Computational Statistics and Data Analysis 59 (2013) 13–27

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Efficient maximum likelihood estimation of multiple membership linear mixed models, with an application to educational value-added assessments

Andrew T. Karl[a,*], Yan Yang[a], Sharon L. Lohr[b]

[a] School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287-1804, United States
[b] Westat, 1600 Research Blvd., Rockville, MD, 20850, United States

ARTICLE INFO

Article history:
Received 31 January 2012
Received in revised form 4 October 2012
Accepted 4 October 2012
Available online 12 October 2012

Keywords:
EM algorithm
Generalized persistence model
Sparse matrices
Unbalanced data
Variable persistence model

ABSTRACT

The generalized persistence (GP) model, developed in the context of estimating "value added" by individual teachers to their students' current and future test scores, is one of the most flexible value-added models in the literature. Although developed in the educational setting, the GP model can potentially be applied to any structure where each sequential response of a lower-level unit may be associated with a different higher-level unit, and the effects of the higher-level units may persist over time. The flexibility of the GP model, however, and its multiple membership random effects structure lead to computational challenges that have limited the model's availability. We develop an EM algorithm to compute maximum likelihood estimates efficiently for the GP model, making use of the sparse structure of the random effects and error covariance matrices. The algorithm is implemented in the package GPvam in R statistical software. We give examples of the computations and illustrate the gains in computational efficiency achieved by our estimation procedure.

1. Introduction

Multilevel mixed models are popular for describing data with complex dependence structure. The units on which primary measurements are taken (usually those at the lowest level) each belong to one or more units at higher levels. In a nested (hierarchical) two-level model, each unit at the lowest level belongs to exactly one higher-level unit. In a multiple membership structure (Browne et al., 2001), a lower-level unit may be associated with multiple higher-level units. This structure is common with non-static populations, and we study multiple membership models in which a lower-level unit is sequentially associated with different higher-level units. Thus, a child in foster care may live with multiple families; a patient may see multiple doctors; a deer may visit multiple salt licks; a worker may have multiple employers; a person may attend multiple therapy groups; a student may have multiple teachers. Fielding and Goldstein (2006) describe multiple membership models and give examples of their use. The multiple membership structure induces a complex dependence structure in the data. Lower-level units are correlated whenever they share any higher-level unit, so the covariance matrix will not have a block diagonal structure as in the nested model.

The complex covariance structure of multiple membership mixed models makes computations challenging, particularly with large data sets. Computational methods that have been developed for nested hierarchical models and other special

* Corresponding author.
E-mail address: akarl@asu.edu (A.T. Karl).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2012.10.004

cases of linear mixed models often will not work. In this paper we develop an EM algorithm to compute maximum likelihood estimates for a class of longitudinal multiple membership models that are applicable in many settings. In the class of models considered, lower-level units are associated with multiple higher-level units in sequence, and a response is recorded on a lower-level unit after the association with each higher-level unit. If the population contains a large number of higher-level units, and the number of lower-level units associated with each higher-level unit is bounded, the covariance matrix will be sparse. The algorithm exploits sparseness of the covariance matrix to speed computations. This sparseness is achieved in many multiple membership settings since, for example, there are upper bounds on the number of patients a doctor can see or the number of students in a teacher's class.

The application motivating this research comes from value-added models (VAMs) in educational evaluation. A VAM score for a teacher is intended to estimate the "value added" by that teacher to students' knowledge—how much more (or less) students' scores changed under that teacher than they would be expected to change under an "average" teacher—by apportioning students' progress on standardized tests to the teachers or schools that have taught those students. Braun et al. (2010) describe some of the potential uses of VAMs and discuss issues associated with using them to evaluate teachers and schools.

While a variety of different models are used (see Lohr, 2012 for a review of common VAMs), in this paper we primarily consider the generalized persistence (GP) model developed by Mariano et al. (2010), one of the most flexible models in the literature. In the GP model, each student is followed over T grades with a different teacher in each grade, and receives a score on a standardized test at the end of each grade. Each student therefore "belongs" to up to T different teachers, resulting in a multiple membership structure. The GP model, like other mixed models used in the value-added context (Sanders et al., 1997; Rowan et al., 2002; McCaffrey et al., 2003, 2004, 2005; Lockwood et al., 2007), uses a longitudinal database of student scores and models the scores with random teacher intercepts. Under this scenario, the empirical best linear unbiased predictors (EBLUPs) for random teacher intercepts are the teacher VAM scores. In this paper we use the term "teacher effect" to represent the VAM score of a teacher but note that, as observed by Lockwood et al. (2007), these teacher effects measure "unexplained heterogeneity at the classroom level", and not necessarily the causal effect of the teacher.

The GP model is distinguished from others in the VAM literature by how it attributes a student's performance to current and prior teachers. If the effects of good teaching persist, one would expect that students of a good teacher in year 1 would do well on the test in year 1 and would continue to do well on the tests in future years. The Educational Value-Added Assessment System (EVAAS) model (Sanders et al., 1997), a complete persistence model, assumes that the effect of a teacher persists undiminished over all subsequent years on his or her students' achievement. This complete persistence assumption, also proposed by Raudenbush and Bryk (2002), implies that each teacher has one VAM score: the effect of a teacher in year g on his or her students' test scores is the same for their tests in each of years $g, \ldots, T$. The complete persistence assumption simplifies the covariance structure, and the EVAAS model is implemented in SAS software (Wright et al., 2010). A model proposed by McCaffrey et al. (2004) allows the effect of a teacher on students' scores to decay in future years, though the effects are otherwise perfectly correlated. Lockwood et al. (2007) refer to this structure as variable persistence (VP). In the VP model, each teacher has one estimated effect, but the impact on students' future year scores is reduced by a multiplicative factor in each year. The multipliers, called persistence parameters, are estimated from the data.

The GP model allows a much more general structure for the effects of a current teacher on future test scores. In the GP model, a teacher in year g has a different effect on his or her students' scores in each year from $t = g, \ldots, T$, and the $(T - g + 1)$ effects of that teacher have an unstructured covariance matrix to allow the effects to be correlated. The EVAAS model is a special case of the GP model in which the current and future effects of a teacher are assumed to be identical. The general correlation structure in the GP model allows much more detailed exploration of the patterns of teacher effects, but greatly complicates the problem of computing estimates.

Hill and Goldstein (1998) estimate a class of multiple membership models using an iterative generalized least squares algorithm, and Browne et al. (2001) employ Monte Carlo Markov chain techniques. Likewise, Mariano et al. (2010) use Bayesian methods to estimate the parameters for the GP model using data from a large urban school district. To obtain a proper posterior distribution, however, a Bayesian approach to computations requires that an informative prior distribution be adopted for the covariance parameters. As investigated in their paper, different priors often result in different estimates of model parameters and teacher effects. A maximum likelihood (ML) approach avoids the need for priors, although ML estimation of even the simpler VP model has been "practically infeasible for all but small data sets" (Lockwood et al., 2007) up to this point. In this paper we use the sparseness of the covariance and design matrices to develop an efficient EM algorithm for calculating ML estimates of parameters in the GP and VP models. We implement the method in the user-friendly GPvam package (Karl et al., 2012) in R statistical software (R Development Core Team, 2012). This development makes the GP and VP models more accessible for use in practice, and provides an alternative to the Bayesian calculations implemented by Mariano et al. (2010).

While the GP model was developed for educational applications, the model and the computational methods in this paper apply in many other settings as well. For example, Ash et al. (2012) note the similarity between the problems of evaluating teacher performance on the basis of student outcomes, and evaluating hospital and physician performance on the basis of patient outcomes. The multiple membership structure also arises in social network data (Airoldi et al., 2008). In another example, Browne et al. (2001) and Goldstein et al. (2000) describe a multiple membership model used to study Belgian household migration with complete persistence, measuring the propensity of individuals to change household membership. The GP model is a good candidate for the Belgian household data since the similarity of former roommates may decrease