ELSEVIER

Computer Methods and Programs in Biomedicine 49 (1996) 229–252

computer methods and programs in biomedicine

MIXREG: a computer program for mixed-effects regression analysis with autocorrelated errors

Donald Hedeker*[a], Robert D. Gibbons[b]

[a]*Division of Epidemiology and Biostatistics, (M/C 922), School of Public Health and Prevention Research Center, University of Illinois at Chicago, 2121 West Taylor Street, Room 510, Chicago, IL 60612-7260, USA*

[b]*Department of Psychiatry and School of Public Health, University of Illinois at Chicago, 912 South Wood Street, Chicago, IL 60680, USA*

Received 8 September 1995; accepted 29 January 1996

Abstract

MIXREG is a program that provides estimates for a mixed-effects regression model (MRM) for normally-distributed response data including autocorrelated errors. This model can be used for analysis of unbalanced longitudinal data, where individuals may be measured at a different number of timepoints, or even at different timepoints. Autocorrelated errors of a general form or following an AR(1), MA(1), or ARMA(1,1) form are allowable. This model can also be used for analysis of clustered data, where the mixed-effects model assumes data within clusters are dependent. The degree of dependency is estimated jointly with estimates of the usual model parameters, thus adjusting for clustering. MIXREG uses maximum marginal likelihood estimation, utilizing both the EM algorithm and a Fisher-scoring solution. For the scoring solution, the covariance matrix of the random effects is expressed in its Gaussian decomposition, and the diagonal matrix reparameterized using the exponential transformation. Estimation of the individual random effects is accomplished using an empirical Bayes approach. Examples illustrating usage and features of MIXREG are provided.

Keywords: Longitudinal data; Clustered data; Random effects; Correlated responses; Multilevel data; Random coefficients models

1. Introduction

Mixed-effects regression models are becoming increasingly popular for analysis of both clustered and longitudinal data, and several books have recently been written on this topic [1–3]. Common to both clustered and longitudinal data is the idea of nesting. In clustered data, subjects are observed and measured clustered or nested within a larger context, for example, a hospital, a school, a clinic, or a firm. In longitudinal data where individuals are repeatedly assessed, the measures are clustered or nested within individuals. In order to take the nesting of the data into account, models with random effects are typically employed. For clustered

* Corresponding author, Tel.: +1 312 9964896; E-mail: HEDEKER@UIC.EDU.

0169-2607/96/$15.00
PII: S0169-2607(96)01723-3

data the random effects represent the cluster effects, while for longitudinal data the random effects represent the subject effects. Since these data are typically unbalanced (numbers of observations per subject are unequal in the longitudinal context, or numbers of subjects per cluster are unequal in the clustered context), the use of standard statistical techniques like analysis of variance is problematic [4]. In this case, mixed-effects models using primarily maximum likelihood or generalized least squares estimation can be utilized to overcome the difficulty induced by the unbalanced design.

Variants of these mixed-effects regression models for unbalanced continuous data have been developed under a variety of names: random-effects models [5,6], variance component models [7], hierarchical linear models [2], multilevel models [1], two-stage models [8], random coefficient models [9], mixed models [10], empirical Bayes models [11], unbalanced repeated-measures models [12], and random regression models [13–17]. The approach to modeling unbalanced data presented in these articles generally involves linear regression models that allow some parameters in addition to the residuals to be random. As the name of this program implies, a mixed-effects regression model generally contains some fixed effects in the model in addition to the random effects.

For longitudinal data, mixed-effects models allow for presence of missing data (i.e., subjects are not assumed to be measured at the same number of timepoints), time-varying or invariant covariates, and subjects measured at different timepoints. Additionally, whereas, the traditional approaches to longitudinal data analysis estimate average change in a population, the mixed-effects approach can also estimate individual change for each subject. This is particularly useful in the biomedical setting where a proportion of subjects may respond to therapy in quite different ways from the average response.

In analysis of clustered data using mixed-effects models, outcomes at the individual level are modeled in terms of both individual and cluster level variables, while concurrently estimating and adjusting for the amount of intraclass correlation present in the data. Further, these models make no assumption regarding cluster sample size, allowing for a varying number of subjects within each cluster.

Much of the work on the mixed-effects regression model has utilized maximum marginal likelihood (MML) estimation for the structural and population parameters and empirical Bayes (EB) estimation for the individual random effects. Marginal maximum likelihood solutions using the EM algorithm have been presented by Dempster, Rubin, and Tsutakawa [7], Laird and Ware [5], and Bryk and Raudenbush [18], while Newton-Raphson and/or Fisher-scoring MML solutions have been proposed by Jennrich and Schluchter [12], Longford [10], Lindstrom and Bates [19], and Bock [8]. These approaches generally assume that, conditional on the random effects, the model residuals are uncorrelated across timepoints, although, Chi and Reinsel [20] and Hedeker [21] both presented a scoring solution for the mixed regression model with autoregressive (AR1) errors.

This paper describes the FORTRAN program MIXREG (mixed-effects regression) for the analysis of repeated or clustered normally-distributed response variables using MML and EB estimation procedures. MIXREG fits a mixed regression model with certain types of autocorrelated error structures (described below). Both an EM solution and a scoring solution are utilized. For the scoring solution, a reparameterization of the covariance matrix of the random effects is performed which utilizes the Gaussian decomposition of a matrix, transforming the resulting diagonal matrix using the exponential transformation. Two examples will illustrate the mixed-effects approach for longitudinal and clustered data, respectively.

Some commercially-based software exists to perform mixed-effects regression analysis, including HLM [22], ML3 [23], VARCL [24], the BMDP 5V procedure, and the SAS procedure MIXED. A detailed comparison of some of these programs is included in Kreft, de Leeuw, and Van Der Leeden [25]. In terms of MIXREG's unique features, MIXREG allows for autocorrelated errors (HLM, ML3, and VARCL do not), and treats time in a continuous manner (BMDP 5V does

not). Additionally, MIXREG is available for the MACINTOSH environment, and allows two autocorrelated error structures not offered by the other software programs: the non-stationary autoregressive error structure described by Mansour, Nordheim, and Rutledge [26], and the first-order mixed autoregressive-moving average error structure.

2. Computational methods

To describe the model in a general way for data which are either clustered or longitudinal, the terminology of multilevel analysis can be used [1]. For this, let i denote the level-2 units (clusters in the clustered data context, or subjects in the longitudinal data context), and let j denote the level-1 units (subjects in the clustered data context, or repeated observations in the longitudinal data context). Assume that there are $i = 1, \ldots N$ level-2 units and $j = 1, \ldots, n_i$ level-1 units nested within each level-2 unit. The mixed-effects regression model for the $n_i \times 1$ response vector y for level-2 unit i (subject or cluster) can be written as:

$$y_i = W_i\alpha + X_i\beta_i + e_i, \quad i = 1, \ldots N, \tag{1}$$

where W_i is a known $n_i \times p$ design matrix for the fixed effects, α is the $p \times 1$ vector of unknown fixed regression parameters, X_i is a known $n_i \times r$ design matrix for the random effects, β_i is the $r \times 1$ vector of unknown individual effects, and e_i is the $n_i \times 1$ error vector. The distribution of the random effects is assumed to be multivariate normal with mean vector μ and covariance matrix Σ and the errors are assumed to be independently distributed as multivariate normal with mean vector 0 and covariance matrix $\sigma^2_\epsilon\Omega_i$. Although Ω_i carries the i subscript, it depends on i only through its dimension n_i, that is, the number of parameters in Ω_i will not depend on i. In the case of independent residuals, $\Omega_i = I_i$, but for our purposes, we will define ω to be the $s \times 1$ vector of autocorrelation terms that Ω_i depends on.

Different types of autocorrelated errors are allowable in MIXREG including the first-order autoregressive process, AR(1), the first-order moving average process, MA(1), the first-order mixed autoregressive-moving average process, ARMA(1,1), and the general autocorrelation structure. A typical assumption in models with autocorrelated errors is that the variance of the errors is constant over timepoints and that the covariance of errors from differing timepoints depends only on the time interval between these timepoints and not on the starting timepoint. This assumption, referred to as the *stationarity* assumption, is assumed for the aforementioned forms. Another form of autocorrelated errors is described by Mansour, Nordheim, and Rutledge [26], who examine autocorrelated errors which follow the first order autoregressive process, however, where the assumption of stationarity is relaxed.

As a result of the above assumptions, the observations y_i and random coefficients β have the joint multivariate normal distribution:

$$\begin{bmatrix} y_i \\ \beta \end{bmatrix} \sim N\left(\begin{bmatrix} X_i\mu + W_i\alpha \\ \mu \end{bmatrix}, \begin{bmatrix} X_i\Sigma X_i' + \sigma^2_\epsilon\Omega_i & X_i\Sigma \\ \Sigma X_i' & \Sigma \end{bmatrix}\right)$$

The mean of the posterior distribution of β, given y_i, yields the empirical Bayes (EB) or EAP ('Expected A Posteriori') estimator of the level-2 parameters,

$$\hat{\beta}_i = [X_i'(\sigma^2_\epsilon\Omega_i)^{-1}\,X_i + \Sigma^{-1}]^{-1}\,X_i'(\sigma^2_\epsilon\Omega_i)^{-1} \times (y_i - X_i\mu - W_i\alpha) + \mu, \tag{2}$$

with the corresponding posterior covariance matrix given by

$$\Sigma_{\beta|y_i} = [X_i'(\sigma^2_\epsilon\Omega_i)^{-1}X_i + \Sigma^{-1}]^{-1}. \tag{3}$$

2.1. Maximum marginal likelihood solution

To obtain the maximum marginal likelihood (MML) solution for the population parameters (μ and Σ) and the structural parameters (α, σ^2_ϵ, and ω), the marginal density of the data y_i,

$$h(y_i) = \int_\beta f_i \cdot g\, d\beta, \tag{4}$$

is maximized. Here,

$$f_i = (2\pi)^{-n_i/2} |\sigma_\epsilon^2 \Omega_i|^{-1/2} \exp[-1/2 e_i' (\sigma_\epsilon^2 \Omega_i)^{-1} e_i], \tag{5}$$

with $e_i = y_i - X_i\beta_i - W_i\alpha$ and the prior distribution g is the normal density with mean μ and variance Σ, that is,

$$g = (2\pi)^{r/2} |\Sigma|^{-1/2} \exp[-1/2(\beta - \mu)' \Sigma^{-1} (\beta - \mu)] \tag{6}$$

The MML solution is then obtained by maximizing the log-marginal likelihood of the data from N subjects,

$$\log L = \sum_{i=1}^{N} \log[h(y_i)] \tag{7}$$

with respect to both the population parameters (μ and Σ) and the structural parameters (α, σ_ϵ^2, and ω). These equations are obtained as:

$$\hat{\mu} = \frac{1}{N} \sum_{i}^{N} \hat{\beta}_i \tag{8}$$

$$\hat{\Sigma} = \frac{1}{N} \sum_{i}^{N} (\hat{\beta}_i \hat{\beta}_i' + \Sigma_{\beta|y_i}) - \hat{\mu}\hat{\mu}' \tag{9}$$

$$\hat{\alpha} = \left[\sum_{i=1}^{N} W_i' \Omega_i^{-1} W_i \right]^{-1} \times \left[\sum_{i=1}^{N} W_i' \Omega_i^{-1} (y_i - X_i \hat{\beta}_i) \right] \tag{10}$$

$$\hat{\sigma}_\epsilon^2 = \left(\sum_{i=1}^{N} n_i \right)^{-1} \sum_{i=1}^{N} \mathrm{tr}[(\Omega_i^{-1}) (\hat{e}_i \hat{e}_i' + X_i \Sigma_{\beta|y_i} X_i')] \tag{11}$$

$$\hat{\omega} = \sigma_\epsilon^{-2} \left[\sum_{i=1}^{N} \frac{\partial \mathrm{vec}' \, \Omega_i}{\partial \omega} (\Omega_i^{-1} \otimes \Omega_i^{-1}) \times \frac{\partial \mathrm{vec} \, \Omega_i}{\partial \omega'} \right]^{-1} \times \sum_{i=1}^{N} \frac{\partial \mathrm{vec}' \, \Omega_i}{\partial \omega} \times \mathrm{vec}[\Omega_i^{-1} (\hat{e}_i \hat{e}_i' + X_i \Sigma_{\beta|y_i} X_i') \Omega_i^{-1}] \tag{12}$$

where, the 'vec' operator stacks the column vectors of a matrix one on top of each other to form-one large column vector, 'tr' denotes the trace of a matrix, and $\otimes$ denotes the Kronecker product.

The EM solution proceeds by assigning starting values for the structural and population parameters in order to estimate the individual statistics $\hat{\beta}_i$ and $\Sigma_{\beta|y_i}$ using equations (2) and (3), respectively. Then, these individual statistics are used with the above Eqs. (8), (9), (10), (11), and (12) to obtain improved parameter estimates. This process is repeated until convergence, which can be very slow, and so, at some point it is better to shift to a Newton algorithm. For this, MIXREG uses the Fisher scoring procedure which utilizes the first derivatives and expected values of the second derivatives to obtain improved parameter estimates. The negative of the expected values of the second derivatives corresponds to the information matrix. The form of the information matrix is illustrated below (the equations for the information matrix can be found in Bock [8] or Longford [3]). Since the off-diagonal information between the fixed component (μ and α) and the variance components (Σ, σ_ϵ^2, and ω) is zero, the information matrix can be inverted using the formulas for inversion of partitioned matrices (see Bock [27] page 38).

$$I = \begin{bmatrix} I(\mu) & & & & \\ I(\alpha,\mu) & I(\alpha) & & & \\ 0 & 0 & I(\Sigma) & & \\ 0 & 0 & I(\sigma_\epsilon^2, \Sigma) & I(\sigma_\epsilon^2) & \\ 0 & 0 & I(\omega, \Sigma) & I(\omega, \sigma_\epsilon^2) & I(\omega) \end{bmatrix}$$

At convergence, the inverse of the information matrix provides the large-sample variances and covariances of the MML estimators which can be used to construct confidence intervals and tests of

hypotheses for the structural and population parameters.

Although the Fisher scoring solution is a significant improvement in terms of speed of convergence over the EM solution used by Laird and Ware [5] and others, it can fail in the estimation of the covariance matrix of the random effects as these terms become very small. Thus, Lindstrom and Bates [19] suggest reparameterizing the variance covariance matrix Σ in terms of the Cholesky factorization, however, a better choice is to reparameterize in terms of the Gaussian factorization of a symmetric matrix (see Bock [27] pages 82–84), utilizing the exponential transformation for the diagonal matrix D corresponding to the variance parameters. Specifically, consider

$$\Sigma = LDL' = L \exp(\Pi)\, L'$$

where, in the case of $r = 3$, we get

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} e^{\pi_1} & 0 & 0 \\ 0 & e^{\pi_2} & 0 \\ 0 & 0 & e^{\pi_3} \end{bmatrix} \begin{bmatrix} 1 & l_{21} & l_{31} \\ 0 & 1 & l_{32} \\ 0 & 0 & 1 \end{bmatrix}$$

The parameters to be estimated are now the diagonal elements in Π and the elements of L below the main diagonal. As a result of the exponential transformation the range of permissible values for Π includes all real numbers, and so, a variance component which is near zero can be approached much easier in the iterative estimation scheme. The additional step of transforming the diagonal elements via the exponential transformation is crucial in allowing the algorithm to estimate variances near zero, and this is the main distinction between our reparameterization and the Cholesky square root transformation used by Lindstrom and Bates [19]. In MIXREG, the scoring solution begins by estimating the reparameterized variance terms in Π and L, and at convergence, performs a final iteration to produce estimates of the unparameterized variance terms in Σ.

At convergence, the MML estimates and their accompanying standard errors can be used to construct asymptotic z-statistics by dividing the parameter estimate by its standard error [28]. The computed z-statistic can then be compared with the standard normal table to test whether the parameter is significantly different from zero. While this use of the standard errors to perform hypothesis tests (and construct confidence intervals) for the fixed effects (μ and α) is generally reasonable, for the variance and covariance components (Σ, ω, and σ_ϵ^2) this practice is problematic (see Bryk and Raudenbush [2] page 55).

Instead, in order to test hypotheses related to the variance and covariance components, as well as the fixed effects, the likelihood-ratio (or difference in log-likelihood) χ^2 test can be used for comparison of nested models. The significance of the additional terms in model A over model B is determined by comparing $-2(\log L_B - \log L_A)$ to a table of the χ^2 distribution with degrees of freedom equal to the number of additional parameters in model A.

3. Program description and usage

MIXREG is currently available in executable form for both MS-DOS and MACINTOSH computers. In the MS-DOS environment, MIXREG can be run in either batch (MIXREGB.EXE) or interactive mode (MIXREG.EXE), while in the MACINTOSH environment only the batch mode (MIXREGB) is possible. For batch processing, the MIXREG instructions must be stored in the file MIXREG.DEF (described below), while in interactive mode the user can specify the various op-

tions using the menu-orientated user interface. Here, we will discuss the procedure for running the program primarily in batch mode. In either batch or interactive processing, MIXREG makes use of the following files:

- input data file;
- MIXREG.DEF — main definition file for analysis options and settings;
- main output file;
- Empirical Bayes estimates results file.

An additional definition file which can be used to save or retrieve analysis options and settings is used in interactive mode only, and is described more fully below.

3.1. Structure of the input data file

This file contains all data (i.e., responses and covariates) read by the program. This file is read in free format and must be a standard text (ASCII) file with no hidden characters or word processing format codes. Variable fields must be separated by one or more blanks. The data are assumed to consist of multiple level 1 observations within a higher-order (2nd level) unit, for example, in the longitudinal data setting there are repeated observations (level 1) within individuals (level 2). There must be a level-2 ID variable for each record and the data must be sorted by this level-2 ID variable. If autocorrelated errors are to be estimated (or fixed at a non-zero value), then the file must additionally be sorted by level-2 ID and time. The repeated measurements of an individual take up as many records in this file as there are measurements for that individual. Thus some individuals will have, for example, four records while others may have two or five records. Alternatively, if missing value codes are utilized, each individual may have data on the same number of records, but some records will contain missing value codes for some (or all) of the variables. Similarly, in the clustered data setting there are nested observations (students, employees, patients) within clusters (schools, firms, clinics). The nested measurements (level 1) of a cluster (level 2) take up as many records in this file as there are level 1 units within that cluster. Thus, some clusters will have, for example, 40 records while others may have 20–50 records.

The fields of variables that are read in, separated by one or more blanks, on a line (or lines) are as follows (the order of variables does not matter):

ID DepVar Xvector Wvector

where, in the longitudinal context, *ID* refers to an individual ID number which does not change across timepoints, *DepVar* is the value of the dependent measure at the particular timepoint, *Xvector* is the part of the design matrix for the random effects at the given timepoint, and *Wvector* is the covariate vector at the timepoint. In the clustered data context, *ID* refers to the cluster ID number which does not change across nested observations, *DepVar* is the value of the dependent measure, *Xvector* is the part of the design matrix for the random effects, and *Wvector* is the covariate vector; all given for each clustered observation. All variables are read as REAL*8 with the exception of the (level-2) IDs which are read as INTEGER. All missing data must have a *numeric* missing value code, in particular, *missing values left as blank fields will definitely cause problems.*

3.2. Analysis options and settings — MIXREG.DEF

This file contains the information to determine which statistical model should be fit to the data in the input data file. Although a word processor can be used to create this file, it must be saved as a standard text (ASCII) file with no hidden characters or word processing format codes. The analysis options and settings that comprise this file are described in Table 1 (a–e).

Except where noted, this file is read in free format. In batch processing, this file is created by the user directly *before* typing the command MIXREGB (or on the MACINTOSH, double-clicking on the MIXREGB file), while in interactive mode (typing the command MIXREG), this file is created using the menu-orientated user interface. For batch mode, this filename and extension

Table 1
Analysis options and settings specified in MIXREG.DEF

(a) Lines 1–6

Line 1 — A title of 60 characters
Line 2 — A subtitle of 60 characters
Line 3 — Name of input data file. Any legal filename of 80 characters or less can be specified.
Line 4 — Name of main output file. Any legal filename of 80 characters or less can be specified.
Line 5 — Name of definition file to be saved or retrieved. Any legal filename of 80 characters or less can be specified. Note that a name for this file must be specified even in batch processing, although in batch processing nothing is done to this file. In interactive mode, after a filename is entered in the appropriate menu field, the program settings and options of the specified file will be retrieved if that file exists. Prior to running the statistical procedure from interactive mode, the current MIXREG options and settings selected by the user will be saved into the specified file.
Line 6 — Name for empirical Bayes estimation results file. Any legal filename of 80 characters or less can be specified. Note that a filename must be specified even if the option for writing out the empirical Bayes estimates (IRES on line 7) is set to 0.

(b) Line 7

Line 7 — NPR NEM NFIELD R P CONV MISS START MEANS IFIN IRES AUTO

NPR = number of level-2 units whose data will be listed on the screen (*usually set to 1*).
NEM = number of EM iterations to perform (*usually set to 10 or 20*).
NFIELD = number of fields of data to read from the input data file.
R = number of random effects.
P = number of fixed effects.
CONV = convergence criterion (*usually set to 0.001 or 0.0001*).
MISS = 0 if no missing values are present in the data, or 1 if missing values are present (codes for which will later be defined).
START = 0 if automatic starting values are to be used, or 1 if user-defined starting values are to be used.
MEANS = 0 if a table of means of the dependent variable broken down by values of another variable is not requested, and 1 if such a mean table is requested.
IFIN = 0 to reparameterize the variance terms during the Fisher scoring solution, or 1 not to reparameterize (*usually set to 0*).
IRES = 0 to avoid writing out the level-2 unit Bayes estimates to a results file, or 1 to write out the level-2 unit Bayes estimates to a results file.
AUTO = 0 for no autocorrelation terms, 1 to fix the autocorrelation terms and estimate all other parameters, and 2 to estimate all terms including the autocorrelation terms.

(c) Lines 8–10

Line 8 — two parameters are to be read on this line: the field of the input data file which contains the (level-2) IDs, followed by the field of the input data file which contains the dependent variable.
Line 9 — R parameters are to be read on this line: the field(s) of the input data file which contain(s) the R random effects.
Line 10 — P parameters are to be read on this line: the field(s) of the input data file which contain(s) the P fixed effects.

(d) Lines after line 10

next line (*if MEANS = 1*) — two parameters and a list of values: the field of the input data file which contains the variable for which means of the dependent variable are to be broken down by, followed by the number of levels of this variable, and a list of the values for all of these levels.
next line (*if MISS = 1*) — missing value code for the dependent variable.
next line (*if MISS = 1*) — R missing value codes for the random-effect variables.
next line (*if MISS = 1*) — P missing value codes for the fixed covariates.
next line — an eight character label for the dependent variable.
next line — R labels for the random effects in 8 character width fields.
next line (*if START = 1*) — R starting values for the means of the random effects.
next line — P labels for the covariates in eight character width fields (a maximum of ten labels per line).
next line (*if START = 1*) — P starting values for the covariate effects.

Table 1 (*Continued*)

next line (*if START = 1*) — ((R * (R+ 1))/2) starting values for the variance and covariance terms of the random effects given in 'packed' form, e.g., for a 3 × 3 covariance matrix, the order of the starting values should be: variance (1), covariance (1,2) variance (2), covariance (1,3), covariance (2,3), and variance (3).

next line (*if START = 1*) — starting value for the error variance.

(e) Remaining lines

The following lines are specified only if AUTO equals 1 or 2.

next line — two parameters and a list of values: the field of the input data file which contains the 'time' variable for which autocorrelated errors are to be generated from, the maximum number of timepoint values that this 'time' variable can take on in the data, and a list of these timepoint values (in increasing order).

next line — two parameters: NS and S.

NS = 0 for stationary AR(1), 1 for non-stationary AR(1) (see reference [26]), 2 for stationary MA(1), 3 for stationary ARMA(1,1), and 4 for a general autocorrelation (Toeplitz) structure.

S = number of autocorrelation terms, which should equal 1 for NS = 0,1,2; 2 for NS = 3; or an integer (≥ 1 and < maximum number of timepoints) for NS equal to 4.

final line (*if START = 1*) — starting values for the S autocorrelation terms.

(MIXREG.DEF) *must* be used and should be in the same directory as the program MIXREGB.EXE or accessible via appropriate PATH statements. In interactive mode, the creation and storage of this file is done by the program.

3.3. Main output file

This file contains descriptive information about the variables as well as results of the specified analysis. Examples of the output file provided below illustrate the contents of this file. In terms of numbers of observations, the number of level-2 units, the total number of level-1 units, and the number of level-1 units for each level-2 unit are all listed. For each variable (except the ID variable), the following descriptive statistics are provided: minimum, maximum, mean, and standard deviation. These descriptive statistics are based on the total number of level-1 observations. An optional listing of means, standard deviations, and sample sizes of the dependent variable by the levels of one of the model covariates may be requested. Starting values, either user-defined or program-generated, are listed for all model parameters.

In terms of the program results, for each model parameter, maximum marginal likelihood estimates, standard errors, *z*-values, and *p*-values are provided. These *p*-values are two-tailed, except for the variance parameters where one-tailed *p*-values are given. As noted earlier, use of the standard errors to perform hypothesis tests for the variance and covariance parameters is controversial (see Bryk and Raudenbush [2] page 55). Number of iterations (both EM and Fisher-scoring iterations) necessary to achieve convergence is listed, as well as the value of the log-likelihood at convergence. The log-likelihood value can be used to perform likelihood-ratio tests. Following the parameter estimates (and associated statistics), MIXREG lists re-expressions of the estimated random-effects variance terms, depending on the type of model specified. If a random-intercepts model is specified then MIXREG calculates and lists the value of the estimated intra-cluster correlation. If a model with more than one random effect is specified, MIXREG expresses the estimated random-effects variance-covariance matrix as a correlation matrix. Finally, correlation matrices are also provided for the estimates of all model parameters. It is important to realize that these correlation matrices are not correlations of the variables themselves, but correlations of the estimated model parameters. These matrices may be helpful in determining the degree to which colinearity is present in terms of the model parameters.

3.4. Empirical Bayes estimates results file

An optional results file is created by MIXREG which contains the empirical Bayes estimates of

the R random effects. If this option is specified, then for each level-2 unit, four pieces of information are given in the following order: (1) the level-2 unit's ID, (2) the number of level-1 observations associated with that level-2 unit, (3) the R empirical Bayes estimates, and (4) the corresponding $R \times R$ posterior variance covariance matrix (in 'packed' form) for these level-2 unit estimates. Note, a 3×3 variance-covariance matrix in 'packed' form is in the following order: variance (1), covariance (1,2), variance (2), covariance (1,3), covariance (2,3), and variance (3). Example 4.1.2. will illustrate specification of this option and the structure of the results file.

3.5. Notes on estimation of autocorrelated errors

There are a few practical considerations in estimating models with autocorrelated error terms. The EM equations are currently programmed for either non-autocorrelated errors or for AR1 errors. Thus, to estimate any other autocorrelated error forms, the user should provide good starting values for all parameters (a good choice is from the model with no autocorrelation) and specify NEM = 0 on line 6. This will then proceed directly with the Fisher scoring solution without any EM iterations. Also, for the correct estimation of the autocorrelation terms the program needs to know which field of the datafile contains the 'time' vector and how many potential values this variable can take on, in this way the program can determine which of all possible timepoints a given individual was measured on. Thus, if there are n potential 'time' values, the autocorrelation matrix from which the estimation is based on is of size $n \times n$. This matrix serves as the reference matrix from which rows and columns are removed for each individual in the presence of missing or non-observed data at specific timepoints for that individual. For instance, if the potential timepoints have values of 0, 1, 2, 3, and 4, then the autocorrelation matrix is of size 5×5 however, if an individual is not observed at time values 1 and 3, then elements in rows 2 and 4 are removed for that individual (since the individual's data provide no information regarding the autocorrelation involving these two time values).

3.6. Some common MIXREG errors

There are a few errors which can prevent MIXREG from running correctly, or even running at all. First, missing values left as blank fields, and not given a specified numeric missing value code, may cause the program to fail or to estimate a model which is incorrect from the user's perspective. To see if this is causing a problem, the user can check the correctness of each variable's descriptive statistics (minimum, maximum, mean, and standard deviation) listed in the output file. If these descriptive statistics are incorrect, the data are not being read into the program correctly and a common reason is that missing values are being left as blank fields in the data file. Second, the MEANS = 1 option is fairly unforgiving at this point. The actual values which must be listed for the levels of the variable *must* be exactly the same as the values that are found in the data file. If a strange error prevents MIXREG from running and this option is selected, the user can set MEANS = 0 to avoid this option. Third, the NPR option (described below), which is used to list data to the screen, can cause MIXREG to stop in certain cases (essentially, when the number of digits to be listed for a variable exceeds the format specification of the program). If the program stops after indicating (on the screen) the number of random and fixed effects in the model, but prior to listing any iterative results to the screen, the user can set NPR = 0 and re-run the program. Finally, if the program 'blows up', it may be that the model that is specified is not estimable. In this case, the user should try fitting a less complicated model by specifying fewer random effects, or fewer covariates. If the number of random effects is 1, and problems still exist, it may be that the random-effect variance cannot be reliably estimated as being different from zero. In this case, a fixed effects model may be warranted.

4. Examples of mixed-effects regression

MIXREG can estimate a variety of models for clustered and longitudinal data. For longitudinal data, an attractive feature of MIXREG is its ability to allow and estimate many forms of autocor-

related errors. An analysis of a longitudinal psychiatric dataset where patients are assessed weekly in terms of their severity of depression is presented first to illustrate features of mixed regression analysis of longitudinal data with autocorrelated errors. This example will allow for individual trends across time as well as examine the influence of a between-subjects grouping factor (endogenous vs. non-endogenous depression) on these trends. A second example, where students are observed nested within classrooms and schools, will illustrate the use of MIXREG for clustered data. These two examples will serve to highlight some of the results that are obtained from mixed-effects analysis, and will be accompanied by listings of specific file setups that are used to run MIXREG.

4.1. Longitudinal analysis with random effects and autocorrelated errors

A study described in Gram et al. [29] and Reisby et al. [30] focused on the longitudinal relationship between imipramine (IMI) and desipramine (DMI) plasma levels and clinical response in 66 depressed inpatients (37 endogenous and 29 non-endogenous). Following a placebo period of 1 week, patients received 225 mg/day doses of imipramine for 4 weeks. In this study, subjects were rated with the Hamilton depression rating scale (HDRS) twice during the baseline placebo week (at the start and end of this week) as well as at the end of each of the four treatment weeks of the study. Plasma level measurements of both IMI and its metabolite DMI were made at the end of each week. The sex and age of each patient was recorded and a diagnosis of endogenous or non-endogenous depression was made for each patient. Although the total number of subjects in this study was 66, the number of subjects with all measures at each of the weeks fluctuated: 61 at week 0 (start of placebo week), 63 at week 1 (end of placebo week), 65 at week 2 (end of first drug treatment week), 65 at week 3 (end of second drug treatment week), 63 at week 4 (end of third drug treatment week), and 58 at week 5 (end of fourth drug treatment week).

For this illustration, we focus on one aspect of the study; namely, is there evidence of differential improvement across time between endogenous and non-endogenous patients. Additional analysis of this dataset can be found in Gibbons, Clark, and Davis [31] and Hedeker et al. [15]. The model fit to the HDRS scores included a random intercept and linear trend across time to allow for differential trends for individual patients, as well as a group effect (endogenous or non-endogenous) and a group by time interaction to examine whether these two groups of patients differed in terms of their initial severity and improvement across time. In terms of the error structure, we will first consider uncorrelated errors and then allow for residual autocorrelation in the model by specifying a stationary AR1 process for the errors. A matrix representation of the model for subject i is given by

$$\begin{bmatrix} \mathrm{HDRS}_{i0} \\ \mathrm{HDRS}_{i1} \\ \mathrm{HDRS}_{i2} \\ \mathrm{HDRS}_{i3} \\ \mathrm{HDRS}_{i4} \\ \mathrm{HDRS}_{i5} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \beta_{i0} \\ \beta_{i1} \end{bmatrix} + \begin{bmatrix} \mathrm{ENDOG}_i & \mathrm{ENDOG}_i \times 0 \\ \mathrm{ENDOG}_i & \mathrm{ENDOG}_i \times 1 \\ \mathrm{ENDOG}_i & \mathrm{ENDOG}_i \times 2 \\ \mathrm{ENDOG}_i & \mathrm{ENDOG}_i \times 3 \\ \mathrm{ENDOG}_i & \mathrm{ENDOG}_i \times 4 \\ \mathrm{ENDOG}_i & \mathrm{ENDOG}_i \times 5 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} + \begin{bmatrix} e_{i0} \\ e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \\ e_{i5} \end{bmatrix}$$

Here, the grouping variable (ENDOG) is a dummy-coded variable indicating whether a subject is endogenous (ENDOG = 1) or non-endogenous (ENDOG = 0). In terms of trend in HDRS scores across time, the model includes a linear effect of week and a group by (linear) week interaction. Note that although the above matrix representation is for a subject with data at all six

Table 2
Data from example 4.1: first ten subjects

101	26	1	0	0	0
101	22	1	1	0	0
101	18	1	2	0	0
101	7	1	3	0	0
101	4	1	4	0	0
101	3	1	5	0	0
103	33	1	0	0	0
103	24	1	1	0	0
103	15	1	2	0	0
103	24	1	3	0	0
103	15	1	4	0	0
103	13	1	5	0	0
104	29	1	0	1	0
104	22	1	1	1	1
104	18	1	2	1	2
104	13	1	3	1	3
104	19	1	4	1	4
104	0	1	5	1	5
105	22	1	0	0	0
105	12	1	1	0	0
105	16	1	2	0	0
105	16	1	3	0	0
105	13	1	4	0	0
105	9	1	5	0	0
106	21	1	0	1	0
106	25	1	1	1	1
106	23	1	2	1	2
106	18	1	3	1	3
106	20	1	4	1	4
106	−9	1	5	1	5
107	21	1	0	1	0
107	21	1	1	1	1
107	16	1	2	1	2
107	19	1	3	1	3
107	−9	1	4	1	4
107	6	1	5	1	5
108	21	1	0	1	0
108	22	1	1	1	1
108	11	1	2	1	2
108	9	1	3	1	3
108	9	1	4	1	4
108	7	1	5	1	5
113	21	1	0	0	0
113	23	1	1	0	0
113	19	1	2	0	0
113	23	1	3	0	0
113	23	1	4	0	0
113	−9	1	5	0	0
114	−9	1	0	0	0
114	17	1	1	0	0
114	11	1	2	0	0
114	13	1	3	0	0
114	7	1	4	0	0
114	7	1	5	0	0
115	−9	1	0	1	0
115	16	1	1	1	1
115	16	1	2	1	2
115	16	1	3	1	3
115	16	1	4	1	4
115	11	1	5	1	5

timepoints, the data from all subjects need not be complete. If a subject's data are missing or incomplete, then they simply would contain less than six observations in the dataset, or their missing data would be coded with numeric missing value codes, which would be identified in the MIXREG.DEF file prior to the statistical analysis. For this problem, a subset of the data are contained in Table 2.

The order of the variables is: the subject's ID, HDRS score, a column of ones for the intercept, the value of week, the subject's endogenous code, and the product of endogenous and week. Notice that there are missing value codes (−9) for some subjects at specific timepoints — the data from these timepoints will not be used in the analysis, however data from these subjects at other timepoints where there are no missing data will be used in the analysis. Thus, for inclusion into the analysis, a subject's data (both the dependent variable

Table 3
MIXREG.DEF file for example 4.1.1

```
Riesby Data — HAMD scores across six timepoints
Two random (int & week) and two fixed (endog & endog*week)
riesby. dat
riesby.out
riesby.def
riesby.res
1 10 6 2 2 0.0001 1 0 1 0 1 0
  1 2
  3 4
  5 6
  4 6 0 1 2 3 4 5
-9
-9 -9
-9 -9
HAMD
InterceptWeek
Endog Endog*Wk
```

Table 4
Output file for example 4.1.1

MIXREG — The program for mixed-effects linear regression analysis
Riesby Data — HAMD scores across six timepoints
Two random (int & week) and two fixed (endog & endog*week)

Numbers of observations

Level 2 observations = 66
Level 1 observations = 375

The number of level 1 observations per level 2 unit are:

6 6 6 6 5 5 6 5 5 5 6 5 6 6 6 6 6 6 6
6 6 5 6 6 6 5 4 6 6 5 6 6 6 5 5 5 6 5
6 6 6 5 6 6 6 6 5 6 6 6 5 5 6 6 5 6 6
6 6 6 6 5 6 6 6 6

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	S.D.
HAMD	0.00000	39.00000	17.63733	7.19006
Intercpt	1.00000	1.00000	1.00000	0.00000
Week	0.00000	5.00000	2.48000	1.68320
Endog	0.00000	1.00000	0.54667	0.49848
Endog*Wk	0.00000	5.00000	1.35200	1.74553

Descriptives for response variable HAMD by the variable Week

HAMD

Week	Mean	S.D.	*N*
0.000	23.44262	4.53330	61
1.000	21.84127	4.69800	63
2.000	18.30769	5.48556	65
3.000	16.41538	6.41505	65
4.000	13.61905	6.97097	63
5.000	11.94828	7.21942	58

Starting values

Mean	22.5181	−2.3784	
Covariates	1.9741	−0.0454	
Var. terms	34.7212	0.0000	17.3606
Residual	34.7212		

Final results — maximum marginal likelihood (MML) estimates

EM iterations = 10
Fisher iterations = 4
Total iterations = 14

Log likelihood = −1107.465

Variable	Estimate	S.E.	*z*	*P*-value
Intercpt	22.47626	0.79435	28.29529	0.00000
Week	−2.36569	0.31181	−7.58695	0.00000
Endog	1.98802	1.06905	1.85962	0.06294
Endog*Wk	−0.02706	0.41947	−0.06450	0.94857

Table 4 (*Continued*)

Random-effect variance covariance term(s)				
Intercpt	11.64122	3.29645	3.53144	0.00021
Covariance	−1.40161	1.00337	−1.39690	0.16244
Week	2.07707	0.50379	4.12286	0.00002
Residual variance				
	12.21847	1.10708	11.03670	0.00000

note: *P*-values are two-tailed except for variances which are one-tailed

Random-effect covariances expressed as correlations

	1 Intercpt	2 Week
1 Intercpt	1.0000	
2 Week	−0.2850	1.0000

Correlation of the MML estimates of the fixed terms

	1 Intercpt	2 Week	3 Endog	4 Endog*Wk
1 Intercpt	1.0000			
2 Week	−0.4510	1.0000		
3 Endog	−0.7430	0.3351	1.0000	
4 Endog*Wk	0.3352	−0.7433	−0.4571	1.0000

Correlation of the MML estimates of variance-related terms

	1 Var-Covl	2 Var-Cov2	3 Var-Cov3	4 Residual
1 Var-Covl	1.0000			
2 Var-Cov2	−0.6008	1.0000		
3 Var-Cov3	0.2287	−0.5979	1.0000	
4 Residual	−0.1891	0.1728	−0.1400	1.0000

and all model covariates being used in a particular analysis) at a specific timepoint must be complete. The number of repeated observations per subject then depends on the number of timepoints for which there are non-missing data for that subject.

As mentioned, two models will be fit to these data: the first model does not allow for autocorrelated errors, while the second model fits an AR1 process to the errors. In the MIXREG.DEF and associated output listings that follow, the 'Intercpt' term represents the initial level of severity (at week 0) as measured by the Hamilton depression rating scale, 'Week' term represents the weekly change in Hamilton depression scores across the six timepoints of the study, the 'Endog' term is a dummy-coded variable indicating whether group membership (0 = non-endogenous and 1 = endogenous) influences the initial severity level, and 'Endog*Wk' is the group by time interaction.

4.1.1. Random-effects and uncorrelated errors

The MIXREG.DEF file for the model without autocorrelated errors is listed in Table 3. Notice, that even though missing values are coded only for the dependent variable in the input data file, numeric missing value codes must be specified in the MIXREG.DEF file for all model terms (if MISS = 1). In this case, the value-9 was specified

for all variables since for the dependent variable this value is the correct missing value code, while for all other model terms (intercept, week, endogenous, and endogenous by week) this value was never observed. Other options that are specified in this DEF file: ten EM iterations, a calculation of descriptive statistics by the different levels of the week variable (MEANS = 1), and listing of the empirical Bayes subject-effects to a file (IRES = 1).

The results for the model without autocorrelated errors are written to the file RIESBY.OUT and are listed in Table 4. Considering the estimated fixed effects (the population mean vector μ and the covariate effects α), the initial level of severity for non-endogenous patients is approximately 22.5 (μ_0) on the HDRS, while endogenous patients start about 2 units higher (α_0). This difference in initial severity is marginally significant ($P < 0.063$). The reason that the intercept and endogenous effect reflect HDRS levels at week 0 is due to the coding of week that was used, namely, 0–5. The meaning of the intercept is the predicted value of the dependent variable when all independent variables equal 0, while the meaning of the endogenous effect in this model is the predicted group difference when the week variable equals 0. Using other codings of week, for example, orthogonal polynomials, would change the meaning of these regression coefficients.

Both groups exhibit an overall weekly rate of improvement of roughly 2 units (μ_1) which is highly significant, and there is no evidence that the groups differ in terms of this rate of improvement (α_1). In terms of the random-effect variance and covariance terms, there is significant individual variation in both patients' initial severity and rate of improvement, however, there does not seem to be any significant overall covariation between these two terms.

4.1.2. Random-effects and autocorrelated errors

As previously mentioned, the most reliable way of fitting models with autocorrelated errors is to bypass the EM iterations, proceeding directly to the Fisher-scoring iterations, using the model estimates from the uncorrelated error model as starting values. In addition to the settings from the previous run, the MIXREG.DEF file listed in Table 5 thus specifies zero EM iterations, the inclusion of starting values, and residual AR1 autocorrelation. Notice also that new names are given for the OUT, DEF, and RES files so that the previous files are not accidentally over-written.

The starting values for the model parameters agree with the results from the previous analysis, with the exception that for the AR1 parameter the arbitrary value of 0.1 was chosen as the starting value. Results from this analysis are listed in Table 6.

The addition of the AR1 parameter to the model is significant by the likelihood ratio test ($\chi^2 = -2\ [-1107.465-(-1103.419)] = 8.1$ $df = 1$, $P < 0.005$). The estimated stationary autocorrelation is moderately large ($\hat{\rho} = 0.37$), indicating a fair amount of autocorrelation among the model residuals. Relative to the model without autocorrelated errors, the estimated values for the coefficients of the fixed covariates α and the population mean vector μ change slightly in this model;

Table 5
MIXREG.DEF file for example 4.1.2

```
Riesby data-HAMD scores across six timepoints — AR1 errors
Two random (int & week) and two fixed (endog & endog*week)
riesby.dat
riesbyar.out
riesbyar.def
riesbyar.res
1  0  6  2  2  0.0001  1  1  1  0  1  2
   1  2
   3  4
   5  6
   4  6  0  1  2  3  4  5
  -9
  -9  -9
  -9  -9
HAMD
IntercptWeek
22.476   -2.366
Endog   Endog*Wk
1.988   -0.027
11.641   -1.402   2.077
12.218
  4  6  0  1  2  3  4  5
  0  1
  0.100
```

Table 6
Output file for example 4.1.2

MIXREG — The program for mixed-effects linear regression analysis
Autocorrelated error structure: AR(1)
Riesby data — HAMD scores across 6 timepoints — AR1 errors
Two random (int & week) and two fixed (endog & endog*week)

Numbers of observations

Level 2 observations = 66
Level 1 observations = 375

The number of level 1 observations per level 2 unit are:

6 6 6 6 5 5 6 5 5 5 6 5 6 6 6 6 6 6 6
6 6 5 6 6 6 5 4 6 6 5 6 6 6 5 5 5 6 5
6 6 6 5 6 6 6 6 5 6 6 6 5 5 6 6 5 6 6
6 6 6 6 5 6 6 6 6

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	S.D.
HAMD	0.00000	39.00000	17.63733	7.19006
Intercpt	1.00000	1.00000	1.00000	0.00000
Week	0.00000	5.00000	2.48000	1.68320
Endog	0.00000	1.00000	0.54667	0.49848
Endog*Wk	0.00000	5.00000	1.35200	1.74553

Descriptives for response variable HAMD by the variable Week

HAMD

Week	Mean	S.D.	*N*
0.000	23.44262	4.53330	61
1.000	21.84127	4.69800	63
2.000	18.30769	5.48556	65
3.000	16.41538	6.41505	65
4.000	13.61905	6.97097	63
5.000	11.94828	7.21942	58

Starting values

Mean	22.4760	−2.3660	
Covariates	1.9880	−0.0270	
Var. terms	11.6410	−1.4020	2.0770
Residual	12.2180		
Auto terms	0.1000		

Final results — maximum marginal likelihood (MML) estimates

EM iterations = 0
Fisher iterations = 12
Total iterations = 12
Log likelihood = −1103.419

Table 6 (*Continued*)

Variable	Estimate	S.E.	z	P-value
Intercpt	22.46235	0.78692	28.54462	0.00000
Week	−2.32843	0.30288	−7.68770	0.00000
Endog	1.87040	1.06009	1.76438	0.07767
Endog*Wk	−0.01632	0.40807	−0.04000	0.96810
Random-effect variance covariance term(s)				
Intercpt	3.90105	5.30699	0.73508	0.23115
Covariance	0.33980	1.26397	0.26883	0.78806
Week	1.27555	0.57996	2.19939	0.01392
Residual variance				
	15.48884	1.92501	8.04610	0.00000
Autocorrelation term(s)				
	0.37148	0.12212	3.04186	0.00235

note: *P*-values are two-tailed except for variances which are one-tailed

Random-effect covariances expressed as correlations

	1 Intercpt	2 Week
1 Intercpt	1.0000	
2 Week	0.1523	1.0000

Correlation of the MML estimates of the fixed terms

	1 Intercpt	2 Week	3 Endog	4 Endog*Wk
1 Intercpt	1.0000			
2 Week	−0.4468	1.0000		
3 Endog	−0.7423	0.3316	1.0000	
4 Endog*Wk	0.3316	−0.7422	−0.4537	1.0000

Correlation of the MML estimates of variance-related terms

	1 Var-Covl	2 Var-Cov2	3 Var-Cov3	4 Residual	5 AutoCorl
1 Var-Covl	1.0000				
2 Var-Cov2	−0.7883	1.0000			
3 Var-Cov3	0.5639	−0.7408	1.0000		
4 Residual	−0.6999	0.5997	−0.5301	1.0000	
5 AutoCorl	−0.7637	0.6101	−0.5391	0.6854	1.0000

although the conclusions drawn from these estimates do not change. All of the variance parameters are effected by the inclusion of autocorrelation into the model, most notably, the variance in the initial severity levels ($\sigma^2_{\beta_0}$) is considerably reduced when autocorrelation among the residuals is allowed for. Interestingly, this variance term is no longer statistically significant by the Wald test, though some authors [2] argue against using the standard errors to construct hypothesis tests for the random-effect variance terms, particularly when these terms are near-zero. In this model, the

Table 7
Empirical Bayes estimates from example 4.1.2: first ten subjects

101	6	
22.244633	−3.549327	
2.564865	−0.368318	0.422972
103	6	
24.151560	−2.186749	
2.564865	−0.368318	0.422972
104	6	
22.607310	−3.449218	
2.564865	−0.368318	0.422972
105	6	
21.848278	−2.490634	
2.564865	−0.368318	0.422972
106	5	
22.667130	−1.660923	
2.574147	−0.331666	0.567709
107	5	
21.715624	−2.875389	
2.576047	−0.354760	0.439411
108	6	
21.093887	−3.196092	
2.564865	−0.368318	0.422972
113	5	
23.243411	−1.074630	
2.574147	−0.331666	0.567709
114	5	
21.450498	−2.950283	
2.779575	−0.416029	0.433574
115	5	
21.596773	−2.437574	
2.779575	−0.416029	0.433574

standard errors for the variance parameters increase when positive autocorrelation is allowed for.

4.1.3. Empirical Bayes estimates

Both MIXREG.DEF files from the previous two analyses specified the option (on line 7) of writing out to a file (named RIESBY.RES and RIESBYAR.RES, respectively) the EB (empirical Bayes) estimates of the individual random terms. A partial listing of file RIESBYAR.RES is given in Table 7.

The output file RIESBYAR.RES contains four pieces of information per individual: (1) the individual's ID, (2) the number of repeated observations for that individual, (3) the two empirical Bayes estimates, and (4) the associated 2 × 2 posterior variance covariance matrix for these individual estimates (this matrix is output in 'packed' form, thus there are three unique elements of this matrix given).

Table 8
Data from example 4.2: first 18 students from five classrooms and two schools

403	403101	3	1	2	1	0	0
403	403101	4	1	4	1	0	0
403	403101	3	1	4	1	0	0
403	403101	4	1	3	1	0	0
403	403101	4	1	3	1	0	0
403	403101	3	1	4	1	0	0
403	403101	2	1	2	1	0	0
403	403101	4	1	4	1	0	0
403	403101	5	1	5	1	0	0
403	403101	4	1	3	1	0	0
403	403101	3	1	3	1	0	0
403	403101	4	1	3	1	0	0
403	403101	3	1	1	1	0	0
403	403101	4	1	2	1	0	0
403	403101	2	1	2	1	0	0
403	403101	4	1	1	1	0	0
403	403101	4	1	4	1	0	0
403	403101	3	1	3	1	0	0
403	403101	3	1	0	1	0	0
403	403101	4	1	3	1	0	0
403	403102	2	1	0	1	0	0
403	403102	5	1	1	1	0	0
403	403102	3	1	5	1	0	0
404	404101	3	1	1	1	1	1
404	404101	4	1	2	1	1	1
404	404101	2	1	4	1	1	1
404	404101	3	1	3	1	1	1
404	404101	2	1	1	1	1	1
404	404101	1	1	1	1	1	1
404	404101	3	1	2	1	1	1
404	404101	5	1	0	1	1	1
404	404101	2	1	1	1	1	1
404	404101	3	1	2	1	1	1
404	404101	4	1	2	1	1	1
404	404102	3	1	1	1	1	1
404	404102	1	1	1	1	1	1
404	404102	0	1	0	1	1	1
404	404102	4	1	4	1	1	1
404	404102	5	1	1	1	1	1
404	404102	3	1	2	1	1	1
404	404102	3	1	1	1	1	1
404	404102	4	1	2	1	1	1
404	404102	2	1	2	1	1	1
404	404103	3	1	2	1	1	1
404	404103	2	1	1	1	1	1
404	404103	2	1	2	1	1	1
404	404103	3	1	1	1	1	1
404	404103	3	1	1	1	1	1

Table 9
MIXREG.DEF file for example 4.2.1 — student-level analysis ignoring clustering

```
TVSFP study — Post-test tobacco and health knowledge scale
Student-level analysis ignoring the clustering
TVSFP2b.dat
TVSFPs1.out
TVSFPs1.def
TVSFPs1.RES
1  10  8  0  5  0.00010  0  0  1  0  0  0
1  3

4  5  6  7  8
5  7  0.0000  1.0000  2.0000  3.0000  4.0000  5.0000  6.0000
THKScore

Intrcpt PreTHKS   CC      TV      CC*TV
```

4.2. *Analysis of a clustered dataset*

Hedeker, Gibbons, and Flay [32] illustrated use of mixed-effects regression for clustered data using a dataset where students were clustered within classrooms and schools. In that article, the mixed-effects approach was compared with both individual-level analysis which ignores the clustering of the data, and classroom-level analysis which aggregates the individual data. Here, we present use of MIXREG in the clustered data context to perform the mixed-effects analysis, as well as individual-level analysis. Data for this example are from the Television School and Family Smoking Prevention and Cessation Project (TVSFP) [33] which was designed to test independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use prevention and cessation. The initial study sample consisted of seventh-grade students who were pretested in January, 1986. Students who took the pretest completed an immediate post-intervention questionnaire in April, 1986, a 1-year follow-up questionnaire (April, 1987), and a second-year follow-up (April, 1988). The study involved students of schools from Los Angeles and San Diego. Randomization to various design conditions was at the school level, while much of the intervention was delivered to students within classrooms.

For this illustration of the mixed-effects model, a subset of the TVSFP data was used. We concentrated on students from 28 Los Angeles schools, where schools were randomized to one of four study conditions: (a) a social-resistance classroom curriculum, (b) a media (television) intervention, (c) a social-resistance classroom curriculum combined with a mass-media intervention, and (d) a no-treatment control group. These conditions form a 2 × 2 design of social-resistance classroom curriculum (CC = yes or no) by mass-media intervention (TV = yes or no). A tobacco and health knowledge scale (THKS) score was one of the primary study outcome variables, and the one chosen for this analysis. The scale consisted of seven questionnaire items used to assess student tobacco and health knowledge. A student's score on this scale was defined as the sum of the items that the student answered correctly. Only data from the pretest and post-intervention timepoints were analyzed, so subjects were included if they had complete data on the tobacco and health knowledge scale at these two timepoints. In all, there were 1600 students from 135 classrooms and 28 schools who met these criteria. The resulting dataset was highly unbalanced with a range of 1–13 classrooms per school, and 1–28 students per classroom. A partial list of these data is given in Table 8. The variables are, in order, school ID, class ID, post-intervention THKS, column of ones, pre-intervention THKS, CC (with yes = 1 and no = 0), TV (with yes = 1 and no = 0), and the product of CC and TV.

4.2.1. *Fixed-effects regression ignoring data clustering*

Before proceeding with the mixed-effects analysis of these data, we will present a fixed-effects analysis which ignores the clustering of the students. Using MIXREG for this type of analysis is essentially equivalent to performing a traditional multiple linear regression analysis, with the exception that MIXREG uses maximum likelihood, and not least squares, estimation. As noted by Hedeker, Gibbons, and Flay [32] and others,

Table 10
Output file for example 4.2.1 — student-level analysis ignoring clustering

MIXREG — The program for mixed-effects linear regression analysis
TVSFP study — Post-test tobacco and health knowledge scale
Student-level analysis ignoring the clustering

Numbers of observations

Level 2 observations = 1600
Level 1 observations = 1600

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	S.D.
THKScore	0.00000	7.00000	2.66188	1.38293
Intrcpt	1.00000	1.00000	1.00000	0.00000
PreTHKS	0.00000	6.00000	2.06937	1.26018
CC	0.00000	1.00000	0.47687	0.49962
TV	0.00000	1.00000	0.49938	0.50016
CC*TV	0.00000	1.00000	0.23938	0.42684

Descriptives for response variable THKScore by the variable PreTHKS

THKScore			
PreTHKS	Mean	S.D.	*N*
0.000	2.06667	1.24602	150
1.000	2.33890	1.30684	419
2.000	2.61331	1.30709	481
3.000	2.90964	1.32520	332
4.000	3.28834	1.35046	163
5.000	3.70833	1.73767	48
6.000	4.57143	1.51186	7

Starting values

Covariates	1.6613	0.3252	0.6406	0.1987	−0.3216
Residual	1.6929				

Final results — maximum marginal likelihood (MML) estimates

EM iterations = 10
Fisher iterations = 1
Total iterations = 11
Log likelihood = −2688.962

Variable	Estimate	S.E.	*z*	*P*-value
Intrcpt	1.66126	0.08423	19.72363	0.00000
PreTHKS	0.32518	0.02581	12.59805	0.00000
CC	0.64055	0.09196	6.96551	0.00000
TV	0.19871	0.08982	2.21222	0.02695
CC*TV	−0.32162	0.13005	−2.47309	0.01339
Residual variance	1.68763	0.05967	28.28427	0.00000

note: *P*-values are two-tailed except for variances which are one-tailed

Table 10 (*Continued*)

Correlation of the MML estimates of the fixed terms

	1 Intrcpt	2 PreTHKS	3 CC	4 TV	5 CC*TV
1 Intrcpt	1.0000				
2 PreTHKS	−0.6595	1.0000			
3 CC	−0.5364	0.0286	1.0000		
4 TV	−0.5423	0.0188	0.4858	1.0000	
5 CC*TV	0.3653	0.0011	−0.7065	−0.6904	1.0000

ignoring the data clustering often results in statistical tests which are too liberal, resulting in falsely rejecting the null hypothesis too often. In this analysis, the post-intervention THKS score is modeled in terms of baseline THKS score and effects of CC, TV, and CC by TV interaction. Table 9 lists the MIXREG.DEF file for these analyses.

Note that zero random effects are specified in the DEF file, and that blank lines are present for the lines that define the fields and labels of the random effects. In terms of options, a table of post-intervention THKS means by baseline levels is requested. The results of this analysis are listed in Table 10. As seen in the output in Table 10, when zero random effects are requested, MIXREG indicates the same number of level-1 and level-2 observations. For these data, this number is simply the number of students. This analysis, which ignores the data clustering, indicates significant effects of all model covariates: PreTHKS, CC, TV, and CC by TV.

Table 11
MIXREG.DEF file for example 4.2.2 — students-within-classrooms analysis

```
TVSFP study — Post-test tobacco and health knowledge scale
Students nested within Classrooms — 1 random effect
TVSFP2b.DAT
TVSFPc.OUT
TVSFPc.DEF
TVSFPc.RES
1  10  8  1  4  0.00010  0  0  1  0  0  0
2  3
4
5  6  7  8
5  7  0.0000  1.0000  2.0000  3.0000  4.0000  5.0000  6.0000
THKScore
Intrcpt
PreTHKS  CC  TV  CC*TV
```

4.2.2. *Mixed-effects regression including data clustering*

Two mixed-effects regression models can be considered for these data: students within schools, and students within classrooms. At present, MIXREG does not allow a three-level analysis which would consider the students as nested within both classrooms and schools concurrently. To perform the students-within-classrooms analysis the class ID (the second variable field in the datafile) would be indicated as the cluster ID on line 8 of the DEF file, while to perform the students-within-schools analysis the school ID (the first variable field) would be indicated as the cluster ID. Table 11 lists the MIXREG.DEF file for the students-within-classrooms analysis. Again, in this analysis, the THKS score is modeled in terms of baseline THKS score and effects of CC, TV, and CC by TV interaction. However, in contrast to the analysis of the previous section, a random classroom effect is included to account for the data clustering.

As described in Hedeker, Gibbons, and Flay [32], results from the two 2-level mixed-effects regression models are very similar though the students-within classroom analysis accounts for a larger clustering effect than does the students-within schools analysis. Table 12 lists the MIX-

Table 12
Output file for example 4.3 — students-within-classrooms analysis

MIXREG — The-program for mixed-effects linear regression analysis
TVSFP study — Post-test Tobacco and Health Knowledge Scale
Students nested within Classrooms — one random effect

Numbers of observations

level 2 observations = 135
Level 1 observations = 1600

The number of level 1 observations per level 2 unit are:

20	3	11	9	5	26	11	10	15	12	12	10	21	10	17	19	2	4	21
16	15	13	2	14	13	1	12	18	21	17	16	15	16	21	21	27	17	3
2	15	7	24	22	15	19	7	12	8	6	11	7	7	8	3	5	8	3
8	9	8	2	11	9	21	13	12	12	14	9	6	11	10	12	11	6	6
14	10	14	2	3	2	4	3	6	10	14	11	6	22	4	7	22	18	23
19	14	5	14	28	15	15	11	12	11	11	15	17	24	20	15	6	8	14
5	11	9	17	14	11	17	15	6	7	14	10	14	18	4	9	7	12	15
11	10																	

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	S.D.
THKScore	0.00000	7.00000	2.66188	1.38293
Intrcpt	1.00000	1.00000	1.00000	0.00000
PreTHKS	0.00000	6.00000	2.06937	1.26018
CC	0.00000	1.00000	0.47687	0.49962
TV	0.00000	1.00000	0.49938	0.50016
CC*TV	0.00000	1.00000	0.23938	0.42684

Descriptives for response variable THKScore by the variable PreTHKS

THKScore			
PreTHKS	Mean	S.D.	*N*
0.000	2.06667	1.24602	150
1.000	2.33890	1.30684	419
2.000	2.61331	1.30709	481
3.000	2.90964	1.32520	332
4.000	3.28834	1.35046	163
5.000	3.70833	1.73767	48
6.000	4.57143	1.51186	7

Starting values

Mean	1.6613			
Covariates	0.3252	0.6406	0.1987	−0.3216
Var. terms	0.3386			
Residual	1.6929			

Final Results — Maximum marginal likelihood (MML) estimates

EM iterations = 10
Fisher iterations = 7
Total iterations = 17
Log likelihood = −2679.982

Table 12 (*Continued*)

Variable	Estimate	S.E.	*z*	*P*-value
Intrcpt	1.67763	0.09881	16.97808	0.00000
PreTHKS	0.31157	0.02580	12.07625	0.00000
CC	0.63298	0.11863	5.33580	0.00000
TV	0.15966	0.11670	1.3681?	0.17126
CC*TV	−0.27469	0.16780	−1.63698	0.10163
Random-effect variance & covariance term(s)				
Intrcpt	0.08697	0.02765	3.14595	0.00083
Residual variance	1.60301	0.05893	27.20003	0.00000

note: *P*-values are 2-tailed except for variances which are 1-tailed

Calculation of the intracluster correlation

residual variance = 1.603
cluster variance = 0.087
intracluster correlation = 0.087/(0.087 + 1.603) = 0.051

Correlation of the MML estimates of the fixed terms

	1 Intrcpt	2 PreTHKS	3 CC	4 TV	5 CC*TV
1 Intrcpt	1.0000				
2 PreTHKS	−0.5586	1.0000			
3 CC	−0.5885	0.0277	1.0000		
4 TV	−0.5931	0.0190	0.4857	1.0000	
5 CC*TV	0.4082	−0.0055	−0.7066	−0.6953	1.0000

Correlation of the MML estimates of variance-related terms

	1 Var-Cov1	2 Residual
1 Var-Cov1	1.0000	
2 Residual	−0.1656	1.0000

REG results from the students-within classroom analysis.

Notice that the intracluster correlation (in this case, the intraclassroom correlation) from this analysis equals 0.051, while for the students-within-schools analysis (not shown) the intracluster correlation equals 0.022. Also, in contrast to the analysis ignoring the clustering of the students (in Table 10), mixed-effects regression analysis indicates that neither the TV effect nor the interaction of CC by TV are statistically significant at the $P < 0.05$ level. Thus, conclusions regarding model terms can change if the clustering of the data is not appropriately accounted for.

5. Hardware and software specifications

MIXREG is written in standard FORTRAN-77 with double arithmetic precision. It was originally developed for MS-DOS personal computers and later ported over to the MACINTOSH environment. As a result, it's use on MACINTOSH personal computers does not take advantage of the system's menu-orientated interface. For use in

either the MS-DOS or MACINTOSH environment, MIXREG requires a math coprocessor. The program stores all necessary matrices and vectors in a single one-dimensional array. Thus there are no fixed limitations on the numbers of level-2 units, level-1 units, or model variables. MIXREG utilizes many MATCAL subroutines [34] for performing the matrix algebra operations.

6. Availability

The MIXREG program is available at no charge. Those interested in obtaining a copy of the program should contact the authors by electronic mail at HEDEKER@UIC.EDU or send a blank diskette by normal mail to Ann Hohmann, Ph.D., M.P.H., NIMH Services Research Branch, 5800 Fishers Lane, Room #10C-06, Rockville MD, 20857.

Acknowledgements

The development of MIXREG was supported by the National Institutes of Mental Health Grant MH44826-01A2. The authors wish to thank Dr. R. Darrell Bock for many valuable discussions, comments, and suggestions during the preparation of this program; the authors also thank Dr. Bock for providing the code for the MATCAL [34] matrix algebra subroutines. The authors also thank Dr. Brian Flay for providing the clustered data-set, Dr. N. Reisby for the longitudinal data-set, and especially thank David Patterson of Discerning Systems, Inc. for development of the interactive user interface.

References

[1] H. Goldstein, Multilevel Models in Educational and Social Research (Oxford University Press, New York, 1987).

[2] A.S. Bryk and S.W. Raudenbush, Hierarchical Linear Models: Applications and Data Analysis Methods (Sage Publications, Inc., Newbury Park CA, 1992).

[3] N.T. Longford, Random Coefficient Models, (Oxford University Press, New York, 1993).

[4] S.R. Searle, Linear Models for Unbalanced Data (Wiley, New York, 1987).

[5] N.M. Laird and J.H. Ware, Random-effects models for longitudinal data, Biometrics 40 (1982) 961–971.

[6] R.D. Bock, Within-subject experimentation in psychiatric research, in: Statistical and Methodological Advances in Psychiatric Research, Eds. R.D. Gibbons and M. Dysken, pp. 59–90 (Spectrum, New York, 1983).

[7] A.P. Dempster, D.B. Rubin and R.K. Tsutakawa, Estimation in covariance components models, J. Am. Stat. Assoc. 76 (1981) 341–353.

[8] R.D. Bock, R.D., Measurement of human variation: a two stage model, in: Multilevel Analysis of Educational Data, Eds. R.D. Bock, pp. 319–342 (Academic Press, New York, 1989).

[9] J. DeLeeuw and I. Kreft, Random coefficient models for multilevel analysis, J. Educ. Stat. 11 (1986) 57–85.

[10] N.T. Longford, A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects, Biometrika 74 (1987) 817–827.

[11] J.F. Strenio, H.I. Weisberg and A.S. Bryk, Empirical Bayes estimation of individual growth curve parameters and their relationship to covariates, Biometrics 39 (1983) 71–86.

[12] R.I. Jennrich and M.D. Schluchter, Unbalanced repeated-measures models with structured covariance matrices, Biometrics 42 (1986) 805–820.

[13] R.D. Bock, The discrete Bayesian, in: Principles of Modern Psychological Measurement, Eds. H. Wainer and S. Messick, (Earlbaum, Hillsdale NJ, 1983).

[14] R.D. Gibbons, D. Hedeker, C. Waternaux and J.M. Davis, Random regression models: A comprehensive approach to the analysis of longitudinal psychiatric data, Psychopharmacol. Bull. 24 (1988) 438–443.

[15] D. Hedeker, R.D. Gibbons, C. Waternaux and J.M. Davis, Investigating drug plasma levels and clinical response using random regression models, Psychopharmacol. Bull. 25 (1989) 227–231.

[16] D. Hedeker, R.D. Gibbons and J.M. Davis, Random regression models for multicenter clinical trials data, Psychopharmacol. Bull. 27 (1991) 73–77.

[17] R.D. Gibbons, D. Hedeker, I. Elkin, C. Waternaux, H.C. Kraemer, J.B. Greenhouse, M.T. Shea, S.D. Imber, S.M. Sotsky and J.T. Watkins, Some conceptual and statistical issues in analysis of longitudinal psychiatric data, Arch. Gen. Psychiatry 50 (1993) 739–750.

[18] A.S. Bryk, and S.W Raudenbush, Applications of hierarchical linear models to assessing change, Psychol. Bull. 101 (1987) 147–158.

[19] M.J. Lindstrom, and D.M. Bates, Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data, J. Am. Stat. Assoc. 404 (1988) 1014–1022.

[20] E.M. Chi, and G.C. Reinsel, Models for longitudinal data with random effects and AR(1) errors, J. Am. Stat. Assoc. 84 (1989) 452–459.

[21] D. Hedeker, Random Regression Models with Autocorrelated Errors: Investigating Drug Plasma Levels and Clinical Response, Doctoral dissertation (The University of Chicago, Chicago, 1989).

[22] A.S. Bryk, S.W. Raudenbush, and M. Seltzer, An in-

troduction to HLM: computer program and users' guide, (Scientific Software, Inc., Chicago, 1994).

[23] R. Prosser, J. Rasbash, and H. Goldstein, ML3 software for three-level analysis, users' guide for v.2., (Institute of Education, University of London, London UK, 1991).

[24] N.T. Longford, VARCL — Interactive software for variance component analysis: applications for survey data, Prof. Stat. 5 (1986) 28–32.

[25] I.G.G. Kreft, J. De Leeuw, R. Van Der Leeden, Review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML3, VARCL, Am. Stat. 48 (1994) 324–335.

[26] H. Mansour, E.V. Nordheim, and J.J. Rutledge, Maximum likelihood estimation of variance components in repeated measures designs assuming autoregressive errors, Biometrics 41 (1985) 287–294.

[27] R.D. Bock, Multivariate Statistical Methods in Behavioral Research (McGraw-Hill, New York, 1975).

[28] A. Wald, Tests of statistical hypotheses concerning several parameters when the number of observations is large, Trans. Am. Math. Soc. 54 (1943) 426–482.

[29] L.F. Gram, N. Reisby, I. Ibsen, A. Nagy, S.J. Dencker, P. Bech, G.O. Petersen, and J. Christiansen, Plasma levels and antidepressive effect of imipramine, Clin. Pharmacol. Ther. 19 (1976) 318–324.

[30] N. Reisby, L.F. Gram, P. Bech, A. Nagy, G.O. Petersen, J. Ortmann, I. Ibsen, S.J. Dencker, O. Jacobsen, O. Krautwald, I. Søndergaard, and J. Christiansen, Imipramine: clinical effects and pharmacokinetic variability, Psychopharmacology 54 (1977) 263–272.

[31] R.D. Gibbons, D.C. Clark, and J.M. Davis, A statistical model for the classification of imipramine response in depressed inpatients, Psychopharmacology 78 (1982) 185–189.

[32] D. Hedeker, R.D. Gibbons, and B.R. Flay, Random-effects regression models for clustered data: with an example from smoking prevention research, J. Consult. Clin. Psychol. 62 (1994) 757–765.

[33] B.R. Flay, B.R. Brannon, C.A. Johnson, et al. The television, school and family smoking cessation and prevention project: I. theoretical basis and program development, Prev. Med. 17 (1989) 585–607.

[34] R.D. Bock and B.H. Repp, MATCAL: Double Precision Matrix Operations Subroutines (National Educational Resources, Chicago, 1974).