5. The above will then implement TERRACE-TWO with the specified data set. Use the dialogs to specify the structure of the multilevel model as done in the previous section.
6. Have fun!
7. Mail comments or suggestions to afsharto@math.ucla.edu or jhilden@math.ucla.edu

5 Summary

Researchers from various fields are beginning to appreciate the intuitive appeal of viewing nested data through multilevel models. However, this intuitive appeal is felled by practical limitations. Even amongst statisticians, experience with multilevel models is quite rare. Thus, the researcher interested in employing multilevel analysis is often left at an impasse. This guide has hopefully provided the first step out of such impasse, clearing the way for further work. TERRACE-TWO, through its use of dialogs, provides an interactive environment for model estimation. In addition, since TERRACE-TWO is based upon the XLISP-STAT's object oriented structure, all the advantages of using XLISP-STAT naturally follow. For those interested in estimating the models illustrated in this paper, the necessary files have been included in this directory. The files are as follows:

- ter2.lsp (The TERRACE-TWO program).
- nels.lsp (A file containing the school and student data matrices, along with the corresponding variable labels).
- ter2-mod.lsp (A file that directs the data into TERRACE-TWO and implements the program.)

The last file has been altered to allow the user to sample from the schools (1052 total) so that the program runs faster. One needn't worry about sampling from the student data since this will be performed automatically by TERRACE-TWO. Thus, after one has the aforementioned files, the following steps need to be performed.

1. Get into XLISP-STAT
2. Load ter2-mod.lsp
3. Type (sample-schools n), where n is the desired number of schools one wishes to sample from the entire 1052 schools.
4. Pick the desired variables from the dialogs that pop up, making sure not to forget the dependent variable, math score (BYTXMSTD).

```
TERRACE-TWO:  Full Maximum Likelihood Estimates


Parameters        Estimates     (S.E.)        T

Intercept
  By Intercept      51.0799    (  0.9301)    54.9212
  By Private        10.5206    (  4.1617)     2.5280

BYSES
  By Intercept       4.0932    (  0.5461)     7.4956


Sigma^2:   71.4388

          Tau (covariance)

Intercept      13.3731
```

Thus, we see that the removal of the error term for the slope equation is satisfactory.

The example presented here is intentionally simple, merely touching on some of the capabilities of TERRACE-TWO. Other options available from the dialogs window include group-mean centering of level-one variables and centering of level-two variables. However, competent use of such options requires a thorough understanding of multilevel model theory. TERRACE-TWO is also equipped with a plethora of diagnostic capabilities. To be sure, multilevel model diagnostics are not nearly as well understood as ordinary linear regression diagnostics. Thus, a thorough discussion of the do's and don'ts of multilevel model diagnostics is beyond the scope of this guide. The interested reader may consult —– by James Hilden-Minton.

```
          Tau (correlation)

Intercept       1.0000   0.8575
BYSES           0.8575   1.0000
```

The average Math score for non-private schools is 49.78; Private schools have a 8.75 point advantage. Moreover, the average math score has a variance of 12.62 from school to school. Regarding the strength of the association between SES and Math score, the average SES slope is 4.04, ie., on average, a one unit increase in SES translates into a 4.04 unit increase in Math score. The SES-Math association has a 1.04 school to school variance. Given this low school to school variance of SES slope, one may question whether this coefficient should be modeled as random. We may test this hypothesis by eliminating the appropriate error term: under "Covariance Parameters," we click on SES. The dialog indicates that Intercept and SES are the chosen covariance parameters to estimate with SES. Since we desire a non-randomly varying level-two equation for SES, we click on these variables again to unselect them. Thus, no covariance parameters are estimated with SES. We click on the OK box, then once again click on Recompute Estimates. Our new model is as follows:

```
Maximizing Likelihood...

                       Deviance    Method

      Iteration  10:  3630.6440    EM, init.
      Iteration  11:  3630.5886    Fisher
      Iteration  12:  3630.5882    Fisher
Final Iteration  13:  3630.5882    Fisher

Summary of Change in Deviance...

      Old Dev.:   3629.0044   df: 499.0
      New Dev.:   3630.5882   df: 501.0
        Change:      1.5838   df: 2.0      P: 0.45297

Butthead says, "Cool."
```

```
> Maximizing Likelihood...

                         Deviance    Method

      Iteration   1:  4291.2876    EM, init.
      Iteration   2:  4259.4213    Fisher
      Iteration   3:  4251.7654    Fisher, 1 half-steps
      Iteration   4:  4249.0796    Fisher, 1 half-steps
      Iteration   5:  4247.7860    Fisher
      Iteration   6:  4247.6570    Fisher
      Iteration   7:  4247.5375    Fisher, 1 half-steps
      Iteration   8:  4247.5108    Fisher
      Iteration   9:  4247.5062    Fisher
Final Iteration  10:  4247.5082    Fisher


TERRACE-TWO:  Full Maximum Likelihood Estimates


Parameters        Estimates     (S.E.)        T

Intercept
  By Intercept      49.7827    (  0.8251)    60.3364
  By Private         8.7482    (  2.5445)     3.4380

BYSES
  By Intercept       4.0372    (  0.5741)     7.0318


Sigma^2:   62.8794

          Tau (covariance)

Intercept      12.6249   3.1035
BYSES           3.1035   1.0375
```

ables this regressor's error term is correlated with. In essence, this dialog is used to specify the structure of the T matrix mentioned earlier.

Let's say one is interested in the following model:

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \beta_{1j}\text{SES} + r_{ij} \\ \beta_{0j} &= \gamma_{00} + \gamma_{01}\text{Private} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned}$$

where we assume the following:

$$\begin{aligned} r_{ij} &\sim N(0, \sigma^2) \\ \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} &\sim N(0, T) \end{aligned}$$

and,

$$\begin{aligned} T &= \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} \\ Cov(u_{0j}, r_{ij}) &= Cov(u_{1j}, r_{ij}) = 0 \end{aligned}$$

We estimate this model as follows. Under "Linear Parameters" we first click on Intercept. Since we want a level-two equation for the intercept, we specify the desired level-two equation by clicking on Intercept and Private in the new box that pops up. Next, we choose SES from "Linear Parameters," but we do not cross SES with any level-two variables. Finally, we specify the error structure of the covariance terms, choosing both level-one coefficients to be random and correlated with each other. Thus, under "Covariance Parameters" we cross each level-one coefficient with itself (making it random, i.e., vary) and the other level-one coefficients (making it covary). After clicking on the "Recompute Estimates" box, TERRACE-TWO attempts to estimate the model, through either Fisher Scoring or the EM Method (the method used at each iteration is specified in the output). If the convergence criterion has not been met after five iterations, the user is prompted for instructions concerning how many more iterations should be carried out. Of course, one could obtain the current estimates by specifying no further iterations. After ten iterations, we get the following output:

this particular example, the program is sitting in a remote directory which must be specified.[1] The next line, (load "nels"), loads the file nels.lsp. This file should contain several things. It should contain the two matrices for the two levels of data, and it should contain two lists for the variable names at each level. The command in XLISP-STAT was used to collect these items into this file. The next line channels the data into TERRACE-TWO, thereby creating a multilevel object called ter1. The names for the two matrices and two lists of variable names are given as school-data, student-data, x-labels, and labels, respectively.

After the above file is loaded into XLISP-STAT, the TERRACE-TWO program will allow the user to estimate multilevel models on one's data set through the use of dialog windows. Moreover, the dialogs incorporate the variable names of one's data set, thereby customizing the output. One merely needs to understand the dialogs in order to harness the capability of TERRACE-TWO.

4 An example: NELS:88

Let us walk through a multilevel model example to illustrate the use of TERRACE-TWO. Once the aforementioned set-up file has been created for the multilevel data set, this file must be loaded into XLISP-STAT. This will initiate the TERRACE-TWO program. A dialog window will then pop up, which contains several sub-boxes. One of these is entitled "Linear Parameters." This box contains the possible level-one variables to include in the model. Click on all the variables that one desires to include in the level-one model. When one clicks on a given linear parameter, another dialog window appears. The reader is prompted to "choose the group variables to cross with . . ." (i.e., the the chosen linear parameter) from all the possible level-two variables. Thus, one specifies the level-two model for a given level-one variable accordingly. The other main box in the original dialog is titled "Covariance Parameters" and includes a list of all possible level-one variables. For each level-one variable chosen, one determines two things: 1) whether or not its coefficient is modeled as random, and 2) which other level-one vari-

[1]If you place the program in the directory in which you run XLISP-STAT, such specification will be unnecessary.

zero). Such parameters will be called *fixed*. This is distinct from the usage in the context of mixed models where effects are said to be fixed, random, or non-randomly varying. Such modeling choices may be activated by *fixing* parameters or including new (indicator) variables. Thus, the computational problem of model specification becomes a problem of restricted parameter estimation.

3 Getting Started with TERRACE-TWO

Given the theory covered in the previous section, there could be a variety of issues of interest. For example, one might be mainly interested in using multilevel models to get improved estimates of individual effects (i.e., the β's). Or, one might be interested in using multilevel models to estimate fixed effects, γ's, in order to examine the effect of the level-two units on level-one regressions (e.g., the effect of the schools on the students in our educational example).

Regardless of one's research interests, one must manipulate one's data within XLISP-STAT to produce two matrices, one for level-one units and one for level-two units. The former matrix should be $N \times P$, where N is the total number of level-one units and P is the total number of level-one variables. The latter matrix should be $J \times F$, where J is the total number of level-two units and F is the total number of level-two variables.

Next, the user must write an XLISP-STAT file so that his/her data can be incorporated into the TERRACE-TWO program appropriately. Here is an example of such a file:

```
(load "/c2/laplace/sgrad/jhilden/terrace/ter2")
(load "nels")

(setf ter1 (make-terrace school-data student-data
             :x-labels stud-labels
             :z-labels schl-labels))
```

When loaded within XLISP-STAT, the above file does several things. The first line of the file loads the TERRACE-TWO program, ter2.lsp. In and

$$W_j = \begin{pmatrix} z_j & 0 & \cdots & 0 \\ 0 & z_j & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & z_j \end{pmatrix}.$$

Note that W_j is a $(P \times QP)$-matrix. Now our second level model may be completely specified by (3) when we impose that each $u_j \overset{iid}{\sim} (0, \tau)$. Also we impose that $(u_j, r_{j'}) = 0$ for any j and j'. This model will be referred to as level-two model, and we will also define an object to represent this model. Level-one objects will be embedded in a level-two object. Thus, the multilevel model will be represented as a network of models.

Combining (1) and (3), one obtains

$$Y_j = X_j W_j \gamma + X_j u_j + r_j. \tag{4}$$

Evidently,

$$Y_j - X_j W_j \gamma = X_j u_j + r_j \sim (0, V_j), \tag{5}$$

where

$$V_j = X_j \tau X_j' + \sigma^2 I_{n_j}.$$

Thus, the full log-likelihood for the jth unit is

$$L_j(\sigma^2, \tau, \gamma) = -\frac{n_j}{2}\log(2\pi) - \frac{1}{2}\log|V_j| - \frac{1}{2}d_j' V_j^{-1} d_j, \tag{6}$$

where $d_j = Y_j - X_j W_j \gamma$. Since the J units are independent, we write the log-likelihood for the entire model as a sum of unit log-likelihoods, *i.e.*,

$$L(\sigma^2, \tau, \gamma) = \sum_{j=1}^{J} L_j(\sigma^2, \tau, \gamma). \tag{7}$$

This decomposition of the log-likelihood function, implies that the information matrix and score function (first and second derivatives) may likewise be written as sums of unit components. These facts motivate the object structure discussed in the next section.

Thus far, I have presented what may be called a maximal model, meaning all parameters in σ^2, τ, and γ are to be estimated or *free*. Reduced models may be specified by setting certain parameters to arbitrary values (usually of the school may increase/decrease the slopes of the student level equation. In more practical terms, students may be significantly helped/hindered by their school.

This paper consists of three parts. Section two formally lays out the theory of multilevel models. Section three discusses how to get started with TERRACE-TWO. Section four takes the reader through an illustration using TERRACE-TWO to an actual data set. It is assumed that the reader has a basic understanding of XLISP-STAT.

2 Description of a multilevel model

Suppose we have N subjects naturally grouped into J units, where there are n_j subjects in the jth unit and $\sum_{j=1}^{J} n_j = N$. Further suppose that for the J units we want to regress the response variable Y_j on matrix of P predictor variables X_j. Thus, for the jth unit we model

$$Y_j = X_j \beta_j + r_j, \tag{1}$$

where each X_j has dimensions $n_j \times P$, and

$$r_j \sim n(0, \sigma^2 I_{n_j}).$$

These models will be referred to as level-one models, and we will define objects to represents these models.

At the next level, we want to model each β_{jp} ($j = 1, 2, \ldots, J$, and $p = 1, 2, \ldots, P$) with

$$\beta_{jp} = z_j' \gamma_p + u_{jp}, \tag{2}$$

where z_j is an vector of Q background variable on the jth unit and $u_{jp} \sim (0, \tau_{pp})$. But the u_{jp} are not independent; to get at their covariance structure, we "stack" the equations in (2) to obtain

$$\beta_j = W_j \gamma + u_j, \tag{3}$$

where

$$u_j = (u_{j1}, u_{j2}, \ldots, u_{jP})',$$

$$\gamma = (\gamma_1', \gamma_2', \ldots, \gamma_P')',$$

Terrace-Two User's Guide: An XLISP-STAT Package for Estimating Multi-Level Models

David Afshartous
James Hilden-Minton

August 29, 1994

1 Introduction

Statisticians and social scientists must often analyze data that comes in hierarchical form. A classic example is educational data, where students are nested within schools. There exists a growing literature concerning the statistical techniques that should be employed to analyze such data. To be sure, one technique is to ignore the hierarchical structure in the data and merely employ conventional techniques, by either aggregating or disaggregating the data. However, the problems with such a route are well documented. (ecological fallacy, observations correlated by definition, etc.) The estimation of multilevel models provides an outlet for those wishing to perform more sophisticated data analysis.

In a multilevel model, there are regression equations for each level of data. Moreover, the coefficients for all but the highest level of data are treated as random variables (note that a fixed constant is a special case of a random variable.) For instance, in educational research, we have a two-level model. At level-one, one specifies a student regression equation, where the coefficients of the regression are random variables. At level-two, these coefficients are the response variable in school-level regressions. Thus, the characteristics