyber (`cyber`; experiences involving being sent threatening or hurtful messages via email, social media, etc.).

The IYS data set considered for this application contains approximately 253,000 observations and 6 attributes, and includes surveys taken in 2005, 2008, and 2010. The binary outcomes of interest are significant exposure to each of the four bullying types, where bullying is defined as three or more incidents of victimization in the month prior to taking the survey. The candidate fixed effect predictor variables include survey year (`surveyyear`; categorical in [2005, 2008, 2010]), grade of the student (`grade`; categorical in $[6^{\text{th}}, 8^{\text{th}}, 11^{\text{th}}]$), gender of the student (`gender`; 0 = female, 1 = male), ethnicity of the student (`ethnicity`; 0 = white, 1 = African American, 2 = Native American, 3 = Asian, 4 = Hispanic, 5 = other/mixed), living situation of the student (`livingsituation`; 0 = with parent(s), 1 = with grandparent(s), 2 = with foster parent(s), 3 = in shelter care, 4 = group home, 5 = independent living, 6 = other), and frequency of teacher intervention in bullying incidents (`intervene`; 0 = almost never, 1 = once in a while, 2 = sometimes, 3 = often or almost always). Following the work of Ramirez et al. (2016), the ordinal variable `intervene` is treated as quantitative as opposed to qualitative. A random effect variable, school district (`schooldist`; categorical containing 412 unique values), will be included in all of the models. We chose to treat school district as a random effect in order to reduce the number of parameters in each model.

Since we have 6 potential fixed effect predictor variables, the all subsets setting will consider a candidate collection comprised of $2^6 - 1 = 63$ models. Once each model is fitted using both the CMC and FMC techniques, AIC and BIC will be calculated and used to ascertain the models favored by each criterion. The "best" model will be determined in the same way as in the simulation: minimum AIC (minAIC), most parsimonious model within two units of minimum AIC (minAIC2), minimum BIC (minBIC), and most parsimonious model within two units of minimum BIC (minBIC2). For the four CMC and FMC criteria, the selected models favored for each type of bullying are featured in Tables 9 and 10, respectively.

There is a noticeable difference in model selections between the CMC and FMC techniques. Under the CMC, the model selections for each outcome and criterion pairing are of order one. For each different bullying outcome, each of the four criteria select the same model. Under the FMC, the model selections for each outcome and criterion pairing are the full model with the exception of modeling verbal bullying and using the minBIC2 criterion, where the chosen model includes all candidate predictors except gender. These results strongly parallel the results from the investigative simulation presented earlier in the paper. Specifically, although we are inclined to believe that the FMC criteria may be selecting overspecified

Table 9: CMC model selections by criterion

Outcome	Criterion	Selected variables
Psych	minAIC	`surveyyear`
Psych	minAIC2	`surveyyear`
Psych	minBIC	`surveyyear`
Psych	minBIC2	`surveyyear`
Verb	minAIC	`gender`
Verb	minAIC2	`gender`
Verb	minBIC	`gender`
Verb	minBIC2	`gender`
Phys	minAIC	`surveyyear`
Phys	minAIC2	`surveyyear`
Phys	minBIC	`surveyyear`
Phys	minBIC2	`surveyyear`
Cyber	minAIC	`ethnicity`
Cyber	minAIC2	`ethnicity`
Cyber	minBIC	`ethnicity`
Cyber	minBIC2	`ethnicity`

Table 10: FMC model selections by criterion

Outcome	Criterion	Selected variables
Psych	minAIC	surveyyear grade gender ethnicity livingsituation intervene
Psych	minAIC2	surveyyear grade gender ethnicity livingsituation intervene
Psych	minBIC	surveyyear grade gender ethnicity livingsituation intervene
Psych	minBIC2	surveyyear grade gender ethnicity livingsituation intervene
Verb	minAIC	surveyyear grade gender ethnicity livingsituation intervene
Verb	minAIC2	surveyyear grade gender ethnicity livingsituation intervene
Verb	minBIC	surveyyear grade gender ethnicity livingsituation intervene
Verb	minBIC2	surveyyear grade ethnicity livingsituation intervene
Phys	minAIC	surveyyear grade gender ethnicity livingsituation intervene
Phys	minAIC2	surveyyear grade gender ethnicity livingsituation intervene
Phys	minBIC	surveyyear grade gender ethnicity livingsituation intervene
Phys	minBIC2	surveyyear grade gender ethnicity livingsituation intervene
Cyber	minAIC	surveyyear grade gender ethnicity livingsituation intervene
Cyber	minAIC2	surveyyear grade gender ethnicity livingsituation intervene
Cyber	minBIC	surveyyear grade gender ethnicity livingsituation intervene
Cyber	minBIC2	surveyyear grade gender ethnicity livingsituation intervene

models, the results are favorable compared to those for the CMC method, which tends to prefer simple models (often of order one) for outcomes following the Bernoulli distribution.

These selections may still raise questions as to which models are optimal; unlike the simulation study, the absence of a “true” model complicates definitive conclusions. As a supplementary comparison, two additional selection approaches are considered. First, we fit GLMMs by employing the more computationally expensive method that approximates the marginal likelihood using the Laplace method. This allows us to obtain AIC and BIC values where the goodness-of-fit term is based on the empirical marginal likelihood. Second, we fit ordinary GLMs by omitting the random intercept, resulting in traditional AIC and BIC values where the goodness-of-fit term is based on the empirical likelihood under independence.

The latter approach, which ignores clustering effects within school districts, can be justified since the contribution of the random intercept to the model fit is of debatable importance. To investigate the importance of this contribution, using Gaussian linear mixed models based on the pseudo-data, we obtained estimates of the variance for the random intercept and for the residual, and performed Wald tests with these estimates. Additionally, we inspected ratios of the variance for the random intercept over the variance for the residual, to assess the magnitude of the between-cluster variability relative to the within-cluster variability. These ratios were all less than 0.01, indicating that the practical importance of the clustering effect is quite modest. However, the Wald tests were highly significant, in part due to the large size of the sample.

Table 11: Model selections by criterion for GLMMs fit using the Laplace approximation and GLMs fit under independence

Outcome	Criterion	Selected variables
Psych	minAIC	surveyyear grade gender ethnicity livingsituation intervene
Psych	minAIC2	surveyyear grade gender ethnicity livingsituation intervene
Psych	minBIC	surveyyear grade gender ethnicity livingsituation intervene
Psych	minBIC2	surveyyear grade gender ethnicity livingsituation intervene
Verb	minAIC	surveyyear grade ethnicity livingsituation intervene
Verb	minAIC2	surveyyear grade ethnicity livingsituation intervene
Verb	minBIC	surveyyear grade ethnicity livingsituation intervene
Verb	minBIC2	surveyyear grade ethnicity livingsituation intervene
Phys	minAIC	surveyyear grade gender ethnicity livingsituation intervene
Phys	minAIC2	surveyyear grade gender ethnicity livingsituation intervene
Phys	minBIC	surveyyear grade gender ethnicity livingsituation intervene
Phys	minBIC2	surveyyear grade gender ethnicity livingsituation intervene
Cyber	minAIC	surveyyear grade gender ethnicity livingsituation intervene
Cyber	minAIC2	surveyyear grade gender ethnicity livingsituation intervene
Cyber	minBIC	surveyyear grade gender ethnicity livingsituation intervene
Cyber	minBIC2	surveyyear grade gender ethnicity livingsituation intervene

Reassuringly, the criterion selections were identical under the aforementioned GLMM and GLM approaches. The selected models for the four criteria are provided in Table 11. Of the 16 model selections featured, 13 match the previous results for the FMC technique. The remaining three models only differ in selection by the same single variable. This agreement indicates that the FMC criterion selections are quite appropriate, making them favorable to the CMC criterion selections. This finding further supports the conclusions drawn from the simulation study (Tables 5, 6, 7, and 8).

6 Conclusion

This paper has introduced a new technique for constructing GLMM criteria for model selection in the pseudo-likelihood framework. In SAS, the technique can be implemented using the GLIMMIX and MIXED procedures. Compared to the default criterion construction with the GLIMMIX procedure, the new technique shows considerable improvement in model selection, as illustrated in the simulation study featured in this paper. Under the assumption that the candidate collection contains the generating model, the criteria based on the new technique select the generating model much more frequently than analogous criteria based on the GLIMMIX default construction. Of course, selection criteria have been developed in estimation frameworks that employ other likelihood constructs, such as conditional AIC (cAIC; Vaida and Blanchard, 2005), and these will often outperform their pseudo-likelihood counterparts. This is particularly true in modeling applications where the pseudo-data is poorly described by the normal distribution, such as logistic regression based on a Bernoulli outcome. However, pseudo-likelihood estimation remains popular due to its computational efficiency and numerical stability, making the FMC criteria quite useful, especially for larger data sets.

A SAS macro has been written by the corresponding author that fits models with all possible subsets of predictor variables and selects an appropriate model based on criteria constructed using the new technique. This macro is generalizable for any set of p variables, providing the $2^p - 1$ fitted candidate models based on predictor subsets. The SAS macro is available upon request.

Future work involves the development of an R package that parallels the functionality of our SAS macro. We also hope to explore a comparison of the CMC and FMC criteria for additional distributions not considered in this paper. Additionally, we will further characterize the behavior of the CMC and FMC criteria based on GLMMs with more complex random effect structures than merely the random intercept.

Acknowledgments. We wish to thank our referees for their valuable feedback, which served to improve the original version of this manuscript.

References

AKAIKE, H. (1973). *Information theory and an extension of the maximum likelihood principle.* Akademia Kiado, Budapest. 2nd International Symposium on Information Theory, PETROV, B. N. and CSAKI, F. (eds.), p. 267–281.

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723.

BENGTSSON, T. and CAVANAUGH, J. E. (2006). An improved Akaike information criterion for state-space model selection. *Computational Statistics and Data Analysis* **50**, 2635–2654.

BOLKER, B. M., BROOKS, M. E., CLARK, C. J., GEANGE, S. W., POULSEN, J. R., STEVENS, M. H. H. and WHITE, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution* **24**, 127–135.

BOZDOGAN, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* **52**, 345–370.

BURNHAM, K. P. and ANDERSON, D. R. (2002). *Model Selection and Multimodel Inference.* Springer, New York.

CAVANAUGH, J. E. and NEATH, A. A. (1999). Generalizing the derivation of the Schwarz information criterion. *Communications in Statistics – Theory and Methods* **28**, 49–66.

CAVANAUGH, J. E. and SHUMWAY, R. H. (1997). A bootstrap variant of AIC for state-space model selection. *Statistica Sinica* **7**, 473–496.

DAVIES, S. L., NEATH, A. A. and CAVANAUGH, J. E. (2005). Cross validation model selection criteria for linear regression based on the Kullback-Leibler discrepancy. *Statistical Methodology* **2**, 249–266.

Dean, C. B. and Nielsen (2007). Generalized linear mixed models: a review and some extensions. *Lifetime Data Analysis* **13**, 497–512.

Haughton, D. M. A. (1988). On the choice of a model to fit data from an exponential family. *The Annals of Statistics* **6**, 342–355.

Hurvich, C. M., Shumway, R. H. and Tsai, C. L. (1990). Improved estimators of Kullback-Leibler information for autoregressive model selection in small samples. *Biometrika* **77**, 709–719.

Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297–307.

Ishiguro, M., Sakamoto, Y. and Kitagawa, G. (1997). Bootstrapping log likelihood and EIC, an extension of AIC. *Annals of the Institute of Statistical Mathematics* **49**, 411–434.

Kashyap, R. L. (1982). Optimal choice of AR and MA parts in autoregressive moving-average models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **4**, 99–104.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.

Kullback, S. (1968). *Information Theory and Statistics*. Dover, New York.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics* **22**, 79–86.

Leonard, T. (1982). Comments on A simple predictive density function, by M LeJeune and GD Faulkenberry. *Journal of the American Statistical Association* **77**, 657–658.

Linhart, H. and Zucchini, W. (1986). *Model Selection*. Wiley, New York.

Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics* **57**, 120–125.

Ramirez, M., ten Eyck, P., Peek-Asa, C., Onwuachi-Willig, A. and Cavanaugh, J. E. (2016). Effectiveness of Iowa's anti-bullying law in preventing bullying. *Injury Epidemiology* **3**, 1, 15.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *The Annals of Statistics* **8**, 147–164.

Shibata, R. (1981). An optimal selection of regression variables. *Biometrika* **68**, 45–54.

Shibata, R. (1997). Bootstrap estimate of Kullback-Leibler information for model selection. *Statistica Sinica* **7**, 375–394.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society Series B* **39**, 44–47.

Stone, M. (1979). Comments on model selection criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society Series B* **41**, 276–278.

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics* **A7**, 13–26.

Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models. *Mathematical Sciences* **153**, 12–18. (in Japanese).

Vaida, F. and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika* **92**, 351–370.

Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* **48**, 233–243.

Patrick Ten Eyck
Institute for Clinical and
Translational Science,
The University of Iowa,
Iowa City, IA, USA
E-mail: patrick-teneyck@uiowa.edu

Joseph E. Cavanaugh
Department of Biostatistics,
The University of Iowa,
Iowa City, IA, USA

Paper received: 18 July 2016.